A CHEN-FLIESS APPROXIMATION FOR DIFFUSION FUNCTIONALS

CHRISTIAN LITTERER AND HARALD OBERHAUSER

ABSTRACT. We show that an interesting class of functionals of stochastic differential equations can be approximated by a Chen-Fliess series of iterated stochastic integrals and give a $L^2$ error estimate, thus generalizing the standard stochastic Taylor expansion. The coefficients in this series are given a very intuitive meaning by using functional derivatives, recently introduced by B. Dupire.

1. Introduction

The expansion of the SDE solution

\[ dY_t = \sum_{i=1}^{d} V_i(Y_t) \circ dB_i^t, \quad Y(0) \in \mathbb{R}^e \]

in a stochastic (Stratonovich) Taylor series

\[ f(Y_t) - f(Y_s) = \sum_{k=1}^{N} \sum_{(i_1, \ldots, i_k) \in \{1, \ldots, d\}^k} V_{i_1} \cdots V_{i_k} f(Y_s) \int_{\Delta^k(s,t)} \circ dB^{(i_1, \ldots, i_k)} + R_N(s,t) \]

is an important tool in the numerical analysis of stochastic differential equations, cf. [14]. In this short note we consider a Stratonovich stochastic Taylor expansion in the spirit of Kloeden and Platen [14] and generalize the classic formula (1.2) that applies to smooth functions $f : \mathbb{R}^e \to \mathbb{R}$ by giving meaning to such an approximation when $f$ is replaced by a sufficiently regular, nonanticipative functional $F : [0,T] \times C([0,T], \mathbb{R}^e) \to \mathbb{R}$ and derive a $L^2$ bound for the remainder of the expansion. The crucial tool in obtaining these results are ideas of B. Dupire [9] on derivatives of nonanticipative functionals (cf. the extensions of Cont&Fournie [8, 7] and the work on pathdependent viscosity PDEs [16, 17, 10]). Dupire showed that one can define a time derivative $\partial_t F$ and a space derivative $\nabla F$ of $F$ such that one arrives at the approximation (for simplicity assume $e = 1$)

\[ dF \approx \partial_t F dt + \nabla F dB_t + \frac{\nabla^2 F}{2} (dB_t)^2. \]

The intuition behind these derivatives is that the time derivative measures the infinitesimal drift of $F$ whereas $\nabla F$ measures the response of $F$ to instantaneous changes in the underlying path. In section 3 we show that these derivatives allow to give proper meaning to the expression $V_{i_1} \cdots V_{i_k} F(s,Y)$ if one interprets a vector field in a corresponding way as a derivation in terms of these functional derivatives (e.g. if $F(t,x) = f(x_t)$ this simply coincides with the standard notion of $V_{i_1} \cdots V_{i_k}$ as a differential operator). This subsequently allows us to adapt the standard estimates for (1.2) in the functional setting.

The functional perspective of the stochastic Taylor expansion suggests some interesting connections to problems studied in deterministic, classic control theory: many systems are modeled

Key words and phrases. Chen-Fliess series, stochastic differential equations, stochastic Taylor expansion.
as a nonlinear, causal response to a $d$-dimensional input $(u^1, \ldots, u^d) \in C([0,T], \mathbb{R}^d)$ and control theory provides a powerful toolbox to deal with functionals of the form

$$F : [0,T] \times C([0,T], \mathbb{R}^d) \rightarrow \mathbb{R}$$

which are nonanticipative in the sense that $F(t, u^1, \ldots, u^d)$ depends only on values of $(u^1, \ldots, u^d)$ up to time $t$ and depend continuously on the input $(u^1, \ldots, u^d)$. The basic example being the solution map of a controlled differential equation

$$\frac{dy}{dt}(t) = \sum_{i=1}^{d} V_i(y(t)) \ u^i(t), \ y(0) \in \mathbb{R}$$

i.e. $F(t, u^1, \ldots, u^d) = y(t)$. We mention - pars pro toto - the work of Brockett, Fliess and Sussmann, [3, 12, 20], which emphasized the importance of the signature of a path in the sense of Chen [5, 6] as a central object in control theory. It is a well known fact to any system theorist or electrical engineer that a functional $F$ that is linear 

$$F(t, \sum_{i=1}^{d} \alpha_i u^i + \beta_i v^i) = \sum_{i=1}^{d} \alpha_i F(t, u^i) + \beta_i F(t, v^i)$$

and additionally time-invariant is completely described by its response to a Dirac impulse $\delta$ at time 0 (clearly, the functional given by (1.3) does not fit into this framework). As mentioned above, Dupire’s functional calculus rests on two derivatives and the space derivative is, at least formally, closely related to the well known method in control theory of response to a Dirac delta (however in integrated form “$\int_0^t (u_r + \epsilon \delta) \ dr = \int_0^t u_r \ dr + \epsilon, u = \frac{dB}{dt}$” and at running time). Our interest lies in the stochastic case, but these functional derivatives might already be of interest to the classic control theory case for non-linear, non-time-invariant systems of the form (1.3). In this context we recall in section 2 a result of M. Fliess, [12] on the existence of approximations to continuous, nonanticipative functionals (this result of Fliess is the reason why we refer to the expansion in section 3 as a Chen-Fliess approximation instead of stochastic Taylor expansion for functionals).

Of course, one can see the finite expansion (1.2) as a small piece in the seminal work initiated by Azencott and Ben Arous [1, 2] on asymptotic expansions of SDE flows and subsequent generalizations of e.g. Castell and Hu [3, 13] using the generalized Campbell-Baker-Hausdorff formula of Strichartz [18]. From this point of view, the present note is certainly only a first step but we should note that our interest for a Taylor expansion for functionals stems from its potential use in a cubature scheme in the sense of Kusuoka-Lyons-Victoir [15] (which we hope to address in forthcoming work) where it replaces (1.2) and, similar to (1.2), it might be of independent interest. Finally, let us mention the classic results of Stroock, [14], who considered a Taylor expansion of functionals in the spirit of Malliavin’s calculus of variations. However, our results are different and tailored to the study of the smaller class of continuous (in uniform norm) and nonanticipative functionals. In fact, thanks to the extended functional Ito formula of Cont&Fournie, it is easy to relax the continuity assumption to include e.g. functionals of the quadratic variation (the same arguments then apply since the domain of the functional changes but only perturbations of the path are relevant, cf. [8, theorem 4.1]). Nevertheless, we focus our presentation on smooth and continuous functionals since they already cover interesting applications like pathdependent financial derivatives, cf. [9], and keep the (necessary) new notation to a minimum, allowing for a concise presentation of the present approach.
2. Approximating Functionals of bounded variation paths

In this section we adapt some techniques from control theory, introduce notation and recall a classical existence theorem which demonstrates that continuous functionals on continuous, bounded variation paths can be uniformly approximated by linear combinations of iterated integrals, the (truncated) Chen-Fliess series. In section 3 we then extend this result to diffusion functionals using a new approach with Dupire’s functional derivatives. Set

\[ \Lambda_{1-\text{var}} = [0, T] \times C^{1-\text{var}} ([0, T], \mathbb{R}^d) \]

with \( C^{1-\text{var}} ([0, T], \mathbb{R}^d) \) being the space of continuous paths \( b : [0, T] \rightarrow \mathbb{R}^d \) which are of bounded variation,

\[ |b|_{1-\text{var};[0,T]} = \sup_\mathcal{D} \sum_\mathcal{D} |b(t_i) - b(t_{i-1})| < \infty \]

with the sup_\mathcal{D} taken over all finite dissections of \([0, T]\). We are interested in functionals \( F : \Lambda_{1-\text{var}} \rightarrow \mathbb{R} \) which are nonanticipative in the sense that \( F(t, b) \) depends only on \( \{b(r), r \leq t\} \). A basic example is the solution map

\[ F(t, b^1, \ldots, b^d) = y(t) \]

mapping the “input” \( b = (b^1, \ldots, b^d) \) to the real valued solution \( y \) of the integral equation

\[ y(t) = y(0) + \int_0^t V(y_r) \, db_r. \]

\[ = y(0) + \sum_{i=1}^d \int_0^t V_i(y_r) \, db^i_r \]

Such a map is continuous (provided the vector fields are bounded, Lipschitz) if \([0, T] \times C^{1-\text{var}} ([0, T], \mathbb{R}^d) \) is equipped with the product topology. However, for later purposes it will be convenient to work with the weaker topology on \( \Lambda_{1-\text{var}} \) induced by the metric

\[ \rho_{1-\text{var}} ((t, b), (s, e)) = |t - s| + |a_t b - a_s e|_{1-\text{var};[0,T]} \]

where \( a_t \) denotes the stopping operator

\[ a_t : C^{1-\text{var}} ([0, T], \mathbb{R}^d) \rightarrow C^{1-\text{var}} ([0, T], \mathbb{R}^d) \]

\[ (a_t x)(r) = \begin{cases} x(r), & r \leq t \\ x(t), & r > t. \end{cases} \]

The proposition below then follows from standard arguments.

**Proposition 1.** The space \( (\Lambda_{1-\text{var}}, \rho_{1-\text{var}}) \) is a complete metric space and the topology on \( \Lambda_{1-\text{var}} \) given by the metric \( \rho_{1-\text{var}} \) is strictly weaker than the topology induced by the product topology of \([0, T] \times C^{1-\text{var}} ([0, T], \mathbb{R}^d) \), the latter equipped with \( |.|_{1-\text{var}} \) convergence.

We will in the following (unless stated otherwise) always work in the topology on \( \Lambda_{1-\text{var}} \) induced by \( \rho_{1-\text{var}} \). Set

\[ C(\Lambda_{1-\text{var}}) = \{ F : \Lambda \rightarrow \mathbb{R} | F \text{ is nonanticipative and continuous wrt } \rho_{1-\text{var}} \} \]
Definition 2. If \( b \in C^{1-var}([0,T], \mathbb{R}^d) \), define for every word \((i_1, \ldots, i_k) \in \{0,1,\ldots,d\}^k\), and \( s, t \in [0,T], s \leq t \), the iterated integral \( \int_{\Delta^k(s,t)} db^{(i_1,\ldots,i_k)} \) recursively as

\[
\int_{\Delta^1(s,t)} db^{(0)} = t - s \quad \text{and} \quad \int_{\Delta^1(s,t)} db^{(i)} = b^i(t) - b^i(s) \quad \text{if } i \neq 0
\]

\[
\int_{\Delta^{k+1}(s,t)} db^{(i_1,\ldots,i_k)} = \int_s^t \int_{\Delta^k(s,r)} db^{(i_1,\ldots,i_{k-1})} db^{i_k}.
\]

We now show that the special class of polynomial functionals are dense in \( C(\Lambda_{1-var}) \).

Definition 3. Call a functional \( P : \Lambda_{1-var} \to \mathbb{R} \) a polynomial functional if there exists a \( N \in \mathbb{N} \) and \((p_I)_{I=(i_1,\ldots,i_k)} \subset \mathbb{R} \) s.t.

\[
P(t, b^1, \ldots, b^d) = \sum_{k=1}^N \sum_{I \in \{0,\ldots,d\}^k} p_I \int_{\Delta^k(0,t)} db^I.
\]

Remark 4. The term polynomial is made more rigorous by using a morphism with the non-commutative algebra \( \mathbb{R}[X] \) over a finite alphabet \( X = \{x_0, \ldots, x_d\} \), cf. [12].

Proposition 5. Every polynomial \( P : \Lambda_{1-var} \to \mathbb{R} \) is continuous with respect to \( p_{1-var} \), i.e. an element of \( C(\Lambda_{1-var}) \). Further, the set of polynomials is a subalgebra of \( C(\Lambda_{1-var}) \).

The theorem below is a straightforward extension of a well-known theorem of control theory, cf. [20, 12, theorem II.5] to bounded variation (but not necessarily absolutely continuous) paths in the topology \( p_{1-var} \). Note that it relies on the Stone-Weierstrass theorem. In section 3 we show that under stronger assumptions on the functional (and in a semimartingale context) one can choose the coefficients as functional derivatives in the sense of Dupire (this is similar to the relation of the classic Taylor expansion, applicable to smooth functions, to the Weierstrass approximation theorem which applies to continuous functions).

Theorem 6. Let \( K \) be a compact subset of \( C^{1-var}([0,T], \mathbb{R}^d) \) and

\[ F : [0,T] \times K \to \mathbb{R} \]

be continuous with respect to \( p_{1-var} \). Then \( F \) can be arbitrarily close approximated in uniform topology by polynomial functionals, that is \( \forall \epsilon > 0 \) exists a polynomial functional \( P : \Lambda_{1-var} \to \mathbb{R} \) such that

\[ \sup_{(t,b) \in [0,T] \times K} |F(t,b) - P(t,b)| < \epsilon. \]

Proof. The set \([0,T] \times K\) is a compact subset of \([0,T] \times C^{1-var}([0,T], \mathbb{R}^d)\) if the latter set is equipped with the product topology of Euclidean distance on \([0,T]\) and \( |\cdot|_{1-var}\). To see that it is also a compact subset in the topology given by \( p_{1-var} \), take a sequence

\[ (t^n, x^n) \subset [0,T] \times K \]

By compactness there exists a \( (t,x) \in [0,T] \times K \) and a sub-sequence \((n_k) \) s.t. \( |t^{n_k} - t| \to 0 \) and \( |x^{n_k} - x|_{1-var;[0,T]} \to x \). Therefore also

\[ \rho(t^{n_k}, x^{n_k}) = |t^{n_k} - t| + |a_t x_t^{n_k} - a_t x|_{1-var;[0,T]} \to 0, \]

hence \([0,T] \times K\) is under the \( p_{1-var} \) metric a compact subset of \( \Lambda_{1-var} \). We now use the Stone-Weierstrass theorem: \([0,T] \times K, p_{1-var}\) is a compact, Hausdorff space and from proposition we know that the space of polynomials is a subalgebra of \( C(\Lambda_{1-var}) \), hence (strictly speaking the space polynomials with domains restricted to \([0,T] \times K\) are also a subalgebra of \( C([0,T] \times K) \).
Further, the space of polynomials contains the nonzero, constant functions and it remains to verify that it separates points. Given two elements of $[0, T] \times \mathcal{K}$,

$$(t, b^1, \ldots, b^d) \neq (s, e^1, \ldots, e^d),$$

we can choose a polynomial $P$ s.t. $P(t, b^1, \ldots, b^d) \neq P(s, e^1, \ldots, e^d)$:

- if $s \neq t$: take the functional $P(t, b) = 1$,
- if $s = t$ and $b^i \neq e^i$ for some $i \in \{1, \ldots, d\}$: consider the polynomials

$$P(t, b^1, \ldots, b^d) = \int_{\Delta^k(0,t)} db^i,$$

with $I = (0, \ldots, 0, i)$ for $k \geq 0$. Then $P(t, b - e) = \int_0^t \rho_k d(b - e)^i$. Recall that a continuous path $x : [0, t] \to \mathbb{R}$ is of finite variation if and only if $x(.) = \int_0^t \mu(dr)$ for a signed measure $\mu$ on $[0, t]$ with finite mass and no atoms. The signed measure $\mu$ defines a linear functional on the space $C([0, t], \mathbb{R})$. Hence, if $\int_0^t \mu(dr) = 0$ for all polynomials $p : [0, t] \to \mathbb{R}$ this would imply $\int_0^t f(r) \mu(dr) = 0 \forall f \in C([0, t], \mathbb{R}),$ by the Stone-Weierstrass theorem on $C([0, T], \mathbb{R})$, therefore also $x(s) = \int_{[0, a]} \mu(dr) = 0$, $\forall s \in [0, t]$.

Applying this reasoning to $x = b^i - e^i$ shows that $P(t, b - e) = \int_0^t \rho_k d(b - e)^i = 0$ for every $k \geq 0$ leads to a contradiction to $b^i \neq e^i$.

\[\square\]

**Remark 7.** An important generalization of theorem\[\text{5}\] to rough path functionals has been obtained by Lyons et al. \[\text{11}\]

## 3. Approximating Functionals of a Diffusion

Theorem\[\text{6}\] applies to bounded variation paths and gives no description of the approximating functionals. We now follow a different approach using Dupire’s functional derivatives to show that under some additional smoothness assumption on $F$ one can also approximate functionals of semi-martingales trajectories and give an explicit description of the coefficients in the expansion. Let us fix notation: We are interested in functionals of the form

$$F : [0, T] \times D([0, T], \mathbb{R}^e) \to \mathbb{R},$$

$D([0, T], \mathbb{R}^e)$ denoting the space of càdlàg paths on $[0, T]$ with values in $\mathbb{R}^e$ and we write $\Lambda = [0, T] \times D([0, T], \mathbb{R}^e)$. Define a metric $\rho_\infty$ on $\Lambda$,

$$\rho_\infty((t, x), (s, y)) = |t - s| + |a_t x - a_s y|_{\infty;[0, T]}$$

($a_t$ denotes as in section\[\text{2}\] the stopping operator now applied to càdlàg paths) and set

$$C(\Lambda) = \{ F : \Lambda \to \mathbb{R}, \text{ progressive measureable wrt } \mathcal{F}, \text{ continuous wrt to } \rho_\infty \},$$

$\mathcal{F}$ being the $\sigma$-algebra generated by the coordinate process on $D([0, T], \mathbb{R}^e)$. Following Dupire \[\text{9}\], we say that $F \in C(\Lambda)$ has a time derivative $\partial_t F(t, x)$ at $(t, x) \in \Lambda$ if the limit

$$\partial_t F(t, x) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} (F(t + \epsilon, a_t x) - F(t, x))$$

exists and a space derivative $\partial_i F(t, x)$ in direction $e_i$, $i = 1, \ldots, e$, ($e_i$ denoting the standard basis on $\mathbb{R}^e$) at $(t, x) \in \Lambda$ if the limit

$$\partial_i F(t, x) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} (F(t, x + \epsilon e_i 1_{\geq t}) - F(t, x))$$


\[1\]T. Lyons personal communication (cf. thesis of T. Fawcett \[\text{11}\] and the transfer report of A. Janssen).
exists. Similarly, define the higher order derivatives $\partial_i \partial_j F$ etc. Denote $C^{m,n}(\Lambda)$ the subset of $C(\Lambda)$ of functionals which are at least $m$ times differentiable in time, at least $n$ times differentiable in space and continuous with respect to $\rho_\infty$. More formally, we define a set of multi-indices by letting

$$\mathcal{E}_{m,n} := \left\{ (\beta_1, \ldots, \beta_k) \in \bigcup_{k=0}^{\infty} \{0, \ldots, e\}^k : \text{card}(j : \beta_j = 0) \leq m \text{ and card}(j : \beta_j \neq 0) \leq n \right\}$$

(by convention $\{0, \ldots, e\}^0 = \emptyset$) and say $F \in C^{m,n}(\Lambda)$ if $\partial_{\alpha_1} \ldots \partial_{\alpha_n} F \in C(\Lambda)$ for all $(\alpha_1, \ldots, \alpha_n) \in \mathcal{E}_{m,n}$. Furthermore we define sets of bounded functions with bounded derivatives by

$$C^{m,n}_b(\Lambda) = \left\{ F \in C^{m,n}(\Lambda) : \sup_{(r,y) \in \Lambda} |\partial_{\alpha_1} \ldots \partial_{\alpha_n} F (r, y)| < \infty, (\alpha_1, \ldots, \alpha_n) \in \mathcal{E}_{m,n} \right\}.$$

The functional derivatives behave similar to the standard derivatives, we make repeatedly use of two properties.

**Lemma 8.** Let $m, n \in \mathbb{N}$ and suppose $F, G \in C^{m,n}_b(\Lambda)$.

1. Then $FG \in C^{m,n}_b(\Lambda)$ and $\partial_i (FG) = (\partial_i F) G + F (\partial_i G)$ for $i \in \{0, 1, \ldots, e\}$.
2. If $F(t, x) = f(t, x)$ for $f \in C^{1,1}(\mathbb{R} \times \mathbb{R}^e, \mathbb{R})$, then $F \in C^{1,1}(\Lambda)$ and $(\partial_i F)(t, x) = \frac{df}{dt}(t, x)$ and $(\partial_i G)(t, x) = \frac{df}{dx}(t, x)$ for $i \in \{0, 1, \ldots, e\}$

**Theorem 9** (Dupire’s functional Itô formula). Let $F \in C^{1,2}_b(\Lambda)$, $Y$ a continuous, $\mathbb{R}^e$-valued semi-martingale. Then $F(Y)$ is a continuous semi-martingale and a.s.

$$F_t(Y) - F_s(Y) = \int_s^t \partial_i F_r(Y) \, dr + \sum_{i=1}^e \int_s^t \partial_i F_r(Y) \, dY^i_r + \frac{1}{2} \sum_{i,j=1}^e \int_s^t \partial_{ij} F_r(Y) \, d[Y]^j_r.$$  

**Proof.** See [9] and [8].

**Remark 10.** Above formula also holds when the boundedness is replaced by boundedness on bounded sets and $F$ depends only continuous wrt to $Y$ and an extra variable e.g. the quadratic variation of $Y$, cf. [8].

We actually need the Stratonovich version with the Stratonovich integral as usual defined as

$$\int X \circ dY := \int X \, dY + \frac{1}{2} [X, Y].$$

Above formula then reads

**Corollary 11.** Let $F \in C^{1,3}_b(\Lambda)$, $Y$ a continuous, $\mathbb{R}^e$-valued semi-martingale. Then we have a.s.

$$F_t(Y) - F_s(Y) = \int_s^t \partial_i F_r(Y) \, dr + \sum_{i=1}^e \int_s^t \partial_i F_r(Y) \circ dY^i_r.$$  

**Proof.** We need to show that for $i \in \{1, \ldots, e\}$,

$$[\partial_i F, Y^i]_t = \sum_{j=1}^e \int_s^t \partial_j \partial_i F_r(Y) \, d[Y]^j_r.$$

By assumption $\partial_i F \in C^{1,2}_b$, hence from theorem [9]

$$d(\partial_i F_t) = \partial_t (\partial_i F_t) \, dt + \sum_{j=1}^e \partial_j \partial_i F_t \, dY^j_t + \frac{1}{2} \sum_{j,k=1}^e \partial_{jk} \partial_i F_t \, d[Y]^k_r,$$
and it follows that,
\[ [\partial_t F, Y^i]_t = \sum_{j=1}^c \left[ \int \partial_j \partial_t F(t, Y^j) \, dY^j, Y^i \right]_t \]
which equals \( \sum_{j=1}^c \int \partial_j \partial_t F(t, Y) \, d[Y]^{j,i}_t \).

3.1. Vector fields as derivations of functionals. Let \( M \) be a smooth manifold, then one can regard a vector field \( V \) on \( M \) as a derivation, \( V \in \text{Der}_\mathbb{R}(C^\infty(M)) \). We need to find a similar concept for the (infinite dimensional) space \( \Lambda \) which respects the additional structure (of being the product of a time coordinate and a pathspace). It seems natural to define for a given vector field \( V \in C_b^\infty(\mathbb{R}^c, \mathbb{R}^c) \),
\[ V \cdot F(t, y) = \sum_{i=1}^c V^i(y_i) \partial_i F(t, y) \text{ for } F \in C_b^\infty(\Lambda) . \]

However, above formulation does not take into account the time decay in functionals measured by the functional time derivative \( \partial_0 \). We therefore instead consider maps \( \nabla : \mathbb{R}^c \to \mathbb{R}^{c+1} \) with an additional 0th coordinate \( \nabla = (\nabla^j)_{j=0}^c \) and let \( \nabla \) act on \( C_b^\infty(\Lambda) \) as
\[ \nabla \cdot F(t, y) = \sum_{i=0}^c \nabla^i(y_i) \partial_i F(t, y) \text{ for } F \in C_b^\infty(\Lambda) . \]

**Lemma 12 (and Definition).** Given a map \( \nabla \in C_b^\infty(\mathbb{R}^c, \mathbb{R}^{c+1}) \), we can identify \( \nabla \) as an element of \( \text{Der}_\mathbb{R}(C_b^\infty(\Lambda)) \), the derivations on the \( \mathbb{R} \)-algebra \( C_b^\infty(\Lambda) \), by setting
\[ \nabla \cdot F(t, y) := \sum_{i=0}^c \nabla^i(y_i) \partial_i F(t, y) \text{ for } F \in C_b^\infty(\Lambda) . \]

**Proof.** By lemma \[ \nabla \] maps \( C_b^\infty(\Lambda) \) to \( C_b^\infty(\Lambda) \) functionals. One need to verify that \( C_b^\infty(\Lambda) \) is an \( \mathbb{R} \)-algebra and that \( \nabla \cdot (FG) = (\nabla \cdot F) \cdot G + F \cdot (\nabla \cdot G) \), but this follows again directly from lemma \[ \nabla \]. \( \square \)

**Proposition 13.** Let \( X = (X^i)_{i=0}^d \) be a continuous, semi-martingale with \( X^i_0 = t, V_i = (V_i^j)_{j=1}^c \in C_b^\infty(\mathbb{R}^c, \mathbb{R}^c) \), \( i \in \{0, 1, \ldots, d\} \) and \( Y = (Y^j)_{j=1}^d \) be the unique strong solution of the SDE
\[ dY_t = V(Y_t) \circ dX_t \]
\[ = V_0(Y_t) \, dt + \sum_{i=1}^d V_i(Y_t) \circ dX_t^i . \]

Define \( \nabla_i \in C_b^\infty(\mathbb{R}^c, \mathbb{R}^{c+1}) \) as
\[ \nabla_i = (\delta_{0i}, V_i^1, \ldots, V_i^c)^T \text{ for } i \in \{0, \ldots, d\} . \]

Then for every \( F \in C_b^\infty(\Lambda) \) we have a.s.
\[ F_t(Y) = F_s(Y) + \sum_{i=0}^d \int_s^t \nabla_i \cdot F_r(Y) \circ dX_r^i \]
\[ = F_s(Y) + \int_s^t \nabla \cdot F_r(Y) \circ dX_r . \]

\[ ^2 \delta_{i,j} \text{ the usual Kronecker delta, } \delta_{i,j} = 1 \text{ if } i = j, \text{ otherwise equal 0} \]
Proof. Applying corollary [11] to \( F \) gives
\[
F_t(Y) = F_s(Y) + \int_s^t \partial_0 F_r(Y) \, dr + \sum_{j=1}^c \int_s^t \partial_j F_r(Y) \circ dY^j_r
\]
\[
= F_s(Y) + \int_s^t \partial_0 F_r(Y) \, dr + \sum_{j=1}^c \sum_{i=0}^d \int_s^t \partial_j F_r(Y) V^j_i(Y) \circ dX^i_r
\]
\[
= F_s(Y) + \int_s^t \partial_0 F_r(Y) \, dr + \sum_{j=1}^c \int_s^t \partial_j F_r(Y) V^j_0(Y) \circ dr
\]
\[
+ \sum_{j=1}^c \sum_{i=1}^d \int_s^t \partial_j F_r(Y) V^j_i(Y) \circ dX^i_r
\]
Identifying \( V_0 \) and \( V_1, \ldots, V_d \) as a derivation, as in lemma [12] (recall that \( V^0_i = 0 \) for \( i \neq 0 \)), this becomes
\[
F_t(Y) = F_s(Y) + \int_s^t V_0 \cdot F_r(Y) \circ dX^0_r + \sum_{i=1}^d \int_s^t V_i \cdot F_r(Y) \circ dX^i_r
\]
\[\square\]

To identify compositions of the vector fields we define the set of multi-indices \( A \) by
\[
A := \bigcup_{k=1}^\infty \{0, \ldots, d\}^k
\]
and let \( I = (\alpha_0, \ldots, \alpha_k) \in A \) be a multi-index. Given \( I \in A \) set \(|I| = \text{card}(I)\) and also define \( \|I\| := |I| + \text{card}(j : \alpha_j = 0) \). Finally let
\[
A(j) = \{I \in A : \|I\| \leq j\}.
\]

Given a multi-index \( I = (\alpha_0, \ldots, \alpha_k) \in A \) we define \( V_I := V_{\alpha_0} \cdots V_{\alpha_k} \).

**Definition 14.** Set
\[
\Delta_k(s,t) = \{(t_1, \ldots, t_k) \in [s,t]^k, s \leq t_1 \leq \cdots \leq t_k \leq t\}
\]
Let \( X = (X^i)_{i=0}^d \) be a continuous semi-martingale with \( X^0_t = t \). Define the iterated Stratonovich integrals of \( X \) as
\[
\int_{\Delta_{k}(s,t)} \circ dX^I = \int_{\Delta_{k}(s,t)} \circ dX^\alpha_1 \cdots \circ dX^\alpha_k \text{ with } I = (\alpha_0, \ldots, \alpha_k) \in A.
\]

**3.2. The functional Taylor series and a remainder estimate.** We can now formulate our main theorem. The reader will notice that applied to a functional of the form \( F(t,x) = f(t,x_t) \), one recovers the usual stochastic Taylor expansion, [14] [15].

**Theorem 15.** Let \( X = (X^i)_{i=0}^d \) be a continuous, semi-martingale with \( X^0_t = t \), \( V = (V_i)_{i=1}^d \) with \( V_i \in C_b^\infty(\mathbb{R}^e, \mathbb{R}^e) \) and let \( Y \) be the unique strong solution of the SDE
\[
dY_t = V(Y_t) \circ dX_t \equiv \sum_{i=0}^d V_i(Y_t) \circ dX^i_t.
\]
Define \( V_i \in C_b^\infty(\mathbb{R}^e, \mathbb{R}^{e+1}) \) as
\[
V_i = (\delta_{0i}, V^1_i, \ldots, V^d_i)^T \text{ for } i \in \{0, \ldots, d\}.
\]
Fix $m \in \mathbb{N}$. Then for every $F \in C^\infty_b(\Lambda)$ we have

$$F_t(Y) - F_s(Y) = \sum_{I \in \mathcal{A}(m)} \nabla_I \cdot F_s(Y) \int_{\Delta_I(s,t)} \circ dX^I + R_{st}^m,$$

where

$$R_{st}^m = \sum_{(i_1, \ldots, i_N) \notin \mathcal{A}(m)} \int_{s<r_1<\ldots<r_N<t} \nabla_{i_1} \cdots \nabla_{i_N} \cdot F_r(Y) \circ dX_{r_1}^{i_1} \cdots \circ dX_{r_N}^{i_N}.$$

Proof. Applying proposition 13 gives

$$F_t(Y) = F_s(Y) + \sum_{i=0}^d \nabla_i \cdot F_r(Y) \circ dX^i.$$

Now by assumption on $F$ and $V$, the functional $\nabla_i \cdot F$ is again in $C^\infty_b$ and applying proposition 13 to $\nabla_i \cdot F$ gives

$$F_t(Y) = F_s(Y) + \sum_{i=0}^d \nabla_i \cdot F_s(Y) \int_s^t \circ dX^i + \sum_{i_1, i_2=0}^d \int_{s<r_2<r_1<t} \nabla_{i_2} \cdot \nabla_{i_1} \cdot F_{r_2}(Y) \circ dX_{r_2}^{i_2} \circ dX_{r_1}^{i_1}.$$

Iterating this procedure, we finally arrive at

$$F_t(Y) = F_s(Y) + \sum_{I \in \mathcal{A}(m)} \nabla_I \cdot F_s(Y) \int_{\Delta_I(s,t)} \circ dX^I + \sum_{(i_1, \ldots, i_N) \notin \mathcal{A}(m)} \int_{s<r_1<\ldots<r_N<t} \nabla_{i_1} \cdots \nabla_{i_N} \cdot F_{r_1}(Y) \circ dX_{r_1}^{i_1} \cdots \circ dX_{r_N}^{i_N}. \tag*{\square}$$

Remark 16. The choice of truncation in the Taylor approximation in Theorem 15 reflects how the semi-martingale part of the noise $X$ scales compared to the time component $X^0 = t$. If $X$ is a Brownian motion, then (cf. [15]) for $I = (\alpha_1, \ldots, \alpha_k)$

$$\int_{\Delta_I(s,t)} \circ dX^I$$

is equal in law to $\sqrt{t} \cdot \int_{\Delta_{\{0,1\}}} \circ dX_I$. This explains the particular truncation of the Taylor expansion via the multi-index sets $\mathcal{A}(m)$ and allows to arrive at the $L^2$ error estimate in corollary 15 which is important in numerical applications, e.g. [15].

Remark 17. Even if there is no drift vector field in the SDE, that is in above notation $V_0 = (0, \ldots, 0)^T$, we still have to deal with iterated integrals which involve $dX^0 = dt$. This is in contrast to the stochastic Taylor expansion for functions $f : \mathbb{R}^c \to \mathbb{R}^c$ due to the time decay in the functional $F \in C^\infty_b(\Lambda)$ which is picked up by the functional derivative $\partial_0$ via the first coordinate of $V_0 = (1, 0, \ldots, 0)^T$.

Of special importance is the case of Brownian noise in which case we can give a $L^2$ estimate for the remainder term.
Corollary 18. Let \((B^i)_{i=1}^d\) be a \(d\)-dimensional Brownian motion and set \(B = (B^i)_{i=0}^d\) with \(B^0_t = t\). Let \(V = (V_i)_{i=1}^d\) with \(V_i \in C_b^\infty(\mathbb{R}^e, \mathbb{R}^e)\) and \(Y\) be the unique strong solution of the Stratonovich SDE with drift,

\[
dY_t = V(Y_t) \circ dB_t = \sum_{i=0}^d V_i(Y_t) \circ dB^i_t.
\]

Define \(\mathbf{V}_i \in C_b^\infty(\mathbb{R}^e, \mathbb{R}^{e+1})\) as \(\mathbf{V}_i = (\delta_{0i}, V^1_i, \ldots, V^e_i)^T\) for \(i \in \{0, \ldots, d\}\). Fix \(m \in \mathbb{N}\). Then for every \(F \in C_b^{\infty, \infty}(\mathcal{A})\) we have

\[
F_t(Y) - F_s(Y) = \sum_{t \in \mathcal{A}(m)} \mathbf{V}_i \cdot F_s(Y) \int_{\Delta_t(s, t)} \circ dB^i + R^m_{st}.
\]

Moreover, there exists a constant \(c = c(d, m)\), only depending on \(d\) and \(m\), such that

\[
\mathbb{E} \left[ |R^m_{st}|^2 \right]^{1/2} \leq c \sum_{j=m+1}^{m+2} \sup_{j \in \mathcal{A}(j) \setminus \mathcal{A}(j-1)} \sup_{(r, y) \in \mathcal{A}} |\mathbf{V}_i \cdot F(r, y)|.
\]

Proof. Given \(G \in C_b^{1,2}(\mathcal{A})\), \(k \in \mathbb{N}\) and \(I = (\alpha_1, \ldots, \alpha_k) \in \mathcal{A}\). We show that there exists a constant \(C = C(d, k)\), only depending on \(d\) and \(k\), s.t. \(\forall t > 0\)

\[
\mathbb{E} \left( \int_{0 < t_1 < \ldots < t_k < t} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_k}_{t_k} \right)^2 \leq C \left( \|G\|_2^2 + (1 - \delta_{0, \alpha_1}) \|G\|_\infty \|V_{\alpha_1}G\|_\infty^2 \right).
\]

The result then follows from Theorem 15 by applying this to \(G = \mathbf{V}_{i_1} \cdots \mathbf{V}_{i_N} \cdot F\) (without loss of generality one can take \(s = 0\)). We assume for induction that for any \(j \leq k\),

\[
\mathbb{E} \left( \int_{0 < t_1 < \ldots < t_j < t} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_j}_{t_j} \right)^2 \leq C(j, d) \left( \|G\|_2^2 + (1 - \delta_{0, \alpha_1}) \|G\|_\infty \|V_{\alpha_1}G\|_\infty^2 \right).
\]

for all \(t > 0\). We will prove the claim for \(I = (\alpha_1, \ldots, \alpha_{k+1})\). First assume that \(\alpha_{k+1} = 0\). In this case we have

\[
\mathbb{E} \left( \int_0^t \int_{0 < t_1 < \ldots < t_k < t_{k+1}} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_k}_{t_k} dt_{k+1} \right)^2
\]

\[
= \mathbb{E} \left( \int_0^t \int_{0 < t_1 < \ldots < t_k < t_{k+1}} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_k}_{t_k} dt_{k+1} \right)^2
\]

\[
\leq t^2 \mathbb{E} \left[ \int_0^1 \left( \int_{0 < t_1 < \ldots < t_k < t_{k+1}} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_k}_{t_k} \right)^2 dt_{k+1} \right]
\]

\[
= t^2 \int_0^1 \mathbb{E} \left( \int_{0 < t_1 < \ldots < t_k < t_{k+1}} G(t_1, Y) \circ dB^{\alpha_1}_{t_1} \ldots \circ dB^{\alpha_k}_{t_k} \right)^2 dt_{k+1}
\]
by using the Jensen inequality. Assuming the inductive hypothesis we estimate that last line as
\[
\leq C(k,d) t^2 \int_0^1 \left( (ts)^{\|\alpha_1,\ldots,\alpha_k\|} \|G\|_\infty^2 + (1 - \delta_{0,\alpha_1}) (ts)^{\|\alpha_1,\ldots,\alpha_k\|+1} \|\bar{V}_{\alpha_1} G\|_\infty^2 \right) ds
\leq C(k + 1, d) \left( t^{\|\alpha_1,\ldots,\alpha_{k+1}\|} \|G\|_\infty^2 + (1 - \delta_{0,\alpha_1}) t^{\|\alpha_1,\ldots,\alpha_{k+1}\|+1} \|\bar{V}_{\alpha_1} G\|_\infty^2 \right).
\]

Now suppose \(\alpha_{k+1} \neq 0\). We have if \(\alpha_k \neq 0\) (it easy to see that \(\alpha_k = 0\) follows from a similar calculation)
\[
\mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k+1} < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k+1}}^{\alpha_{k+1}} \right)^2
= \mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k+1} < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k+1}}^{\alpha_{k+1}} + \frac{1}{2} \left[ \int_{0 < t_1 < \ldots < t_k < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_k}^{\alpha_k} \right] \right)^2
\leq 2 \mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k+1} < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k+1}}^{\alpha_{k+1}} \right)^2
+ 2 \delta_{\alpha_k, \alpha_{k+1}} \mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k-1} < s < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k-1}}^{\alpha_{k-1}} ds \right)^2
\]
Note that we can bound the second term in the sum using the same arguments as in the case \(\alpha_{k+1} = 0\). For the first term we estimate the Ito integral,
\[
\mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k+1} < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k+1}}^{\alpha_{k+1}} \right)^2
= \int_0^t \mathbb{E} \left( \int_{0 < t_1 < \ldots < t_{k+1} < t} G(t_1, Y) \circ dB_{t_1}^{\alpha_1} \ldots \circ dB_{t_{k+1}}^{\alpha_{k+1}} \right)^2 dt_{k+1}
\leq C(k, d) \int_0^t \left( t^{\|\alpha_1,\ldots,\alpha_k\|} \|G\|_\infty^2 + (1 - \delta_{0,\alpha_1}) t^{\|\alpha_1,\ldots,\alpha_k\|+1} \|\bar{V}_{\alpha_1} G\|_\infty^2 \right) dt_{k+1}
\leq C(k + 1, d) \left( t^{\|\alpha_1,\ldots,\alpha_{k+1}\|+1} \|G\|_\infty^2 + (1 - \delta_{0,\alpha_1}) t^{\|\alpha_1,\ldots,\alpha_{k+1}\|+2} \|\bar{V}_{\alpha_1} G\|_\infty^2 \right)
\]
using the inductive hypothesis. Finally we prove the base case of the induction: Recall that \(Y\) is \(\mathbb{R}^c\)-valued and \(\bar{V}_t = (0, V_{1t}^1, \ldots, V_{ct}^c)\). Similar to proposition \(\mathbb{L}\) we see
\[
G_t (Y) - G_0 (Y) = \sum_{j=1}^d \int_0^t \nabla_j \cdot G(t, Y) dB^j_t + BV_t
\]
with \(BV\) a continuous bounded variation process. Thus for \(\alpha_1 \neq 0\) we have
\[
d[G(\cdot, Y), B^{\alpha_1}]_t = \sum_{i=1}^d \nabla_i \cdot G(t, Y) dB^i_t + \nabla_{\alpha_1} \cdot G(t, Y) dt.
\]
Hence,
\[
\int_0^t G(s, Y) \circ dB_s^{\alpha_1} = \int_0^t G(r, Y) dB_r^{\alpha_1} + \frac{1}{2} [G(\cdot, Y), B]_t
\]
\[
= \int_0^t G(r, Y) dB_r^{\alpha_1} + \frac{1}{2} \int_0^t \nabla_{\alpha_1} G(r, Y) dr
\]
which implies the base case of the induction for \( \alpha_1 \neq 0 \). The estimate for the case \( \alpha_1 = 0 \), i.e. \( B_t^{\alpha_1} = t \) is obvious. \( \square \)

To demonstrate that above theorem leads to concrete, easy to use, estimates we conclude with a toy example which can be verified (by a longer calculation) with standard Ito calculus.

**Example 19.** Take a 1-dimensional Brownian motion \( B, f, g, V_1 \in C_b^\infty (\mathbb{R}, \mathbb{R}) \) and let \( dY = V_1 (Y_t) \circ dB \). Consider
\[
F(t, x) = f \left( \int_0^t g(x_r) \, dr \right).
\]
We immediately get \( \partial_0 F(t, x) = g(x_t) f' \left( \int_0^t g(x_r) \, dr \right) \), \( \partial_1 F(t, x) = 0 \), \( \partial_0 \partial_0 F(t, x) = g'(x_t) f' \left( \int_0^t g(x_r) \, dr \right) \) and similarly one can calculate higher derivatives. The first nontrivial case of corollary [18] applied to this example is the step \( m = 3 \) approximation,
\[
F_3(Y) - F_s(Y) = \partial_0 F_s(Y) \int_{s<r<t} \circ dr + V_1 (Y_s) \partial_1 \partial_0 F_s (Y) \int_{s<r_1<r_2<t} \circ dB_{r_1} \circ dr_2 + R_{3t}^3
\]
with
\[
\mathbb{E} \left[ \left| R_{3t}^3 \right|^2 \right]^{1/2} \leq c \sup_{I \in \{(1,1,0),(1,0,0),(0,1,0),(0,0,0)\}} \left( |t-s|^{1/2} \left| \nabla^I \cdot F(r, y) \right| \right).
\]

**Acknowledgment.** Christian Litterer and Harald Oberhauser were partially supported by the European Unions Seventh Framework Programme, ERC grant agreement 258237.

**References**

[1] Robert Azencott. Formule de Taylor stochastique et développement asymptotique d’intégrales de Feynman. In *Seminar on Probability, XVI, Supplement*, pages 237–285. Springer, Berlin, 1982.

[2] Gérard Ben Arous. Flots et séries de Taylor stochastiques. *Probab. Theory Related Fields*, 81(1):29–77, 1989.

[3] Roger W. Brockett. Volterra series and geometric control theory. *Automatica*, 12(2):167 – 176, 1976.

[4] Fabienne Castell. Asymptotic expansion of stochastic flows. *Probab. Theory Related Fields*, 96(2):225–239, 1993.

[5] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math.* (2), 65:163–178, 1957.

[6] Kuo-Tsai Chen. Integration of paths—a faithful representation of paths by non-commutative formal power series. *Trans. Amer. Math. Soc.*, 89:395–407, 1958.

[7] R. Cont and D. Fournie. Change of variable formulas for non-anticipative functionals on path space. *Journal of Functional Analysis*, 259:1043–1072, 2010.

[8] R. Cont and D. Fournie. Functional Ito calculus and stochastic integral representation of martingales. *Annals of Probability*, 2011. In press.

[9] B. Dupire. Functional Ito calculus. 2010. SSRN.

[10] I. Ekren, C. Keller, N. Touzi, and J. Zhang. On Viscosity Solutions of Path Dependent PDEs. *ArXiv e-prints*, September 2011.

[11] Thomas Fawcett. *Problems in stochastic analysis: connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2003.

[12] M. Fleiss. Fonctionnelles causales non linéaires et indéterminées non commutatives. *Bull. Soc. Math. France*, 109(1):3–40, 1981.

[13] Yao Zhong Hu. Série de Taylor stochastique et formule de Campbell-Hausdorff, d’après Ben Arous. In *Séminaire de Probabilités, XXVI*, volume 1526 of *Lecture Notes in Math.*, pages 579–586. Springer, Berlin, 1992.

[14] Peter Kloeden and Eckhard Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin, 1992.
[15] Terry Lyons and Nicolas Victoir. Cubature on Wiener space. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 460(2041):169–198, 2004.
[16] S. Peng. Note on viscosity solution of path-dependent pde and g-martingales. ArXiv e-prints, 2011.
[17] S. Peng and F. Wang. BSDE, Path-dependent PDE and Nonlinear Feynman-Kac Formula. ArXiv e-prints, August 2011.
[18] Robert S. Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. J. Funct. Anal., 72(2):320–345, 1987.
[19] D. W. Stroock. Homogeneous chaos revisited. Seminaire de probabilites (Strasbourg), 21:1–7, 1987.
[20] H.J. Sussmann. Semigroup representations, bilinear approximation of input-output maps, and generalized inputs. Mathematical systems theory, 131, 1975.

TU Berlin, Fakultät II, Institut für Mathematik, MA 7-2, Strasse des 17. Juni 136, 10623 Berlin.

E-mail address: h.oberhauser@gmail.com