LIMITS OF THE WONG-ZAKAI TYPE WITH A MODIFIED DRIFT TERM

HÉCTOR J. SUSSMANN*

Abstract. We study Stratonovich stochastic differential equations driven by an m-dimensional Wiener process W, with $m \geq 2$. If W is approximated by processes W^ν with more regular sample paths, then it is known that the solutions of the equations driven by the W^ν will converge to the solution of the equation driven by W, provided that the approximations satisfy the conditions of the Wong-Zakai theorem. McShane gave an example showing that, if those conditions are not satisfied, then a different limiting equation can arise. Here we describe a large class of equations, obtained from the original one by suitably modifying the drift term, that can arise as limiting equations by some choice of the sequence $\{W^\nu\}$.

1. Introduction. Consider a stochastic differential equation

$$dx = f_0(x)dt + \sum_{i=1}^{m} f_i(x)dW_i \ , \tag{1.1}$$

where x is n-dimensional, $W = (W_1, \ldots, W_m)$ is a standard m-dimensional Brownian motion, the vector fields f_i satisfy appropriate smoothness and growth conditions, and the solutions are always understood to be in the Stratonovich sense.[5]

If we approximate W by a sequence of processes W^ν with more regular (e.g. Lipschitz) sample paths, then the well known Wong-Zakai Approximation Theorem (cf. [9], [10]) says that the solutions $t \to X^\nu(t)$ of the corresponding approximating equation

$$dx = f_0(x)dt + \sum_{i=1}^{m} f_i(x)dW_i^\nu \ , \tag{1.2}$$

with some given initial condition $X^\nu(0) = \bar{X}$, converge to the solution X of (1.1) with the same initial condition, *provided that the approximations W^ν satisfy some extra conditions*, which always hold if $m = 1$, but may fail if $m > 1$. McShane gave an example in [4] showing that, for $m = 2$, the X^ν may indeed fail to converge to X. In this note we investigate the possible limits that can be obtained by taking more general sequences of approximating processes and show that, by a suitable choice of the approximation, it is possible to make the X^ν converge to the solution of an equation

$$dx = (f_0(x) + g(x))dt + \sum_{i=1}^{m} f_i(x)dW_i \ , \tag{1.3}$$

*Mathematics Department, Rutgers University, New Brunswick, New Jersey 08903.Work supported in part by the National Science Foundation under NSF Grant DMS-8902994.

with a different drift term. We will show that g can be chosen to be an arbitrary element of Λ, where Λ is the linear span of all the Lie brackets of the f_i for $i = 1, \ldots, m$ that contain at least two factors. The precise statement is given below in Theorem 6.1. Our result is a generalization of Theorem 7.2 of [3], Chapter 6, where it is shown that, by a suitable choice of the approximating sequence, one can produce a drift term which is an arbitrary linear combination of brackets $[f_i, f_j]$, $i, j > 0$.

We remark that we only prove almost sure convergence for a fixed T and a fixed initial condition. With a more careful analysis, one can prove a.s. convergence uniformly in t for t in any bounded interval, and convergence of the stochastic flows.

2. Differential Equations with Inputs. We let $C_b^\infty(\mathbb{R}^n, \mathbb{R}^n)$ denote the class of all maps $f : \mathbb{R}^n \to \mathbb{R}^n$ of class C^∞ such that all the partial derivatives $\frac{\partial^{\alpha_1+\cdots+\alpha_n} f}{\partial x_1^{\alpha^1} \ldots \partial x_n^{\alpha^n}}$ of all orders (*including* order 0) are bounded on $\mathbb{R}^n$. (In particular, every $f \in C_b^\infty(\mathbb{R}^n, \mathbb{R}^n)$ is globally bounded and globally Lipschitz.)

We assume that $f_0, \ldots, f_m \in C_b^\infty(\mathbb{R}^n, \mathbb{R}^n)$. We let $\mathcal{U}^m$ denote the space of all locally absolutely continuous functions $U : [0, \infty) \to \mathbb{R}^m$ such that $U(0) = 0$. If $U \in \mathcal{U}^m$, then we write $u = \dot{U} = \frac{dU}{dt}$, so $u \in L^1_{\rm loc}([0, \infty), \mathbb{R}^m)$.

Let $U \in \mathcal{U}^m$, and write $u = \dot{U}$. Write $U_0(t) \equiv t$, i.e. $u_0(t) \equiv 1$. Then the ordinary differential equation

$$\dot{x} = f_0(x) + \sum_{i=1}^{m} u_i(t) f_i(x) \ , \tag{2.1}$$

can also be written in the form

$$dx = f_0(x)\, dt + \sum_{i=1}^{m} f_i(x) dU_i$$

or

$$dx = \sum_{i=0}^{m} f_i(x) dU_i$$

It is clear that (2.1) satisfies the conditions of the Carathéodory existence and uniqueness theorem. Moreover, since the f_i are bounded, trajectories do not escape in finite time. So, given any $a \in [0, \infty)$, $\bar{x} \in \mathbb{R}^n$, there exists a unique solution $t \to x(t)$ of (2.1) such that $x(a) = \bar{x}$. For fixed $b \in [0, \infty)$, we will use $\Phi^U_{b,a}$ to denote the map that assigns to each $\bar{x}$ the value $x(b)$ of the corresponding solution. That is, $t \to \Phi^U_{t,a}(x)$ is the solution of (2.1) that goes through x when $t = a$. Each map $\Phi^U_{b,a}$ is a C^∞ diffeomorphism from $\mathbb{R}^n$ onto $\mathbb{R}^n$. Moreover, these diffeomorphisms satisfy $\Phi^U_{a,a} = $ identity, and $\Phi^U_{c,b}\Phi^U_{b,a} = \Phi^U_{c,a}$ for all $a, b, c \in [0, \infty)$.

We now define the *iterated integrals* $\int_a^b u_I$, where $I = (i_1, \ldots, i_r)$ is an arbitrary member of $\mathcal{I}(m)$, the space of all finite sequences of indices $i \in \{0, \ldots, m\}$. (We will write $|I|$ for the *length* of I, i.e. the number r. When $|I| = 1$, so $I = (i)$ for some i, we will just write $\int_a^b u_i$ instead of $\int_a^b u_{(i)}$. The empty sequence $\emptyset$ is a member of $\mathcal{I}(m)$.) The definition is recursive: we let $\int_a^b u_I = 1$ if $I = \emptyset$, and for a general I we write $I = (i, I')$, and define $\int_a^b u_I$ to be equal to $\int_a^b u_i(t)(\int_a^t u_{I'})dt$.

We will also write $U_I(b, a)$ for $\int_a^b u_I$. Notice that the identity $U_I(b, a) = U_i(b) - U_i(a)$ holds when $|I| = 1$, $I = (i)$, but no similar formula is true when $|I| \neq 1$, since in that case the additivity property $U_I(c, a) = U_I(c, b) + U_I(b, a)$ does not hold in general.

We are interested in the derivatives of $U_I(t, s)$ with respect to both variables t and s. Define

$$u_I^{+,s}(t) = \frac{\partial}{\partial t} U_I(t, s)$$

and

$$u_I^{-,t}(s) = -\frac{\partial}{\partial s} U_I(t, s)\,.$$

Then $\int_a^b u_I = \int_a^b u_I^{+,a}(t)\,dt = \int_a^b u_I^{-,b}(s)\,ds$. The functions $u_I^{+,a}(t)$, $u_I^{-,b}(s)$ are equal, respectively, to

$$u_{i_1}(t) \int_a^t \int_a^{t_1} \cdots \int_a^{t_{k-2}} u_{i_2}(t_1) \ldots u_{i_k}(t_{k-1})\, dt_{k-1} \ldots dt_1 \tag{2.2}$$

and

$$u_{i_k}(s) \int_s^b \int_{t_{k-1}}^b \cdots \int_{t_2}^b u_{i_1}(t_1) \ldots u_{i_{k-1}}(t_{k-1})\, dt_{k-1} \ldots dt_1\ . \tag{2.3}$$

if $I = (i_1, \ldots, i_k) \in \mathcal{I}(m)$.

Now suppose that φ is a scalar- or vector-valued function of class C^∞. Write $f_i\varphi$ to denote the result of applying f_i to φ as a first-order differential operator, i.e. $(f_i\varphi)(x) = \lim_{h\to 0} \frac{1}{h}(\varphi(x + hf_i(x)) - \varphi(x))$. More generally, if $I = (i_1, \ldots, i_r) \in \mathcal{I}(m)$, we write $f_I = f_{i_1} f_{i_2} \ldots f_{i_r}$. Then (2.1) implies the equation

$$\varphi(\Phi_{t,a}^U(x)) = \varphi(x) + \sum_{i=0}^m \int_a^t u_i(s)(f_i\varphi)(\Phi_{s,a}^U(x))\, ds\ , \tag{2.4}$$

which is the $k = 0$ case of the general formula

$$\varphi(\Phi_{t,a}^U(x)) = \sum_{|I|\le k} U_I(t, a)(f_{I\#}\varphi)(x) + R_{k,t,a,U,\varphi,f}(x) \tag{2.5}$$

where, for any multiindex $I = (i_1, \ldots, i_r)$, we use $I^\#$ to denote the reversed multiindex, i.e. $I^\# = (i_r, \ldots, i_1)$, and the remainder $R_{k,t,a,U,\varphi,f}(x)$ is given by

$$R_{k,t,a,U,\varphi,f}(x) = \sum_{|I|=k+1} \int_a^t u_I^{-,t}(s)(f_{I^\#}\varphi)(\Phi^U_{s,a}(x))\, ds \ . \tag{2.6}$$

It is easy to see that (2.5) is actually true for all k. (The proof is by induction, using repeated integrations by parts.) A particularly important choice of φ is $\varphi = E^n$, where $E^n : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map. In that case, (2.5) becomes

$$\Phi^U_{t,a}(x) = \sum_{|I|\le k} U_I(t,a) E^f_I(x) + R_{k,t,a,U,E^n,f}(x) \tag{2.7}$$

where $E^f_I = f_{I^\#} E^n$, and

$$R_{k,t,a,U,E^n,f}(x) = \sum_{|I|=k+1} \int_a^t u_I^{-,t}(s) E^f_I(\Phi^U_{s,a}(x))\, ds \ . \tag{2.8}$$

We remark that all the vector-valued functions E^f_I belong to $C^\infty_b(\mathbb{R}^n, \mathbb{R}^n)$.

3. Stochastic Ordinary Inputs. Now assume that $(\Omega, \mathcal{F}, P)$ is a probability space, and $\{\mathcal{F}_t\}$ is an increasing family of sub-σ-fields of $\mathcal{F}$. An m-dimensional *ordinary input process* (OIP) on $(\Omega, \mathcal{F}, P)$ is a stochastic process $U = \{U(t) : t \ge 0\}$ such that all the sample paths $t \to U(t)(\omega)$, $\omega \in \Omega$, belong to $\mathcal{U}^m$. In that case, the derivative $\dot{U}$, the iterated integrals, the solutions of (2.1), and all the other U-dependent objects introduced above are well defined for each $\omega \in \Omega$.

As usual, we call a process U *adapted* if $U(t)$ is $\mathcal{F}_t$-measurable for each t. However, it is also useful to define a weaker concept, namely, that of a π*-adapted* process, where π is a partition of $[0, \infty)$. Precisely, we define a *partition* of $[0, \infty)$ to be an infinite sequence $\pi = \{t_j\}_{j=0}^\infty$ such that $0 = t_0 < t_1 < t_2 < \ldots$ and $\lim_{j\to\infty} t_j = \infty$. The *mesh* $|\pi|$ of a partition π is the number $\sup\{t_j - t_{j-1} : j = 1, 2, \ldots\}$. If π is a partition, then we call U π*-adapted* to $\{\mathcal{F}_t\}$ if, for every j, $U(t)$ is $\mathcal{F}_{t_j}$-measurable whenever $t \le t_j$.

Now suppose that $\pi = \{t_j\}_{j=0}^\infty$ is a partition of $[0, \infty)$ and U is a π-adapted m-dimensional OIP. Let $k > 0$ be an integer, and let $C \in \mathbb{R}$, $C > 0$. We will say that U belongs to the class $OIP(m, k, C, \pi)$ if U satisfies the following three bounds

$$|\mathbb{E}(U_I(t_j, t_{j-1}))/\mathcal{F}_{t_{j-1}}| \le C(t_j - t_{j-1}) \ , \tag{3.1}$$

$$\mathbb{E}(U_I(t_j, t_{j-1})^2/\mathcal{F}_{t_{j-1}}) \le C(t_j - t_{j-1}) \ , \tag{3.2}$$

and

$$|\mathbb{E}(U_I(t,t_{j-1})^2)/\mathcal{F}_{t_{j-1}}| \leq C \ , \tag{3.3}$$

for all choices of $I \in \mathcal{I}(m)$ such that $|I| \leq k$, all $j \in \{1,2,3,\ldots\}$, and all $t \in [t_{j-1}, t_j]$, as well as the bound

$$|u_I^{-,t}(s)| \leq C \tag{3.4}$$

for all $I \in \mathcal{I}(m)$ such that $|I| = k+1$ and all s, t such that $t_{j-1} \leq s \leq t \leq t_j$ for some j.

If $U \in OIP(m,k,C,\pi)$ and $1 \leq |I|, |J| \leq k$, then it follows from the Schwartz inequality for conditional expectations that

$$|\mathbb{E}(U_I(t_j,t_{j-1})U_J(t_j,t_{j-1}))/\mathcal{F}_{t_{j-1}}| \leq C(t_j - t_{j-1}) \ . \tag{3.5}$$

We now define $D_I(f) = \sup\{||x-y||^{-1}||E_I^f(x) - E_I^f(y)|| : x, y \in \mathbb{R}^n \ , x \neq y\}$, and let $D = D(k,f) = \max\{D_I(f) : 1 \leq |I| \leq k+1\}$.

LEMMA 3.1. *For every k, m, f, C there exist constants K, μ, depending only on k, m, D and C, but not on the particular choice of U, X, Y, f or π, with the property that, whenever $\pi = \{t_j\}_{j=0}^{\infty}$ is a partition of $[0,\infty)$, and U is a process in $OIP(m,k,C,\pi)$, then the bound*

$$||\Phi^U_{t_j,t_{j-1}}(X) - \Phi^U_{t_j,t_{j-1}}(Y)||_{L_2} \leq (1 + K\mathrm{e}^{\mu|\pi|}(t_j - t_{j-1}))||X - Y||_{L^2}$$

holds whenever $X, Y : \Omega \to \mathbb{R}^n$ are $\mathcal{F}_{t_{j-1}}$-measurable and square-integrable.

PROOF. Throughout this proof, we will use the notation $\mathcal{E}(X,Y)$ to denote $\mathcal{E}(X) - \mathcal{E}(Y)$, whenever $\mathcal{E}$ is some expression that depends on X. Write $a = t_{j-1}$, $b = t_j$, and use $\mathbb{E}_a$ to denote conditional expectation with respect to $\mathcal{F}_a$. Let $a \leq t \leq b$. Using (2.7), we get

$$\begin{aligned} \Phi^U_{t,a}(X,Y) = X - Y + &\sum_{1\leq|I|\leq k} U_I(t,a)E_I^f(X,Y) \\ &+ R_{k,t,a,U,E^n,f}(X,Y) \, . \end{aligned} \tag{3.6}$$

Using the bound $||E_I^f(x,y)|| \leq D||x-y||$, we get

$$\begin{aligned} \mathbb{E}(U_I(t,a)^2||E_I^f(X,Y)||^2) &= \mathbb{E}(||E_I^f(X,Y)||^2 \cdot \mathbb{E}_a(U_I(t,a)^2) \\ &\leq C^2D^2\mathbb{E}(||X-Y||^2) \, , \end{aligned}$$

so that

$$||U_I(t,a) \cdot E_I^f(X,Y)||_{L^2} \leq CD||X-Y||_{L^2} \tag{3.7}$$

if $1 \leq |I| \leq k$. Similarly, if $|I| = k+1$, we have

$$||u_I^{-,t}(s) \cdot E_I^f(\Phi^U_{s,a}(X,Y))|| \leq CD||\Phi^U_{s,a}(X,Y)|| \tag{3.8}$$

pointwise, so

$$||u_I^{-,t}(s) \cdot E_I^f(\Phi^U_{s,a}(X,Y))||_{L^2} \leq CD||\Phi^U_{s,a}(X,Y)||_{L^2}\ . \tag{3.9}$$

Since $R_{k,t,a,U,E^n,f}(X,Y) = \sum_{|I|=k+1} \int_a^t u_I^{-,t}(s) \cdot E_I^f(\Phi^U_{s,a}(X,Y))\, ds$, we find

$$||R_{k,t,a,U,E^n,f}(X,Y)||_{L^2} \leq \mu \int_a^t ||\Phi^U_{s,a}(X,Y)||\, ds\ , \tag{3.10}$$

where $\mu = (m+1)^{k+1}CD$. Combining (3.6), (3.7) and (3.10), we get

$$||\Phi^U_{t,a}(X,Y)||_{L^2} \leq (1+\nu CD)||X-Y||_{L^2} + \mu \int_a^t ||\Phi^U_{s,a}(X,Y)||_{L^2}\, ds\ , \tag{3.11}$$

where $\nu = m+1+(m+1)^2 + \ldots + (m+1)^k = \frac{(m+1)^{k+1}-m-1}{m}$.

Gronwall's inequality then yields

$$||\Phi^U_{t,a}(X,Y)||_{L^2} \leq (1+\nu CD)\mathrm{e}^{\mu|\pi|}||X-Y||_{L^2}\ . \tag{3.12}$$

If we now use (3.10) again, with $t=b$, together with (3.12), we get

$$||R_{k,b,a,U,E^n,f}(X,Y)||_{L^2} \leq \mu(1+\nu CD)\mathrm{e}^{\mu|\pi|}(b-a)||X-Y||_{L^2}\,. \tag{3.13}$$

Using (3.6) with $t=b$, we can write $\Phi^U_{b,a}(X,Y) = A+B$, where $A = X - Y + \sum_{1\leq|I|\leq k} U_I(b,a)(E_I^f(X,Y))$ and $B = R_{k,b,a,U,E^n,f}(X,Y)$. We have already estimated $||B||_{L^2}$ in (3.13). To get a bound for $||A||_{L^2}$ write

$$\begin{aligned}||A||^2 = ||X-Y||^2 &+ 2 \sum_{1\leq|I|\leq k} U_I(b,a)\langle X-Y, E_I^f(X,Y)\rangle \\ &+ 2 \sum_{1\leq|I|,|J|\leq k} U_I(b,a)U_J(b,a)\langle E_I^f(X,Y), E_J^f(X,Y)\rangle.\end{aligned} \tag{3.14}$$

Then

$$\begin{aligned}\mathbb{E}_a(||A||^2) = ||X-Y||^2 &+ 2 \sum_{1\leq|I|\leq k} \mathbb{E}_a(U_I(b,a))\langle X-Y, E_I^f(X,Y)\rangle \\ &+2 \sum_{1\leq|I|,|J|\leq k} \mathbb{E}_a(U_I(b,a)U_J(b,a))\langle E_I^f(X,Y), E_J^f(X,Y)\rangle\ ,\end{aligned} \tag{3.15}$$

so that

$$\mathbb{E}_a(||A||^2) \leq (1 + (2\nu CD + \nu^2 CD^2)(b-a))||X-Y||^2\ . \tag{3.16}$$

Taking expectations, we get

$$\mathbb{E}(||A||^2) \leq (1 + (2\nu CD + \nu^2 CD^2)(b-a))\mathbb{E}(||X-Y||^2)\ , \tag{3.17}$$

so that

$$||A||_{L^2} \leq (1 + (\nu CD + \frac{1}{2}\nu^2 CD^2)(b-a))||X - Y||_{L^2} . \tag{3.18}$$

Combining (3.18) with the bound for B, we get

$$||\Phi^U_{b,a}(X) - \Phi^U_{b,a}(Y)||_{L_2} \leq (1 + K e^{\mu|\pi|}(b-a))||X - Y||_{L^2} , \tag{3.19}$$

with $K = \nu CD + \frac{1}{2}\nu^2 CD^2 + \mu(1 + \nu CD)$. ▮

4. The Chen-Fliess Series. It is clear from the preceding considerations that it is important to be able to analyze sums of the form

$$\sum_{1 \leq |I| \leq k} U_I(b,a)(f_{I\#}\varphi)(x) .$$

To compute such expressions, we use the formalism of the *Chen-Fliess series* (cf. [2], [6], [7], [8]).

If $\mathcal{X}$ is a nonempty set, we use $\hat{\mathbf{A}}(\mathcal{X})$ to denote the algebra of *noncommutative formal power series* in $\mathcal{X}$, i.e. the set of all infinite linear combinations $\sum_{M \in \mathcal{M}(\mathcal{X})} s_M M$, where $\mathcal{M}(\mathcal{X})$ is the set of all *monomials* in $\mathcal{X}$, that is, the set of all finite sequences of elements of $\mathcal{X}$. The length of a monomial is its *degree*. Monomials are multiplied by just concatenating them, and then the product of two elements of $\hat{\mathbf{A}}(\mathcal{X})$ is well defined. The empty sequence is a monomial of degree 0, and is denoted by 1. Then $1.S = S.1 = S$ for all $S \in \hat{\mathbf{A}}(\mathcal{X})$. A linear combination of monomials of degree k is said to be *homogeneous of degree* k, and the set of all such combinations is denoted by $\mathbf{A}^k(\mathcal{X})$. Clearly, every $S \in \hat{\mathbf{A}}(\mathcal{X})$ has a unique decomposition $S = \sum_{k=0}^{\infty} H_k(S)$ as a sum of homogeneous components. If we regard $\hat{\mathbf{A}}(\mathcal{X})$ as a Lie algebra, with the bracket defined by $[S,T] = ST - TS$, then the Lie subalgebra of $\hat{\mathbf{A}}(\mathcal{X})$ generated by $\mathcal{X}$ is denoted by $\mathbf{L}(\mathcal{X})$ and its elements are known as *Lie polynomials* in $\mathcal{X}$. Those $S \in \hat{\mathbf{A}}(\mathcal{X})$ all whose homogeneous components $H_k(S)$ are in $\mathbf{L}(\mathcal{X})$ are known as *Lie series* in $\mathcal{X}$, and the set of all such series is denoted by $\hat{\mathbf{L}}(\mathcal{X})$. The *order* $\omega(S)$ of a series $S \in \hat{\mathbf{A}}(\mathcal{X})$ is the smallest k such that the k-th homogeneous component of S is $\neq 0$. (If $S = 0$ then $\omega(S)$ is defined to be $+\infty$.) An infinite sum $S_1 + S_2 + S_3 + \ldots$ of series in $\hat{\mathbf{A}}(\mathcal{X})$ such that $\omega(S_j) \to \infty$ as $j \to \infty$ is convergent in an obvious way since, for each k, $H_k(S_j) = 0$ for all but finitely many j's. In particular, the exponential e^S, and the logarithm $\log(1+S)$ are well defined by the usual power series if $\omega(S) \geq 1$. If S is a Lie series then $\omega(S) \geq 1$, so e^S and $\log(1+S)$ are defined. The elements of the form e^S, with $S \in \hat{\mathbf{L}}(\mathcal{X})$, are known as *exponential Lie series* in $\mathcal{X}$.

Given an input $U \in \mathcal{U}^m$, we can consider the differential equation

$$\dot{S}(t) = S(t)(X_0 + u_1(t)X_1 + \ldots + u_m(t)X_m) , \tag{4.1}$$

where $u_i = \dot{U}_i$ and $X_0, \ldots, X_m$ are formal noncommutative indeterminates. We can regard S as evolving in the algebra $\hat{\mathbf{A}}(X_0, \ldots, X_m)$ of noncommutative formal power series in the $m+1$ indeterminates $X_0, \ldots, X_m$. (That is, $\hat{\mathbf{A}}(X_0, \ldots, X_m)$ is the set of all formal infinite sums $S = \sum_{I \in \mathcal{I}(m)} s_I X_I$, where, if $I = (i_1, \ldots, i_r)$, $r > 0$, we define $X_I = X_{i_1} X_{i_2} \ldots X_{i_r}$, and we let $X_\emptyset = 1$.) If we solve (4.1) with initial condition $S(a) = 1$, then the solution is given by

$$S(t) = \sum_{I \in \mathcal{I}(m)} U_I(t,a) X_{I\#} \ . \tag{4.2}$$

We can also consider (4.1) as evolving in $\mathbf{A}_k(X_0, \ldots, X_m)$, the free nilpotent associative algebra of order k in $X_0, \ldots, X_m$, i.e. the set of all sums $S = \sum_{I \in \mathcal{I}(m),\, |I| \le k} s_I X_I$, where monomials are multiplied in the usual way, and every monomial of degree $> k$ is set equal to zero. In this case, the solution is given by

$$S(t) = \sum_{I \in \mathcal{I}(m),\, |I| \le k} U_I(t,a) X_{I\#} \ . \tag{4.3}$$

The value at b of this solution will be denoted by $S_{k,a,b}(U)$, or $S_{k,a,b}(u)$, and referred to as the *Chen-Fliess series of U from a to b, truncated at order k*. Formula (4.3) shows that $S_{k,a,b}(U)$ is just a way of coding all the iterated integrals $U_I(b,a)$, $|I| \le k$, into one algebraic expression.

It is clear that, if a function $t \to S(t)$ is a solution of (4.2), and $Q \in \mathbf{A}_k(X_0, \ldots, X_m)$, then $t \to QS(t)$ is also a solution. In particular, if $a < b < c$, then $t \to S_{k,a,t}(U)$ and $t \to S_{k,a,b}(U) S_{k,b,t}(U)$ are both solutions, whose values at $t = b$ coincide. Hence the identity

$$S_{k,a,c}(U) = S_{k,a,b}(U) S_{k,b,c}(U) \tag{4.4}$$

holds in $\mathbf{A}_k(X_0, \ldots, X_m)$. Notice that, when $k = 1$, Formula (4.4) just amounts to the statement that $U_I(c,a) = U_I(c,b) + U_I(b,a)$ whenever $|I| = 1$, i.e. to the property that the integral is additive with respect to the interval. So (4.4) can be viewed as a generalization to high-order iterated integrals of the additivity property.

We will need the *Campbell-Hausdorff formula* (CHF), cf. [1]. To state the CHF, let A, B be indeterminates. The CHF then says that

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+C(A,B)} \ , \tag{4.5}$$

where $C \in \hat{\mathbf{L}}(A,B)$ is a Lie series in A, B of order 3. Naturally, if $S, T \in \hat{\mathbf{L}}(X_0, \ldots, X_m)$, we can plug them into (4.5) and get $e^S e^T = e^{S+T+\frac{1}{2}[S,T]+C(S,T)}$, so in particular the set of Lie series is closed under multiplication. A similar formula holds in $\mathbf{L}_k(X_0, \ldots, X_m)$ (where $\mathbf{L}_k(X_0, \ldots, X_m)$ is the truncated version of $\hat{\mathbf{L}}(X_0, \ldots, X_m)$, i.e. the Lie subalgebra of $\mathbf{A}_k(X_0, \ldots, X_m)$ generated by the X_i), and in this case the series $C(S,T)$ is actually a finite sum.

5. Construction of Approximating Processes. We now fix m and define, for each k, each formal bracket $B = [X_{i_1}, [\ldots, [X_{i_{r-1}}, X_{i_r}]\ldots]]$, $i_j \in \{1, \ldots, m\}$ for $j = 1, \ldots, r$, each interval $[a, b] \subset [0, \infty)$, and each real number $\tau > 0$, two controls $u(B, a, b, \pm, \tau) : [a, b] \to \mathbb{R}^m$, such that

$$S_{k,a,b}(u(B, a, b, \pm, \tau)) = \mathrm{e}^{(b-a)X_0 \pm \tau^r B + Z(B,a,b,\pm,\tau)} \ , \tag{5.1}$$

where $Z(B, a, b, \pm, \tau) = P_B^{\pm}((b-a)X_0, \tau X_1, \ldots, \tau X_m)$, and $P_B^{\pm}$ are Lie polynomials of order ≥ 2 in indeterminates $Y_0, \ldots, Y_m$, that do not contain monomials in $Y_1, \ldots, Y_m$ of degree $\leq r$ (i.e. $P_B^{\pm}$ are such that $\omega(P_B^{\pm}) \geq 2$ and $\omega(P_B^{\pm}(0, Y_1, \ldots, Y_m)) > r$).

The $u(B, a, b, \pm, \tau) : [a, b] \to \mathbb{R}^m$ are constructed inductively as follows. Assume first that $r = 1$, so $B = X_i$ for some $i \in \{1, \ldots, m\}$. Then (writing $u(i, a, b, \pm, \tau)$ instead of $u(X_i, a, b, \pm, \tau)$) we define $u(i, a, b, \pm, \tau)$ to be the control whose i-th component is constant and equal to $\pm\frac{\tau}{b-a}$, while all the other components are zero. Now assume that $u(B', a, b, \pm, \tau)$ has been defined whenever B' has degree $r - 1$. Pick B of degree r, and write $B = [X_i, B']$. Divide the interval $I = [a, b]$ into four equal subintervals $I_j = [t_{j-1}, t_j]$, $j = 1, \ldots, 4$, where we let $t_j = a + j\delta$, for $j = 0, \ldots, 4$, with $\delta = \frac{1}{4}(b - a)$. Then define $u(B, a, b, +, \tau)$ to be equal to $u(i, t_0, t_1, +, \tau)$ on I_1, to $u(B', t_1, t_2, +, \tau)$ on I_2, to $u(i, t_2, t_3, -, \tau)$ on I_3, and to $u(B', t_3, t_4, -, \tau)$ on I_4. Having defined $u(B, a, b, +, \tau)$, we construct $u(B, a, b, -, \tau)$ by "changing sign and reversing time," that is, by letting $u(B, a, b, -, \tau)(t) = -u(B, a, b, +, \tau)(a + b - t)$ for $a \leq t \leq b$.

With this definition of the $u(B, a, b, \pm, \tau)$, we now show by induction on r that the Chen-Fliess series of $u(B, a, b, \pm, \tau)$ satisfies the desired properties. Consider first the case $r = 1$. In this case, it is obvious that

$$S_{k,a,b}(u(i, a, b, \pm, \tau)) = \mathrm{e}^{(b-a)X_0 \pm \tau X_i} \ . \tag{5.2}$$

Now assume that the desired property holds for $r - 1$. Let B be of degree r, and write $B = [X_i, B']$. In view of (4.4), we have

$$S_{k,a,b}(u(B, a, b, +, \tau)) = S_1 S_2 S_3 S_4 \ , \tag{5.3}$$

where $S_j = S_{k,t_{j-1},t_j}(u(B_j, t_{j-1}, t_j, \theta_j, \tau))$, $B_1 = B_3 = X_i$, $B_2 = B_4 = B'$, $\theta_1 = \theta_2 = +$, $\theta_3 = \theta_4 = -$. Then $S_1 = \mathrm{e}^{\delta X_0 + \tau X_i}$ and $S_3 = \mathrm{e}^{\delta X_0 - \tau X_i}$, where $\delta = \frac{b-a}{4}$. By the inductive hypothesis, we have

$$S_2 = \mathrm{e}^{\delta X_0 + \tau^{r-1} B' + R^+} \qquad \text{and} \qquad S_4 = \mathrm{e}^{\delta X_0 - \tau^{r-1} B' + R^-} \ ,$$

where $R^{\pm} = P_{B'}^{\pm}(\delta X_0, \tau X_1, \ldots, \tau X_m)$, and the $P_{B'}^{\pm}$ are Lie polynomials in $Y_0, \ldots, Y_m$ such that $\omega(P_{B'}^{\pm}) \geq 2$ and $\omega(P_{B'}^{\pm}(0, Y_1, \ldots, Y_m)) \geq r$.

We now repeatedly apply the CHF. (In our case, all the Lie series occurring in the computation are actually Lie polynomials, because we are working in a nilpotent algebra.) We get $S_1 S_2 = \mathrm{e}^{Z^+}$, $S_3 S_4 = \mathrm{e}^{Z^-}$, where

$$Z^{\pm} = 2\delta X_0 \pm \tau X_i \pm \tau^{r-1} B' + \frac{1}{2}\tau^r [X_i, B] + Q^{\pm} \ , \tag{5.4}$$

$$S_{k,a,b}(u(B,a,b,+,\tau)) = S_1S_2S_3S_4 = \mathrm{e}^{(b-a)X_0+\tau^r[X_i,B]+Q} \ , \tag{5.5}$$

$$Q^{\pm} = R^{\pm} \pm \frac{1}{2}\delta\tau^{r-1}[X_0,B'] \pm \frac{1}{2}\delta\tau[X_i,X_0'] + \frac{1}{2}\delta[X_0,R^{\pm}]$$

$$\pm\frac{1}{2}\tau^r[X_i,R^{\pm}] + C(\delta X_0 \pm \tau X_i, \delta X_0 \pm \tau^{r-1}B' + R^{\pm}) \ , \tag{5.6}$$

$$Q = Q^+ + Q^- + \frac{1}{2}[Z^+,Z^-] + C(Z^+,Z^-) \ . \tag{5.7}$$

It is clear that Q is a Lie polynomial in $\delta X_0, \tau X_1, \ldots, \tau X_m$, i.e. $Q = P_B^+(\delta X_0, \tau X_1, \ldots, \tau X_m)$ for some Lie polynomial in $Y_0, \ldots, Y_m$. Moreover, P_B^+ clearly has order ≥ 2. We must now show that $\omega(P_B^+(0, Y_1, \ldots, Y_m)) > r$. That is, we must show that, if we plug in $\delta = 0$ in Q, then the resulting expression is divisible by τ^{r+1}. It is easy to see that, in the right-hand side of (5.7), the only possible terms of degree $\leq r$ in τ must come from the sum $Q^+ + Q^-$. Using (5.6), we conclude immediately that such terms can only arise from the sum $P_{B'}^+(0, \tau X_1, \ldots, \tau X_m) + P_{B'}^-(0, \tau X_1, \ldots, \tau X_m)$. So our conclusion will follow if we show that this sum vanishes, i.e. that $P_{B'}^+(0, Y_1, \ldots, Y_m) + P_{B'}^-(0, Y_1, \ldots, Y_m) = 0$. This in turn follows from the equality $\tilde{S}_2\tilde{S}_4 = 1$, where the $\tilde{S}_j$ are the series obtained from the S_j by setting $X_0 = 0$. These series can be computed by setting $X_0 = 0$ in (4.1) and then solving on the intervals $[t_{j-1}, t_j]$ with input $u(B, a, b, \pm, \tau)$. If we let $\hat{S}_j$ denote the corresponding solutions with initial condition $\hat{S}_j(t_{j-1}) = 1$, then $\tilde{S}_j = \hat{S}_j(t_j)$. By translation invariance, we have $\tilde{S}_2 = S_+(\delta),$, $\tilde{S}_4 = S_-(\delta)$, where $S_\pm$ is the solution of (4.1) on $[0, \delta]$, with input $u(B', 0, \delta, \pm, \tau)$ and initial condition $S_\pm(0) = 1$. Since, as explained above, $u(B', a, b, -, \tau)$ is obtained from $u(B', a, b, +, \tau)$ by changing sign and reversing time, it follows easily that $S_-(\delta) = S_+(\delta)^{-1}$, completing the proof of our conclusion.

We record for future use the trivial fact that

$$|u(B,a,b,\pm,\tau)_i| \leq \frac{4^{r-1}\tau}{b-a} \ . \tag{5.8}$$

We now let $g \in \Lambda$, so we can write $g = \sum_{\mu=1}^{p} g_\mu B_\mu(f)$, where the B_μ are Lie brackets of the form

$$[X_{i_1^\mu}, [\ldots, [X_{i^\mu_{r(\mu)-1}}, X_{i^\mu_{r(\mu)}}]\ldots]] \ , \quad i_k^\mu \in \{1, \ldots, m\} \ ,$$

$r(\mu) \geq 2$, and $B_\mu(f)$ is the vector field obtained by plugging in f_i for X_i for each i. We assume, without loss of generality, that all the numbers g_μ are nonzero.

We now let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{\mathcal{F}_t\}$ be a filtration as above. Let $W = (W_1, \ldots, W_m)$ be an m-dimensional standard Wiener process on $(\Omega, \mathcal{F}, P)$ that has continuous sample paths and is adapted to $\{\mathcal{F}_t\}$, in the sense that W_t is $\mathcal{F}_t$-measurable and $W_t - W_s$ is independent from $\mathcal{F}_s$ whenever $s < t$. For each integer $\nu = 1, 2, \ldots$ we let π_ν be the partition $\{t_j^\nu\}_{j=0}^\infty$, where $t_j^\nu = j2^{-\nu}$. We write $\Delta W_i(j,\nu) = W_i(j2^{-\nu}) - W_i((j-1)2^{-\nu})$.

Using the $u(B, a, b, \pm, \tau)$ defined above, we will construct for each ν a π_ν-adapted input process U^ν. We define U^ν by specifying its derivative $u^\nu = \dot{U}^\nu$. Divide the interval $I_j^\nu = [(j-1)2^{-\nu}, j2^{-\nu}]$ into two equal subintervals $I_j^{\nu,-}$, $I_j^{\nu,+}$. On $I_j^{\nu,-}$, we let the component u_i^ν be equal to $2^{\nu+1}\Delta W_i(j,\nu)$ if $|\Delta W_i(j,\nu)| \leq 2^{-\frac{2\nu}{5}}$, and to zero otherwise. (It then follows, in particular, that $|u_i^\nu(t)| \leq 2^{1+\frac{3\nu}{5}}$.) On $I_j^{\nu,+}$ we proceed as follows. Let

$$\alpha_\mu = \frac{|g_\mu|}{|g_1| + \ldots + |g_p|},$$

so that $0 < \alpha_\mu$ and $\alpha_1 + \ldots + \alpha_p = 1$. Divide $I_j^{\nu,+}$ into intervals $I_j^{\nu,+,\mu}$, $\mu = 1, \ldots, p$, of length $\alpha_\mu 2^{-\nu-1}$. If $I_j^{\nu,+,\mu} = [a(j,\nu,\mu), b(j,\nu,\mu)]$, then we let u^ν be equal to $u(B_\mu, a(j,\nu,\mu), b(j,\nu,\mu), \pm, \tau_{\mu,\nu})$, where the sign is $+$ or $-$ depending on whether g_μ is > 0 or < 0, and the number $\tau_{\mu,\nu}$ is chosen so that $\tau_{\mu,\nu}^{r(\mu)} = |g_\mu| 2^{-\nu}$.

Then, if we apply (5.8) to the controls u_i^ν on an interval $I_j^{\nu,+,\mu}$, we get $|u_i^\nu(t)| \leq \alpha_\mu^{-1} 2^{\nu+1} 4^{r(\mu)-1} |g_\mu|^{\frac{1}{r(\mu)}} 2^{-\frac{\nu}{r(\mu)}}$. We now pick $k \geq 3$ such that $r(\mu) \leq k$ for all μ. We then have the pointwise inequality

$$|u_i^\nu(t)| \leq \kappa 2^{\rho\nu}, \tag{5.9}$$

where $\kappa = 2\max(1, \max\{\alpha_\mu^{-1} 4^{r(\mu)-1} |g_\mu|^{\frac{1}{r(\mu)}} : \mu = 1, \ldots, p\})$ and $\rho = \frac{k-1}{k}$. (This has just been shown to be true on $I_j^{\nu,+}$, but it clearly holds on $I_j^{\nu,-}$ as well, since (a) $\kappa \geq 2$ and (b) $\rho \geq \frac{3}{5}$, because $k \geq 3$.)

We let $I_j^{\nu,-} = [a(j,\nu,0), b(j,\nu,0)]$. Then it is easy to see that

$$S_{k,a(j,\nu,\mu),b(j,\nu,\mu)}(U^\nu) = \mathrm{e}^{2^{-\nu-1}X_0 + \sum_{i=1}^m \tilde{\Delta} W_i(j,\nu) X_i} \tag{5.10}$$

if $\mu = 0$, where $\tilde{\Delta} W_i(j,\nu) = \chi_{\nu,j,i} \Delta W_i(j,\nu)$, and $\chi_{\nu,j,i}$ is the indicator function of the set $\tilde{\mathcal{B}}_{\nu,j,i} = \{\omega \in \Omega : |\Delta W_i(j,\nu)| \leq 2^{-\frac{2\nu}{5}}\}$. If $\mu > 0$, we have

$$S_{k,a(j,\nu,\mu),b(j,\nu,\mu)}(U^\nu) = \mathrm{e}^{\alpha_\mu 2^{-\nu-1} X_0 + g_\mu 2^{-\nu} B_\mu + \cdots}, \tag{5.11}$$

where "..." denotes a Lie polynomial whose coefficients are bounded by a fixed constant times $2^{-\theta\nu}$, where $\theta = \frac{k+1}{k}$. From this, using the Campbell-Hausdorff formula, we conclude that

$$S_{k,(j-1)2^{-\nu},j2^{-\nu}}(U^\nu) = \mathrm{e}^{2^{-\nu}X_0 + \sum_{i=1}^m \tilde{\Delta} W_i(j,\nu) X_i + 2^{-\nu} G + \cdots}, \tag{5.12}$$

where $G = \sum_{i=1}^{p} g_\mu B_\mu$.

Notice that, since $2^{\frac{\nu}{2}}\Delta W_i(j,\nu)$ is normalized Gaussian, we have

$$P(\tilde{\mathcal{B}}_{\nu,j,i}) \geq 1 - \sqrt{\frac{2}{\pi}}2^{-\frac{\nu}{10}}\mathrm{e}^{-2^{\frac{\nu}{5}-1}} \tag{5.13}$$

so that, for any fixed $T > 0$, if we let $\mathcal{B}_{T,N}$ be the event that $\chi_{\nu,j,i} = 1$ for all i, j, ν such that $i \in \{1, \ldots, m\}$, $j2^{-\nu} \leq T$, and $\nu \geq N$, then we have $P(\mathcal{B}_{T,N}) \geq 1 - T\sqrt{\frac{2}{\pi}}\sum_{\nu=N}^{\infty} 2^{\frac{9\nu}{10}}\mathrm{e}^{-2^{\frac{\nu}{5}-1}}$, so that $P(\mathcal{B}_{T,N}) \to 1$ as $N \to \infty$.

We will need the following technical result:

LEMMA 5.1. *The process U^ν is in $OIP(m,k,C,\pi_\nu)$, where C is a fixed constant, independent of ν.*

PROOF. The iterated integrals $U_I^\nu(t_j^\nu, t_{j-1}^\nu)$ for $1 \leq |I| \leq k$ can be obtained from the Chen-Fliess series (5.12) by computing the exponential. Since $|\tilde{\Delta}W_i(j,\nu)| \leq 2^{-\frac{2\nu}{5}}$, it is clear that all the coefficients of $S_{k,(j-1)2^{-\nu},j2^{-\nu}}(U^\nu) - 1$ (i.e. all the $U_I^\nu(t_j^\nu, t_{j-1}^\nu)$ with $1 \leq |I| \leq k$) are pointwise bounded by a fixed constant times $2^{-\frac{2\nu}{5}}$, so that (3.3) holds. Moreover, it follows from (5.12) that

$$S_{k,(j-1)2^{-\nu},j2^{-\nu}}(U^\nu) = 1 + \sum_{i=1}^{m} \tilde{\Delta}W_i(j,\nu)X_i$$

$$+ \sum_{i,i'=1}^{m} \tilde{\Delta}W_i(j,\nu)\tilde{\Delta}W_{i'}(j,\nu)X_iX_{i'} + \ldots\ , \tag{5.14}$$

where "..." denotes a finite sum of terms that are bounded by a fixed constant times $2^{-\nu}$. So the conditions of (3.1) and (3.2) will be trivially verified if we show that, if we let A be any of the variables $V_i = \tilde{\Delta}W_i(j,\nu)$ or $V_{ij} = \tilde{\Delta}W_i(j,\nu)\tilde{\Delta}W_{i'}(j,\nu)$, then $|\mathbb{E}(A)|$ and $\mathbb{E}(A^2)$ are both bounded by a constant times $2^{-\nu}$. (Since A is independent from $\mathcal{F}_{t_{j-1}^\nu}$, we can compute true expectations instead of conditional ones.) And these bounds follow trivially from the fact that $V_i = 2^{-\nu/2}H_i$, where the H_i are obtained by symmetrically truncating normalized Gaussian random variables. This completes the proof that the bounds (3.1), (3.2), (3.3) hold.

As for (3.4), recall that the components $u_i^\nu(t)$ satisfy (5.9). Since every integration over an interval of length $\leq 2^{-\nu}$ improves the bound by a factor of $2^{-\nu}$, we conclude that a k-th order iterated integral of u^ν is bounded by $\kappa^k 2^{(\rho-1)k\nu}$, i.e. by $\kappa^k 2^{-\nu}$. So

$$|(u^\nu)_I^{-,t}(s)| \leq \kappa^{k+1}2^{(\rho-1)\nu}\ , \tag{5.15}$$

and (3.4) holds, since $\rho < 1$. ∎

It is clear from our construction that $U^\nu(j2^{-\nu}) = W(j2^{-\nu})$ on $\mathcal{B}_{T,N}$, if $j2^{-\nu} \leq T$. In view of (5.9), we have $||U^\nu(t) - U^\nu(j2^{-\nu})|| \leq c2^{-\frac{\nu}{k}}$ pointwise, where c is a fixed constant. Since $P(\mathcal{B}_{T,N}) \to 1$ as $N \to \infty$, and W has continuous sample paths, it follows that

$$(5.16) \qquad P(\lim_{\nu\to\infty}(\sup\{||W(t) - U^\nu(t)|| : 0 \leq t \leq T\}) = 0) = 1$$

for every $T > 0$, so the U^ν are indeed approximations of W.

6. Proof of Convergence. We now fix an $\mathcal{F}_0$-measurable square-integrable initial condition $\bar{X} : \Omega \to \mathbb{R}^n$, and let $t \to X(t)$ denote the Stratonovich solution of (1.3) such that $X(0) = \bar{X}$. Also, let W^ν denote the ordinary input process such that $W^\nu(t_j^\nu) = W(t_j^\nu)$ for all j, and W^ν is linear on the intervals $[t_{j-1}^\nu, t_j^\nu]$ of the partition π_ν. Let $w^\nu = \dot{W}^\nu$. Define $\tilde{w}^\nu$ to be the result of truncating w^ν as before, i.e. let $\tilde{w}_i^\nu = w_i^\nu$ on $[t_{j-1}^\nu, t_j^\nu]$ if on that interval $|w_i^\nu| \leq 2^{-\frac{2\nu}{5}}$, and otherwise let $\tilde{w}_i^\nu = 0$. We then let $\tilde{W}^\nu$ be the integral of $\tilde{w}^\nu$.

It is clear that both W^ν and $\tilde{W}^\nu$ are π_ν-adapted OIP's. Moreover, the $\tilde{W}^\nu$ are in $OIP(m, k, C, \pi_\nu)$ for a fixed C, independent of ν, provided that $k \geq 2$. (The proof is analogous to, but easier than that of Lemma 5.1.) It is then easy to see that

$$(6.1) \qquad S_{k,(j-1)2^{-\nu},j2^{-\nu}}(\tilde{W}^\nu) = e^{2^{-\nu}X_0 + \sum_{i=1}^m \tilde{\Delta} W_i(j,\nu)X_i} .$$

We now want to consider the maps $\Phi_{a,b}^U$ defined for an OIP U, using the equation $dx = (f_0(x) + g(x))dt + \sum_{i=1}^m f_i(x)dU_i$ instead of (2.1). We will use $\hat{\Phi}_{a,b}^U$ to denote these maps, so as to avoid any confusion with the $\Phi_{a,b}^U$ that are associated to (2.1).

THEOREM 6.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space endowed with a filtration $\{\mathcal{F}_t\}$. Let W be an m-dimensional standard Wiener process with respect to $\{\mathcal{F}_t\}$. Let $f_0, \dots, f_m \in C_b^\infty(\mathbb{R}^n, \mathbb{R}^n)$, let Λ_0 be the Lie algebra of vector fields generated by $f_1, \dots, f_m$, and let $\Lambda = [\Lambda_0, \Lambda_0]$. Let $g \in \Lambda$. Let $\bar{X} \in L^2((\Omega, \mathcal{F}_0, P); \mathbb{R}^n)$, and let $t \to X(t)$ be the Stratonovich solution of (1.3) with $X(0) = \bar{X}$. Let $\{U^\nu\}$ be the ordinary input processes constructed in §5. Then, for every $T > 0$,*

$$(6.2) \qquad \Phi_{T,0}^{U^\nu}(\bar{X}) \to X(T) \qquad \text{a.s.} \qquad \text{as} \qquad \nu \to \infty .$$

PROOF. Lemma 3.1 gives us estimates

$$(6.3) \qquad ||\Phi_{t_j^\nu,t_{j-1}^\nu}^{U^\nu}(X) - \Phi_{t_j^\nu,t_{j-1}^\nu}^{U^\nu}(Y)||_{L_2} \leq (1 + K2^{-\nu})||X - Y||_{L^2} ,$$

$$(6.4) \qquad ||\hat{\Phi}_{t_j^\nu,t_{j-1}^\nu}^{\tilde{W}^\nu}(X) - \hat{\Phi}_{t_j^\nu,t_{j-1}^\nu}^{\tilde{W}^\nu}(Y)||_{L_2} \leq (1 + K2^{-\nu})||X - Y||_{L^2} ,$$

valid for all ν, j, and all square-integrable $\mathcal{F}^\nu_{t_{j-1}}$-measurable X, Y. (The exponential factor that occurs in the formula of Lemma 3.1 is bounded independently of ν, since $|\pi_\nu| \leq 1$ for all ν.)

Write $X^\nu_j = \hat{\Phi}^{\tilde{W}^\nu}_{t^\nu_j,0}(\bar{X})$, $Y^\nu_j = \Phi^{U^\nu}_{t^\nu_j,0}(\bar{X})$, and $Z^\nu_j = X^\nu_j - Y^\nu_j$. Then $X^\nu_j = \hat{\Phi}^{\tilde{W}^\nu}_{t^\nu_j,t^\nu_{j-1}}(X^\nu_{j-1})$ and $Y^\nu_j = \Phi^{U^\nu}_{t^\nu_j,t^\nu_{j-1}}(Y^\nu_{j-1})$, and therefore $Z^\nu_j = A^\nu_j - B^\nu_j$, where

$$
\begin{aligned}
A^\nu_j &= \hat{\Phi}^{\tilde{W}^\nu}_{t^\nu_j,t^\nu_{j-1}}(X^\nu_{j-1}) - \hat{\Phi}^{\tilde{W}^\nu}_{t^\nu_j,t^\nu_{j-1}}(Y^\nu_{j-1}) \\
B^\nu_j &= \hat{\Phi}^{\tilde{W}^\nu}_{t^\nu_j,t^\nu_{j-1}}(Y^\nu_{j-1}) - \Phi^{U^\nu}_{t^\nu_j,t^\nu_{j-1}}(Y^\nu_{j-1}) \ .
\end{aligned}
$$

From (6.4) we get the bound $||A^\nu_j||_{L^2} \leq (1 + K2^{-\nu})||Z^\nu_{j-1}||_{L^2}$.

We now estimate B^ν_j. Let $\hat{f}_0 = f_0 + g$, $\hat{f} = (\hat{f}_0, f_1, \ldots, f_m)$. Write $a = t^\nu_{j-1}$, $b = t^\nu_j$. We pick $k \geq \max(3, r(1), \ldots, r(p))$, and apply (2.7) for f with $U = U^\nu$, and for $\hat{f}$ with $U = \tilde{W}^\nu$, and let $t = b$. We get

$$\Phi^{U^\nu}_{b,a}(x) = \sum_{|I|\leq k} U^\nu_I(b,a)E^f_I(x) + R_{k,b,a,U^\nu,E^n,f}(x) \ , \tag{6.5}$$

$$\hat{\Phi}^{\tilde{W}^\nu}_{b,a}(x) = \sum_{|I|\leq k} \tilde{W}^\nu_I(b,a)E^{\hat{f}}_I(x) + R_{k,b,a,\tilde{W}^\nu,E^n,\hat{f}}(x) \ . \tag{6.6}$$

In view of (5.15), plus the analogous bound for $\tilde{W}^\nu$, and (2.8), the remainders $R_{k,b,a,U^\nu,E^n,f}(x)$, $R_{k,b,a,\tilde{W}^\nu,E^n,\hat{f}}(x)$ are bounded by a fixed constant times $2^{-\theta\nu}$. (Recall that $\theta = 1 + \frac{1}{k}$.) Moreover, we have

$$\sum_{|I|\leq k} U^\nu_I(b,a)E^f_I(x) = (S_{k,a,b}(U^\nu)(f)E^n)(x) \ , \tag{6.7}$$

$$\sum_{|I|\leq k} \tilde{W}^\nu_I(b,a)E^{\hat{f}}_I(x) = (S_{k,a,b}(\tilde{W}^\nu)(\hat{f})E^n)(x) \ , \tag{6.8}$$

where, for a noncommutative polynomial P in the X_i, $P(f)$ denotes the partial differential operator obtained by plugging in the f_i for the X_i. Using "..." to denote terms that are bounded by a fixed constant times $2^{-\theta\nu}$, we have

$$
\begin{aligned}
S_{k,a,b}(U^\nu) &= 1 + (b-a)(X_0 + G) + V + \ldots \ , \\
S_{k,a,b}(\tilde{W}^\nu) &= 1 + (b-a)X_0 + V + \ldots \ ,
\end{aligned}
$$

where $V = \sum_{i=1}^m \tilde{\Delta}W_i(j,\nu)X_i + \sum_{i,i'=1}^m \tilde{\Delta}W_i(j,\nu)\tilde{\Delta}W_{i'}(j,\nu)X_iX_{i'}$, so that

$$
\begin{aligned}
&\sum_{|I|\leq k} U^\nu_I(b,a)E^f_I(x) = x + (b-a)(f_0(x) + g(x)) + V(f)(x) + \ldots, \\
&\sum_{|I|\leq k} \tilde{W}^\nu_I(b,a)E^{\hat{f}}_I(x) = x + (b-a)\hat{f}_0(x) + V(f)(x) + \ldots,
\end{aligned}
$$

with $V(f) = \sum_{i=1}^{m} \tilde{\Delta} W_i(j,\nu) f_i + \sum_{i,i'=1}^{m} \tilde{\Delta} W_i(j,\nu) \tilde{\Delta} W_{i'}(j,\nu) f_i f_{i'} E^n$.

Since $\hat{f}_0 = f_0 + g$, we conclude that $||B_j^\nu|| \leq \gamma 2^{-\theta\nu}$ pointwise, where γ is a fixed constant.

Therefore $||Z_j^\nu||_{L^2} \leq (1+K2^{-\nu})||Z_{j-1}^\nu||_{L^2} + \gamma 2^{-\theta\nu}$. From this it follows easily by induction on j that $||Z_j^\nu||_{L^2} \leq j \mathrm{e}^{jK2^{-\nu}} \gamma 2^{-\theta\nu}$, i.e.

$$||\hat{\Phi}_{T,0}^{\tilde{W}^\nu}(\bar{X}) - \Phi_{T,0}^{U^\nu}(\bar{X})||_{L^2} \leq \gamma T \mathrm{e}^{KT} 2^{-\frac{\nu}{k}}, \tag{6.9}$$

if $T = t_j^\nu = j2^{-\nu}$. Actually, (6.9) holds for arbitrary T, with the factor γT replaced by $\gamma(T+\lambda)$ for some fixed $\lambda > 0$. (To see this, let $t_{j-1}^\nu \leq T < t_j$, and notice that (2.4) (for $U = U^\nu$, $\varphi = E^n$) together with (5.9) imply the pointwise bound $||\Phi_{T,t_{j-1}}^{U^\nu}(x) - x|| \leq \text{constant} \cdot 2^{-\frac{\nu}{k}}$. A similar bound holds for $\hat{\Phi}_{T,t_{j-1}}^{\tilde{W}^\nu}$.)

It follows from (6.9) that $\hat{\Phi}_{T,0}^{\tilde{W}^\nu}(\bar{X}) - \Phi_{T,0}^{U^\nu}(\bar{X}) \to 0$ almost surely. On the set $\mathcal{B}_{T,N}$, $\hat{\Phi}_{T,0}^{\tilde{W}^\nu}(\bar{X}) = \hat{\Phi}_{T,0}^{W^\nu}(\bar{X})$ for sufficiently large ν. Since $P(\cup_N \mathcal{B}_{T,N}) = 1$, we conclude that $\hat{\Phi}_{T,0}^{\tilde{W}^\nu}(\bar{X}) - \Phi_{T,0}^{W^\nu}(\bar{X}) \to 0$ almost surely. Finally, $\hat{\Phi}_{T,0}^{W^\nu}(\bar{X})$ converges a.s. to $X(T)$ by the Wong-Zakai theorem. So (6.2) holds. ∎

REFERENCES

[1] N. Bourbaki, *Groupes et Algèbres de Lie,* Élements de Mathématique, Fascicule XXXVII, Chap. II et III, Hermann, Paris, 1972.

[2] M. Fliess, *Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives,* Invent. Math. **71**, 1983, p. 521-537.

[3] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 1981.

[4] E. J. McShane, *On the use of stochastic differentials in models of random processes,* Proc. Sixth Berkeley Symp. Math. Statist. Prob. **3**, 1972, p. 263-294.

[5] R. L. Stratonovich, *A new form of representation of stochastic integrals and equations,* SIAM J. Control, 1966, p. 362-371.

[6] H. J. Sussmann, *Lie brackets and local controllability: a sufficient condition for scalar input systems,* SIAM J. Control and Optimization **21**, 1983, p. 686-713.

[7] H. J. Sussmann, *A general theorem on local controllability,* SIAM J. Control and Optimization **25**, 1987, p. 158-194.

[8] H. J. Sussmann, *Exponential Lie series and the discretization of stochastic differential equations,* in "Stochastic Differential Systems, Stochastic Control and Applications," W. H. Fleming and P. L. Lions Eds., I.M.A. vols. in Math. and its Apps. No. 10, Springer-Verlag, 1988, p. 563-582.

[9] E. Wong and M. Zakai, *On the relationship between ordinary and stochastic differential equations and applications to stochastic problems in control theory,* Proc. Third IFAC Congress, 1966, paper 3B.

[10] E. Wong and M. Zakai, *Riemann-Stieltjes approximations of stochastic integrals* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **12**, 1969, p. 87-97.