A PATHWISE VIEW ON SOLUTIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS

Eeva-Maria Sipiläinen

Presented for the degree of Doctor of Philosophy
University of Edinburgh
1993

Abstract

The Ito-Stratonovich theory of stochastic integration and stochastic differential equations has several shortcomings, especially when it comes to existence and consistency with the theory of Lebesque-Stieltjes integration and ordinary differential equations. An attempt is made firstly, to isolate the path property, possessed by almost all Brownian paths, that makes the stochastic theory of integration work. Secondly, to construct a new concept of solutions for differential equations, which would have the required consistency and continuity properties, within a class of deterministic noise functions, large enough to include almost all Brownian paths.

The algebraic structure of iterated path integrals for smooth paths leads to a formal definition of a solution for a differential equation in terms of generalized path integrals for more general noises. This suggests a way of constructing solutions to differential equations in a large class of paths as limits of operators. The concept of the driving noise is extended to include the generalized path integrals of the noise. Less stringent conditions on the Holder continuity of the path can be compensated by giving more of its iterated integrals. Sufficient conditions for the solution to exist are proved in some special cases, and it is proved that almost all paths of Brownian motion as well as some other stochastic processes can be included in the theory.

Declaration

This thesis has been composed by myself, and it is my own work.

30, September 1993

Eeva-Maria Sipiläinen

Acknowledgements

I wish to thank my supervisor Professor T.J. Lyons, Jessica Gaines and many other people at Mathematics Department of Edinburgh University, for their help and their patience. I am very grateful to my employer, the Department of Applied Mathematics at Turku University, Finland, for granting me a leave of absence for my studies in Edinburgh. My special thanks go to my many flatmates throughout the years, especially to Indi for believing in the whole thing and to Lois for helping to control the chaos of the last few days.

This thesis was financially supported by the Academy of Finland (project number 8001166) and by grants from British Council, Suomen Kulttuurirahasto, Jenny ja Antti Wihurin Rahasto, Magnus Ehrnroothin Saatio, Vilho, Yrjo ja Kalle Vaisalan Rahasto, Suomalainen Konkordialiitto, and Turun Yliopistosaatio.

Contents

1 Introduction **3**

2 Background: Classical SI and SDE theories for stochastic processes **7**

2.1 Processes of finite variation 9

2.2 SI and SDE: Ito theory 11

2.2.1 Ito integrals 12

2.2.2 Ito stochastic differential equations 20

2.3 Other stochastic integrals and SDE's 21

2.4 Shortcomings of the classical SI/SDE theory 24

3 Requirements from a more general theory of differential equations **34**

4 Exponential series expansions for solutions of differential equations **40**

4.1 Series expansions for solutions of ODEs and SDEs 42

4.2 About the algebraic structure of iterated integrals 48

5 Construction of a new DE theory **53**

5.1 Notations and assumptions 55

5.2 The sequence $\theta^{(n)}$ and its convergence 61

5.3 Case of linear DE 72

5.4 About the general case 84

5.5 Convergence towards a given flow 85

5.6 Notes and open questions 88

6 Special cases and examples **94**

6.1 Case of stochastic processes 95

6.2 A continuity result for stochastic processes 97

6.3 Brownian motion 110

6.4 Other stochastic processes 115

7 About areas **117**

7.1 Classical and generalised areas 117

7.2 Stochastic areas; P. Levy's area for Brownian motion 120

References **125**

1 Introduction

The theory of Ito and Stratonovich integration and the corresponding differential equations gives a mathematically convenient way of defining stochastic integrals and solutions to stochastic differential equations with respect to martingale and semimartingale noises; however this theory has many limitations. The range of possible integrators and integrands is fairly restricted even if it does include many interesting processes. Due to the entirely probabilistic nature of the construction of the Ito integral, it is impossible to extend the theory to larger classes; at most general we can only integrate predictable processes with respect to semimartingale noises. What is more, the Ito/Stratonovich theory and the theory of ordinary (Lebesque-Stieltjes) integration are not consistent with each other, and all kinds of problems appear in approximation and continuity results.

The aim in this thesis has been to try to redefine the theory of stochastic differential equations, to get a consistent theory which would be available for a much larger class of noises.

The idea has been to isolate the path property, possessed by almost all Brownian paths, that makes the Ito integration theory possible. A new concept of solutions to differential equations is then defined for a class of functions, which is large enough to include almost all Brownian paths. It becomes obvious from considerations on the gap between stochastic differential equations and ordinary differential equations that a consistent definition of a solution should depend on iterated path integrals of the driving noise.

The new point of view is based on the well known idea of expressing the solution for a differential equation explicitly as a continuous functional of iterated path integrals of the driving noise. The algebraic structure of iterated path integrals of smooth paths and their relation to solutions of differential equations has been examined by Lyons [37], and has been shown to extend in a generalized form to more general paths. This result enables us to give a well-defined meaning to solutions of differential equations, with respect to generalised driving noises. The concept of 'noise function' must be extended by giving not just the path but also some of the generalised iterated path integrals of it. The theory obtained can be applied in a large class of functions, including almost all paths of Brownian motion.

Solutions are defined as limits of operators, rather than using the usual Picard type iterations. In fact, once some initial iterated integrals for the path are defined, we need not look at any other integrals with respect to the path.

We will concentrate on differential equations on R^d, with respect to noises in R^n. (However, many of our results will easily extend to differential equations on smooth manifolds.)

Here is an outline of the thesis: In Chapter 2, we give a brief account of the classical Ito and Stratonovich theories of stochastic integration and stochastic differential equations, and some of their generalizations, together with an explanation on why these theories are far from complete. In Chapter 3, we lay out the goal we have in this thesis, namely to construct an alternative theory of differential equations, and we explain what kind of an approach seems to be necessary to achieve this goal.

Our new definition of solution is based on a result by Lyons [37], which clarifies the algebraic structure of the iterated path integrals of smooth paths and their relation to the solutions of differential equations with respect to these paths; these algebraic relations can then formally be extended in a generalized form for more general paths. An account of this result and some related ones is given in Chapter 4.

Our new definitions of 'paths' and 'solutions' are given in Chapter 5, and we prove some results of convergence. In Chapter 6 we look at examples of paths that our results can be applied to, including paths of Brownian motion. We give a lemma which can be used to prove for stochastic processes that they can be included in our new theory, and we prove that the method does converge for almost all Brownian paths.

In many parts of the thesis, we come accross the concept of areas and area integrals. In Chapter 7, we give a short account about what we mean by these.

NOTATIONS AND CONVENTIONS We will use all the standard notations of analysis, algebra and probability theory. We use 'a.s.' and 'a.e.' as abbreviations for 'almost surely' and 'almost everywhere'; and we will often omit these if they are obvious from the context. SDE, ODE, and DE are short for stochastic differential equations, ordinary differential equations and differential equations, respectively. General and unspecified finite constants will be denoted by C, C_1; D, D_1 etc.; these may change from line to line. R is the real line, and R^n the n-dimensional Euclidean space. We define $N = 1, 2, \ldots$ and $R_+ = [0, \infty)$. $R^n \otimes R^m$ is the set of real $n \times m$-matrices, and $\mathcal{B}(R)$ is the Borel σ-field on R. C, C^k, C^∞ denote the sets of continuous, k times continuously differentiable, and smooth functions, respectively. C_b^k denotes the set of k times continuously differentiable functions which are, together with their k first derivatives, bounded.

2 Background: Classical SI and SDE theories for stochastic processes

We will start by giving a short account of the traditional ways of defining a theory of differential equations for stochastic processes. Throughout this chapter and the rest of the thesis, we will work with a fixed filtered probability space $(\Omega, P, \mathcal{F}, (\mathcal{F}_t, t \geq 0))$, i.e. a probability space $(\Omega, P, \mathcal{F})$ with $(\mathcal{F}_t, t \geq 0)$ an increasing family of sub-σ-algebras of $\mathcal{F}$. We will assume that the probability space is complete, $\mathcal{F}_t$ is a right-continuous filtration, and $\mathcal{F}_0$ contains all P-null sets of $\mathcal{F}$. Expected values and conditional expected values of random variables will be denoted by $E(\,\cdot\,)$ and $E(\,\cdot\,|\,\cdot\,)$ respectively, as usual. $\|X\|_{L_p}$ denotes the L_p-norm of a real valued random variable X, $p \in N$:

$$\|X\|_{L_p} = \left(\int_\Omega |X(\omega)|^p \, P(d\omega)\right)^{1/p}.$$

If X_n, $n = 1, 2, \ldots$ and X are random variables, then we say that X_n converges to X in L_p if

$$\|X_n - X\|_{L_p} \to 0, \quad n \to \infty;$$

and that X_n converges to X in probability if

$$\forall \varepsilon > 0, \quad P(\,\|X_n - X\| > \varepsilon) \to 0 \;\; as \;\; n \to \infty;$$

in which case we will write

$$X = l.i.p.\, X_n, \quad as \;\; n \to \infty.$$

All stochastic processes are defined with respect to this set-up $(\Omega, P, \mathcal{F})$. They are assumed to be continuous, and $\mathcal{F}_t$-adapted if necessary. We will only look at stochastic processes with values in R, or more generally in R^n. For a R^n-valued stochastic process X, we use the obvious notations $X = (X_1, \ldots, X_n)$,

$$X : R_+ \to R^n, \quad X = (X(t), t \geq 0)$$

$$X_i : R_+ \to R, \quad X_i = (X_i(t), t \geq 0),\; i = 1 \ldots n,$$

and we will write $X(\omega)$, $X_i(t, \omega)$ etc. when we wish to emphasize that we are looking at a fixed realisation of a random variable or stochastic process. The main sources for this account are Dellacherie-Meyer [12], Durrett [15], Ikeda-Watanabe [24], Ito-McKean [25], McKean [42], Meyer [46], Rogers-Williams [51], Williams [58], and articles in *Stochastic Integrals* [59].

Definition 2.1 *A process $X = (X_t, t \geq 0)$ with values in R^n is*

- *integrable if $E\|X_t\| < \infty \; \forall t \geq 0$,*
- *square-integrable if $E\|X_t\|^2 < \infty \; \forall t \geq 0$,*
- *measurable if the mapping*

$$(t, \omega) \in [0, \infty) \times \Omega \to X_t(\omega) \in R^n$$

is $\mathcal{B}(R_+) \times \mathcal{F}/\mathcal{B}(R^n)$-measurable,

- *$\mathcal{F}_t$-adapted if X_t is $\mathcal{F}_t$-measurable for every $t \geq 0$,*

- *$\mathcal{F}_t$-predictable if the mapping*

$$(t, \omega) \in [0, \infty) \times \Omega \rightarrow X_t(\omega) \in R^n$$

  is $\mathcal{S}/\mathcal{B}(R^n)$-measurable where $\mathcal{S}$ is the predictable σ-field, i.e. the smallest σ-field on $[0, \infty) \times \Omega$ such that all left-continuous, $\mathcal{F}_t$-adapted processes are measurable.

A real valued continuous process $M = (M_t, t \geq 0)$ is an $\mathcal{F}_t$-martingale if M is $\mathcal{F}_t$-adapted, integrable, and

$$E[\, M_t \,|\, \mathcal{F}_s \,] = M_s \quad P - a.s., \quad \forall s < t.$$

An R^n-valued continuous process $B = (B_1, \ldots, B_n)$ is an n-dimensional $\mathcal{F}_t$-Brownian motion if it is $\mathcal{F}_t$-adapted with

$$E[\, \exp(i\langle \xi, B_t - B_s \rangle) | \mathcal{F}_s \,] = \exp[-(t-s)\, |\xi|^2 / 2]$$

P-almost surely, $\forall \xi \in R^n, \quad 0 \leq s \leq t$.

2.1 Processes of finite variation

For stochastic processes with almost all paths of finite variation, it is a simple matter to apply the Lebesque-Stieltjes theory of integration.

Definition 2.2 *A real valued function $A = (A_t, t \geq 0)$ is of finite variation if it can be written as a difference of two non-decreasing functions. For a function A of finite variation, we denote by $V(A) = (V(A)(t), t \geq 0)$ the total variation function* *of A: i.e.*

$$V(A)(t) - V(A)(s) = \int_s^t |dA(u)|.$$

For a function A of finite variation, the Lebesque-Stieltjes integrals

$$\int_0^t Y_s dA_s,\ t \geq 0$$

are well defined for any bounded and measurable function Y, and the integrals are controlled by the inequality

$$\left|\int_0^t Y_s dA_s\right| \leq \int_0^t |Y_s| dV(A)(s) \ \ \forall t. \tag{1}$$

A differential equation

$$dX_t = f(X_t)dA_t,\ t \geq 0$$

driven with a path A of finite variation is now interpreted as the differentiated form of the corresponding integral equation with Lebesque-Stieltjes integrals. Existence and uniqueness of solutions to differential equations like this, with Lipschitz continuous coefficients f, is well known from the theory of ODEs; the solution can be constructed as a limit of a sequence of Picard iterations, where the convergence and uniqueness rely on the inequality (1) above.

In the stochastic setting this gives the following result:

Result 2.3 *Let $A = (A_1, \ldots, A_n)$ be a stochastic process with almost all paths $A_i(\omega) : R_+ \rightarrow R$ of finite variation. Then*

- *for any real valued continuous process Y, the integrals*

$$\int_0^t Y(u) dA_i(u)$$

are a.s. well defined as pathwise Stieltjes integrals, and for almost every $\omega \in \Omega$,

$$\left|\int_0^t Y_u(\omega) dA_i(u,\omega)\right| \leq \int_0^t |Y_u(\omega)| \, dV(A_i)(u,\omega), \ \forall t \geq 0.$$

- *If f_i are Lipschitz continuous functions $R \rightarrow R$ then for P-a.e. $\omega \in \Omega$, the pathwise differential equation*

$$dX_t(\omega) = \sum_{i=1}^{n} f_i(X_t(\omega)) dA_i(t,\omega)$$

has a unique solution $X(\omega)$, obtained by Picard iterations.

This case of stochastic processes of finite variation is trivial, of course. More importantly, the concept of stochastic integration expands the set of possible integrators. Ito's stochastic integral can be defined with paths of martingales as integrators, despite the fact that for non-trivial martingales, almost all sample paths are nowhere differentiable. The Ito integral is defined through probabilistic arguments, using the stochastic nature of martingales, via a Hilbert space argument. This approach leads to a well-working theory of stochastic integration and stochastic differential equations, including an integral-differential calculus for semimartingale processes (Ito's stochastic calculus).

2.2 SI and SDE: Ito theory

Here follows a brief summary of the main points of Ito's theory of stochastic integrals and stochastic differential equations. We are only interested in continuous stochastic processes, so all the martingales in the following are assumed to be continuous.

2.2.1 Ito integrals

Denote by $\mathcal{M}_2$ the set of real valued, square-integrable $\mathcal{F}_t$-martingales $(M_t, t \geq 0)$ with $M(0) = 0$. The key concept in defining Ito integrals is the quadratic variation process $\langle M, M \rangle$ of a martingale $M \in \mathcal{M}_2$.

Definition 2.4 *For $M \in \mathcal{M}_2$, the quadratic variation process $\langle M \rangle$ is defined via the Doob-Meyer decomposition result, as the unique increasing, integrable and $\mathcal{F}_t$ - predictable process $(A_t, t \geq 0)$ for which $A(0) = 0$ and $M_t^2 - A_t$ is an $\mathcal{F}_t$-martingale. For $M, N \in \mathcal{M}_2$, the quadratic covariation process $\langle M, N \rangle_t$ is defined as*

$$\langle M, N \rangle_t = \frac{1}{2} \left(\langle M + N \rangle_t - \langle M \rangle_t - \langle N \rangle_t \right),$$

or

$$\langle M, N \rangle_t = \frac{1}{4} \left(\langle M + N \rangle_t - \langle M - N \rangle_t \right).$$

Finally, the quadratic variation process of an n-dimensional martingale $M = (M_1, \ldots, M_n)$, with components M_i in $\mathcal{M}_2$, is defined as a $R^n \otimes R^n$- valued process $\langle M \rangle$ with

$$\langle M \rangle_{ij}(t) = \langle M_i, M_j \rangle(t).$$

For an n-dimensional Brownian motion $B = (B_1, \ldots, B_n)$, the components B_i are martingales with

$$\langle B_i, B_j \rangle_t = t\delta_{i,j} \quad \forall i, j.$$

(In fact it is well known that this result is sufficient to characterize an n-dimensional Brownian motion - this is P. Levy's characterization of Brownian motion. What is more, every 1-dimensional continuous square-integrable martingale can be obtained from a Brownian motion by a time change with respect to its "internal clock" given by its quadratic variation process.)

The quadratic variation of a martingale controls very strongly the behaviour of the martingale. Obviously $E(M_t^2) = E\langle M\rangle_t$ and

$$E(M_t^2 - M_s^2|\mathcal{F}_s) = E(\langle M\rangle_t - \langle M\rangle_s|\mathcal{F}_s) \ \ \forall s < t.$$

More generally, we have the Burkholder-Davis-Gundy inequalities:

Theorem 2.5 *For any $0 < p < \infty$, there exist universal finite constants c_p, C_p, such that for every continuous square-integrable martingale M,*

$$c_p E((M_t^*)^{2p}) \leq E(\langle M\rangle_t^p) \leq C_p E((M_t^*)^{2p}),$$

where

$$M_t^* = \sup_{s\leq t} |M_s|.$$

Even pathwise the process $\langle M\rangle$ bounds the behaviour of the process M; for example, for a martingale $M \in \mathcal{M}_2$, almost surely

$$|M_t(\omega) - M_s(\omega)| \leq C(\omega)\sqrt{(\langle M\rangle_t(\omega) - \langle M\rangle_s(\omega))\,|\log\left(\langle M\rangle_t(\omega) - \langle M\rangle_s(\omega)\right)|}$$

$$\forall\, 0 < s < t < T$$

where $C(\omega) < \infty$. (This reflects the fact that any square integrable martingale is a time change of a Brownian motion; see Result 6.3 in Chapter 6 about pathwise continuity of a Brownian motion.)

The probabilistic definition of $\langle M\rangle$ above agrees with the "classical" definition of the quadratic variation of a function:

Result 2.6 *Let $(\tau_n, n \in N)$ be a collection of deterministic partitions of $[0, t]$: $\tau_n = \{0 = t_0^{(n)} < t_1^{(n)} < \ldots \leq t\}$, with maximal step length $\delta(n) \to 0$ as $n \to \infty$,*

$$\delta(n) = \max_j \{t_{j+1}^{(n)} - t_j^{(n)}\}.$$

For every $M \in \mathcal{M}_2$,

$$\sum_{t_k \in \tau_n} |M_{t_{k+1}} - M_{t_k}|^2 \to \langle M \rangle_t$$

in probability, as $n \to \infty$. Similarly,

$$\sum_{t_k \in \tau_n} (M_{t_{k+1}} - M_{t_k})(N_{t_{k+1}} - N_{t_k}) \to \langle M, N \rangle_t$$

in probability, as $n \to \infty$, for $M, N \in \mathcal{M}_2$. If M is a Brownian motion, then this holds even almost surely.

Note that $(\tau_n, n \in N)$ here has to be deterministically chosen, since obviously choosing the subdivisions after checking what the path $M(\omega)$ looks like can lead to trouble. For a general martingale, the convergence is at best in probability, or in mean square; almost sure convergence holds for a subsequence of the $(\tau_n, n \in N)$-subdivision.

The Ito integral

$$I(t) = \int_0^t \theta_s \cdot dM(s), t \geq 0$$

can now be defined when θ is an $\mathcal{F}_t$-predictable process with

$$E\left(\int_0^t \theta^2(s, \omega) d\langle M \rangle(s)\right) < \infty \quad \forall t$$

and when M is a square-integrable $\mathcal{F}_t$-martingale, via an isometry result. The reasoning goes briefly as follows. (We assume here that M is real valued; the multidimensional case is similar. We will also assume for simplicity that M is an $\mathcal{L}_2$-bounded martingale.)

- Define a space of martingales: $\mathcal{M}$ = the set of $\mathcal{F}_t$-martingales with

$$\|M\|_{\mathcal{M}} = \left(\sup_t E(M_t^2)\right)^{1/2} < \infty,$$

which is a complete metric space with the metric defined by $\|M - N\|_{\mathcal{M}}$;

- define a set of integrands for a given $M \in \mathcal{M}$ by taking $\mathcal{L}^2(M)$ = the set of $\mathcal{F}_t$-predictable processes θ_t with

$$\|\theta\|_{2,M}^2 = E\left[\int_0^\infty \theta^2(s,\omega)d\langle M\rangle(s,\omega)\right] < \infty.$$

Then $\mathcal{L}^2(M)$ is a complete metric space with the metric given by $\|\theta_1 - \theta_2\|_{2,M}$.

- Finally, define a set $\mathcal{L}_0$ of elementary ("simple" predictable) processes by taking $\mathcal{L}_0$ = the set of all $\theta \in \mathcal{L}_2(M)$ which are of the form

$$\theta(t,\omega) = \begin{cases} f_0(\omega), & t = 0 \\ f_i(\omega), & t \in (t_i, t_{i+1}] \end{cases}$$

for some $0 = t_0 < t_1 < \ldots \to \infty$, where for all i, $f_i(\omega)$ is an $\mathcal{F}_{t_i}$ -measurable random variable and $\sup_i \|f_i\|_\infty < \infty$. Then $\mathcal{L}_0$ is dense in $\mathcal{L}_2(M)$ with respect to the metric $\|\cdot\|_{2,M}$. For $\theta \in \mathcal{L}_0$, the Ito integral of θ with respect to M will be defined as the sum

$$I^M(\theta)(t) = \sum_{i=0}^{n-1} f_i(\omega)(M(t_{i+1}) - M(t_i)) \; + \; f_n(\omega)(M(t) - M(t_n))$$

when $t_n \leq t \leq t_{n+1}$.

- Now, it follows that for $\theta \in \mathcal{L}_0$,

$$I^M(\theta) \in \mathcal{M}$$

and

$$\|I^M(\theta)\|_{\mathcal{M}} = \|\theta\|_{2,M}.$$

This gives an isometry between $\mathcal{L}_0$ and $\mathcal{M}$, which then extends uniquely the mapping

$$\theta \rightarrow I^M(\theta), \quad \mathcal{L}_0 \rightarrow \mathcal{M}$$

to the mapping

$$\theta \rightarrow I^M(\theta), \quad \mathcal{L}_2(M) \rightarrow \mathcal{M}.$$

This mapping defines uniquely for each $\theta \in \mathcal{L}_2(M)$ the stochastic integral of θ with respect to $M \in \mathcal{M}$, denoted by $I^M(\theta)$ and also written as

$$I^M(\theta)(t) = \int_0^t \theta(s) \cdot dM(s), \ t \geq 0.$$

The definition of the stochastic integral can be extended by localization to include locally bounded predictable pocesses as integrands and locally square-integrable martingales as integrators.

The most useful thing about the Ito integral is that if M is a continuous square-integrable $\mathcal{F}_t$-martingale, then the integral of a $\mathcal{F}_t$-predictable process θ with respect to M is also a continuous, square integrable $\mathcal{F}_t$-martingale. Its quadratic variation is given by

$$\langle \int_0^{\cdot} \theta_s dM_s \rangle(t) = \int_0^t \theta^2(u) d\langle M \rangle_u \;\; \forall t, \; P - a.s.$$

and more generally, for two martingales M,N and two predictable processes θ, ψ,

$$\langle \int_0^{\cdot} \theta_s dM_s, \int_0^{\cdot} \psi_s dN_s \rangle(t)$$

$$= \int_0^t (\theta\psi)(u) d\langle M, N \rangle_u, \; \forall t, \quad P - a.s.$$

This gives an L_2-bound for the stochastic integrals: e.g. for all $t \geq 0$,

$$\| \int_0^t \theta_s dM_s \|_{L_2}^2 = E| \int_0^t \theta_s dM_s|^2 = E\left(\int_0^t \theta_s^2 d\langle M \rangle_s \right). \tag{2}$$

The extension of the Ito integral to include semimartingale integrators is obvious.

Definition 2.7 *An $\mathcal{F}_t$-semimartingale is a real valued process X such that*

$$X_t = X_0 + M_t + A_t,$$

where M is a locally square-integrable $\mathcal{F}_t$-martingale, A is an $\mathcal{F}_t$-adapted process with almost all paths of finite variation, and $A_0 = M_0 = 0$. (Note that the decomposition above is unique when X is continuous.) A process $X = (X_1, \ldots, X_n)$ with values in R^n is a semimartingale if each X_i is a semimartingale.

Definition 2.8 *The stochastic Ito integral of an $\mathcal{F}_t$-predictable, (locally) bounded process θ, with respect to a semimartingale X, with $X_t = X_0 + A_t + M_t$, is defined as the sum of an Ito integral with respect to the martingale component M of X*

and a Lebesque-Stieltjes integral with respect to the finite variation part A of X:

$$\int_0^t \theta_s \cdot dX_s = \int_0^t \theta_s \cdot dM_s + \int_0^t \theta_s dA_s.$$

Thus, Ito integrals of semimartingales are themselves semimartingales.

Above, the stochastic Ito integral for square integrable martingale integrators was constructed via an isometry argument, as an element in a martingale space; however the integral can also be constructed in a more concrete way as a limit of Riemannian sums along a suitably chosen sequence of subdivisions. The limit will in general be in mean square or in probability - or, if the subdivisions are good enough, even almost surely. Obviously, once again, the important thing is to choose a sequence of divisions which is independent of the path of the martingale. Here is one formulation of these kinds of results:

Result 2.9 *Let M be a square integrable martingale and θ a predictable locally bounded process. The integral*

$$\int_0^t \theta(s,\omega) \cdot dM(s,\omega)$$

is the L_2- limit as $n \to \infty$ of sums

$$S_n(\omega) = \sum_j \theta(t_j^{(n)},\omega)\left[M(t_{j+1}^{(n)},\omega) - M(t_j^{(n)},\omega)\right]$$

for all deterministic partitions

$$0 = t_1^{(n)} < t_2^{(n)} < \ldots < t_n^{(n)} = t$$

for which $\delta(n) \to 0$ as $n \to \infty$ where $\delta(n)$ is the maximal step length:

$$\delta(n) = \max_{1 \le j \le n} (t_j^{(n)} - t_{j-1}^{(n)}).$$

Along with the definition of the Ito integral comes the theory of Ito calculus, with a chain rule of differentiation now provided by the Ito formula, instead of the Newton-Leibniz differentiation rules.

Theorem 2.10 *(Ito formula)*
Let F be a function of class C^2 on R^n, and X an n-dimensional semimartingale: $X(t) = X(0) + M(t) + A(t)$, where M is the martingale component of X. Then almost surely for all $t \ge 0$,

$$F(X(t)) - F(X(0)) = \sum_{i=1}^{n} \int_0^t D_i F(X(s)) \cdot dM_i(s)$$

$$+ \frac{1}{2} \sum_{i,j=1}^{n} \int_0^t D_{ij} F(X(s)) d\langle M_i, M_j \rangle (s)$$

$$+ \sum_{i=1}^{n} \int_0^t D_i F(X(s)) dA_i(s).$$

Here

$$D_i F = \frac{\partial F}{\partial x_i}, \qquad D_{ij} F = \frac{\partial^2 F}{\partial x_i \partial x_j}.$$

This means that if X is a $\mathcal{F}_t$-semimartingale, then for any $F \in C^2$, the stochastic process $F(X) = (F(X(t)), t \ge 0)$ is also a semimartingale. Thus, the class of semimartingales is closed under action by C^2-functions.

2.2.2 Ito stochastic differential equations

The construction of the Ito integral leads immediately to the theory of stochastic differential equations with respect to martingale and semimartingale noises. An Ito SDE will be the differentiated form of an integral equation where the integrals are interpreted as stochastic Ito integrals. For our purposes, it will suffice to look at the simplest possible SDE's, i.e. time homogeneous, diffusion type SDE's with zero drift. Also, we will only look at solutions in the 'strong' sense, since we are interested here in the mapping: [noise path $M(\omega)$] $\rightarrow$ [solution $Y(\omega)$ of an SDE driven with $M(\omega)$]. So, for M: a square-integrable $\mathcal{F}_t$-martingale with values in R^n, and f: a measurable function $R^d \rightarrow R^d \otimes R^n$, we will denote our Ito differential equation on R^d by

$$dX_t = f(X_t) \cdot dM_t \tag{3}$$

or,

$$dX_i(t) = \sum_{j=1}^{n} f_{ij}(X_t) \cdot dM_j(t), \; i = 1 \ldots d.$$

Definition 2.11 *A stochastic process X is called a solution to the SDE (3) if it is a continuous and $\mathcal{F}_t$-adapted stochastic process, with values in R^d, such that almost surely,*

$$X(t) - X(0) = \int_0^t f(X_s) \cdot dM_s, \; \forall t \geq 0.$$

Definition 2.12 *Pathwise uniqueness holds for (3) if, whenever X and $\tilde{X}$ are both solutions to it and $X(0) = \tilde{X}(0)$ a.s., then $X(t) = \tilde{X}(t)$ $\forall t \geq 0$ almost surely.*

Existence and uniqueness of solutions can now be proved e.g. by the following theorem:

Theorem 2.13 *(Ito theorem of uniqueness and existence of solution)*
Let $f : R^d \to R^d \otimes R^n$ be Lipschitz continuous:

$$\|f(x) - f(y)\| \leq K|x - y| \quad \forall x, y \in R^d.$$

Then the SDE (3) has a pathwise unique solution. For each given $(x_1, \dots x_d) \in R^d$, the solution X with $X_i(0) = x_i$, $i = 1 \dots d$, can be constructed from M as a limit of successive approximations (Picard iterations).

Thus, the method of constructing the unique solution via Picard iterations works even for Ito SDE's. The proof of convergence and uniqueness is based on the fact that the L_2-norm of the Ito integrals can be controlled via equation (2).

The theorem extends naturally to the case where f is locally Lipschitz continuous and M is a locally square integrable martingale; the solution might then explode in finite time, so it would be necessary to modify the definition of solution to include this possibility. Again, definitions and results for Ito SDE's with respect to semimartingale noises are obvious.

2.3 Other stochastic integrals and SDE's

Given two semimartingales X and Y, the Stratonovich integral of Y with respect to X is defined by taking

$$\int_0^t Y_s \circ dX_s = \int_0^t Y_s \cdot dX_s + \frac{1}{2}\langle X, Y\rangle_t$$

where the quadratic covariance $\langle X, Y\rangle$ of two semimartingales is defined to be equal to the quadratic covariance of their martingale components. (Cf. Stratonovich [56]) Here and in the following we will use $\int Y \circ dX$ to denote a Stratonovich integral and $\int Y \cdot dX$ to denote an Ito integral.

Using Stratonovich integrals leads to rules of differentiation which are the same as in the ordinary (Leibniz-Newton) calculus. Thus e.g. for a semimartingale X with values in R, and a real function $F \in C^3$,

$$F(X_t) - F(X_0) = \int_0^t F'(X_s) \circ dX_s.$$

(Note that $F \in C^3$ is needed to keep the integrand in the class of semimartingales.) The Stratonovich integral is "symmetrical", with:

Result 2.14 *For X,Y real valued semimartingales, the Stratonovich integral $\int_0^t Y_s \circ dX_s$ is the limit in probability of sums*

$$S_n = \sum_{i=1}^{n} \frac{Y(t_i) + Y(t_{i+1})}{2} (X(t_{i+1}) - X(t_i))$$

for all deterministic partitions $0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_n^{(n)} = t$ for which $\delta(n) \to 0$ as $n \to \infty$, where

$$\delta(n) = \max_j \{t_j^{(n)} - t_{j-1}^{(n)}\}.$$

Of course, the result above suggests also an infinite number of other possible definitions of stochastic integrals for semimartingales, defined as limits of Riemannian sums. In general, a "λ-integral" with arbitrary $\lambda \in [0,1]$ could be defined as follows: (Cf. for example Kloeden-Platen [28].)

Definition 2.15 *Let X,Y be real valued semimartingales. The λ-integral is defined by*

$$\begin{aligned} & {}_{(\lambda)}\int_0^t Y_s dX_s \\ = \; & l.i.p. \sum_{i=1}^{n} \left((1-\lambda) Y(t_i^{(n)}) + \lambda Y(t_{i+1}^{(n)})\right) \left(X(t_{i+1}^{(n)}) - X(t_i^{(n)})\right) \end{aligned}$$

for any deterministic sequence of partitions as in Result 2.14.

So, $\lambda = 0$ gives the Ito integral and $\lambda = 1/2$ gives the Stratonovich integral. Of course, with $\lambda > 0$, λ-integrals with respect to martingales are no longer martingales. The connection with Ito integrals is as follows:

$$_{(\lambda)}\int_0^t Y_s dX_s = \int_0^t Y_s \cdot dX_s + \lambda\langle X, Y\rangle_t,$$

and the corresponding rule of partial differentiation would then be

$$F(X_t) - F(X_0) =_{(\lambda)} \int_0^t F\prime(X_s)dX_s$$

$$+ (\frac{1}{2} - \lambda)\int_0^t F''(X_s)d\langle X\rangle_s, \ \forall F \in C_3.$$

However, all these integrals are based on the Ito integral, in that proofs that the appropriate Riemannian sums do converge use the fact that the corresponding Riemannian sums converge for the Ito integral. Similarly, to establish properties of these integrals it would be necessary to convert them into Ito integrals for which the powerful martingale theory is available to give e.g. bounds for the sizes of the integrals. In practice, values other than $\lambda = 0$ or $\lambda = 1/2$ have very little interest. The Stratonovich integral will turn out to be the "natural" choice for the stochastic integral in many cases.

By choosing the Stratonovich integral as the stochastic integral, we get the notion of Stratonovich differential equations, for which we will in general use notations like $dY(t) = f(Y(t)) \circ dX(t)$. Control on the sizes of integrals, needed in Picard iteration, is not directly available for the Stratonovich integral, instead proofs of existence and uniqueness of solutions to Stratonovich SDE's are classically obtained by transforming the Stratonovich SDE into an Ito SDE (with drift).

Unfortunately, one extra derivative from the coefficient function is usually needed to carry out this transformation. So, for instance, Lipschitz continuity for the coefficient f is no longer sufficient to prove existence and uniqueness of solution, instead e.g. $f \in C^2$ is needed.

2.4 Shortcomings of the classical SI/SDE theory

The construction of the Ito integral is done with respect to a given, fixed filtration $(\mathcal{F}_t, t \geq 0)$, and the whole theory entirely relies on combining the $\mathcal{F}$-predictability of the integrand with the $\mathcal{F}$-martingale property of the integrator. It is not possible to generalize the definition of the Ito integral further, out of these fairly limited classes of processes. (See for example Dellacherie-Meyer [12] : "The only sensible integrators are semimartingales.") The Stratonovich integral is based on the Ito integral, and for its existence even more is required, since both the integrator and the integrand must be semimartingales. An alternative definition of a stochastic integral is given in Young [64] [65], but even that one is based on probabilistic reasoning and is done along a given filtration. Having to fix the filtration means that the direction of time in integration is strictly restricted in this stochastic theory. Solving stochastic differential equations with respect to semimartingales can only be done forwards in time, along the fixed filtration $(\mathcal{F}_t, t \geq 0)$.

Another major problem with the theory of stochastic integrals and stochastic differential equations for semimartingales appears when one tries to combine it with the corresponding theories for smooth or Lipschitz continuous noises. Both of these theories give well defined maps

$$\text{Noise } X \quad \rightarrow \quad I(X) = \int \theta(X)dX, \qquad \text{for fixed } \theta,$$

$$\text{Noise } X \quad \rightarrow \quad \text{solution of DE } dY = f(Y)dX, \quad \text{for fixed } f$$

where the integrals/differential equations are formally identical but can bc interpreted as Lebesque-Stieltjes, Ito, or Stratonovich ones according to what class of noises we are looking at. It turns out to be impossible to unify the classical Lebesque-Stieltjes theory of integration and differential equations, and the classical (Ito and Stratonovich) theory of stochastic integration and stochastic differential equations in such a way that these maps would in general be continuous with respect to a reasonable topology in the space of noises X.

Wong and Zakai ([61]) first pointed out that if a Brownian path is approximated by a sequence of smooth functions, then the solutions of a given differential equation driven with these noises do not converge towards the solution of the corresponding Ito differential equation but instead towards the solution of the Stratonovich equation. This is to be expected of course; while the Ito integral is mathematically very convenient, it is not the obviously correct stochastic integral to choose since it comes with its own peculiar calculus. The Stratonovich and Lebesque-Stieltjes integration on the other hand both obey the same rules of calculus. So, the solution to this particular problem is just a matter of choosing the correct (Stratonovich) definition for the stochastic integral.

It is fairly straightforward to define a unified theory of differential equations which includes both the stochastic (Stratonovich) theory and the ODE theory (for noises which are Lipschitz continuous), by replacing the differential equation by an appropriate "canonical" functional equation, which reduces to the ODE for Lipschitz noises and to the Stratonovich SDE for semimartingale noises. (See Mc Shane [43], [44]; Marcus [41].) However, there still remains the more serious and fundamental problem of "stability": Uniform convergence of noises $X_n \to X$ does not imply uniform convergence of the corresponding solutions, solution$(X_n) \to$ solution(X). This problem appears already when X_n and X are all functions of bounded variation, as the following example shows.

Example. (Sussman, [53])

Look at the linear differential equation

$$dx_t = A \cdot x_t \, d\omega_1(t) + B \cdot x_t \, d\omega_2(t)$$

where $x_t \in R^n$, $\omega_1(t)$, $\omega_2(t) \in R$, A and B are $(n \times n)$-matrices and $t \in [0,1]$. We will construct a sequence of noise functions $(\omega_1^{(n)}, \omega_2^{(n)})$ as follows: For a given n, we divide [0,1] into n intervals $I_j^{(n)}$, $j = 1 \ldots n$: thus, $I_j^{(n)} = [(j-1)/n, j/n]$; and further we divide each $I_j^{(n)}$ into four subintervals, denoted by $I_{j1}^{(n)}, \ldots, I_{j4}^{(n)}$. Now we can construct two sequences of step functions $u_1^{(n)}$ and $u_2^{(n)}$ on [0,1] by defining for all $j = 1 \ldots n$, their values on interval $I_j^{(n)}$ as follows:

subinterval	*value of* $u_1^{(n)}$	*value of* $u_2^{(n)}$
$I_{j1}^{(n)}$	$+4\sqrt{n}$	0
$I_{j2}^{(n)}$	0	$+4\sqrt{n}$
$I_{j3}^{(n)}$	$-4\sqrt{n}$	0
$I_{j4}^{(n)}$	0	$-4\sqrt{n}$.

Finally, define $\omega_i^{(n)}$ to be the function with $\dot{\omega}_i^{(n)} = u_i^{(n)}$ and $\omega_i^{(n)}(0) = 0$, $i = 1, 2$. So what happens is that for all n, during each time interval $I_j^{(n)}$ of length $1/n$, the path $(\omega_1^{(n)}, \omega_2^{(n)})$ completes a counter-clockwise loop beginning and ending at

(0,0), along a square with sides of length $1/\sqrt{n}$. Now, $\omega_1^{(n)}$ and $\omega_2^{(n)}$ both converge uniformly towards zero as $n \to \infty$. It is easy to check that solutions $x^{(n)}$ of the differential equations

$$dx_t^{(n)} = A \cdot x_t^{(n)} d\omega_1^{(n)}(t) + B \cdot x_t^{(n)} d\omega_2^{(n)}(t)$$

converge towards

$$x(t) = \exp(t\,[\,B, A\,]) \cdot x_0,$$

whereas the solution with $(\omega_1, \omega_2) = (0,0)$ is obviously $x(t) = x_0$. Therefore, uniform convergence of the noises does not imply uniform convergence of the corresponding solutions — unless, that is, $[B, A] = 0$.

Attempts to solve this stability (or continuity, or approximation) problem for the solutions of a differential equation, generally denoted here by

$$dY(t) = \sum_{i=1}^{n} f_i(Y_t) dX_i(t) \tag{4}$$

have usually been along the following lines:

1) Restrict the types of differential equations, usually by only looking either at cases with scalar noise (i.e. n=1) or for multi-dimensional noises, only cases where the vector fields f_i commute (i.e. the Lie brackets $[f_i, f_j] = 0\ \forall\, i, j$). In these cases, we can get very strong continuity results. The solution of (4) can be proved to be a continuous functional of the noise process with respect to the uniform norm (see for example Wong-Zakai [61], [62], McShane [43],Protter [48], [49] and Sussman [53] where the concept of solution to the differential equation is generalized

directly from the set of C^1 noises to the set of continuous noises).

2) An alternative in the general multidimensional case is to restrict the class of permitted approximating noises, and then prove continuity results within these classes. Thus, for example, it can be proved that when a Brownian motion or a martingale X is approximated by a sequence of processes X_n of bounded variation, then the solutions with respect to the X_ns converge towards the solution of the corresponding (Stratonovich) differential equation with respect to X, for a fairly large class of "reasonable" approximations (see, for instance, Wong-Zakai [61], Wong-Zakai [62], Ikeda-Watanabe [24], Nakao-Yamato [47], Protter [48], [49], [50], Konecny [30], Emery [16], Ikeda-Nakao-Yamato [23] — the convergence in these papers is usually in probability or in L_2-norm). This class of "reasonable approximations" of X by a sequence of smooth paths X_n include:

- Polygonal approximations of X along an arbitrary sequence of (fixed and deterministic) subdivisions $(\tau_n, n \in N)$ with step lengths decreasing to zero; i.e. X_n coincides with X at τ_n- division points and is piecewise linear between these;

- generalizations of these, where for each n, the components of X_ns are interpolated over each division interval $[t_k, t_{k+1}]$ of τ_n with given (fixed) interpolating functions ϕ_i. By this we mean that for all i, we will take:

$$X_n^i(t) = X_i(t_k) + \phi_i\left(\frac{t - t_k}{t_{k+1} - t_k}\right)(X_i(t_{k+1}) - X_i(t_k))$$

when $t \in [t_k, t_{k+1}]$, where $\phi_i \in \Phi = \{$continuous, differentiable, nondecreasing functions on [0,1] with $\phi(0) = 0$, $\phi(1) = 1$ $\}$.

Similarly, the example above could be excluded from a set of "reasonable" approximations if we agreed to only permit approximations of bounded variation functions to be by sequences of functions for which the total variation functions are uniformly bounded.

From the point of view of stochastic processes, all these results are fairly limited, in the sense that the method of approximation must be chosen in advance, deterministically. Things are complicated if we wish to pay attention to individual Brownian paths for instance, as this example shows:

Example. (See McShane [43], Ikeda-Watanabe [24])

Let (b_1, b_2) be a two dimensional Brownian motion, and look at the differential equation

$$\begin{cases} dx^1(t) &= db_1(t) \\ dx^2(t) &= x^1(t)db_2(t) \end{cases}$$

over the interval [0,1]. We will now choose any functions ϕ_1 and ϕ_2 from the class Φ such that

$$\phi_1(t) = \begin{cases} 0, & t \leq 0 \\ 1, & t \geq 1/2 \end{cases}$$

$$\phi_2(t) = \begin{cases} 0, & t \leq 1/2 \\ 1, & t \geq 1 \end{cases}$$

in which case

$$\begin{cases} \int_0^1 [1 - \phi_1(s)]\dot{\phi}_2(s)ds &= 0 \\ \int_0^1 [1 - \phi_2(s)]\dot{\phi}_1(s)ds &= 1 \end{cases}$$

(meaning that $\phi_1 = 1$ when ϕ_2 varies, and $\phi_2 = 0$ when ϕ_1 varies.) Now construct two different sequences of approximations for (b_1, b_2) denoted by $(z_1^{(n)}, z_2^{(n)})$ and $(\hat{z}_1^{(n)}, \hat{z}_2^{(n)})$ as follows: For i=1,2,

$$\begin{aligned} z_i^{(n)}(t,\omega) &\quad \text{is interpolated by } \phi_i \text{ from } b_i, \\ \hat{z}_i^{(n)}(t,\omega) &\quad \text{is interpolated by } \phi_{3-i} \text{ from } b_i \end{aligned}$$

over division intervals for which $\triangle_j b_1 \triangle_j b_2 \geq 0$;

$$\begin{aligned} z_i^{(n)}(t,\omega) &\quad \text{is interpolated by } \phi_{3-i} \text{ from } b_i, \\ \hat{z}_i^{(n)}(t,\omega) &\quad \text{is interpolated by } \phi_i \text{ from } b_i \end{aligned}$$

over division intervals for which $\triangle_j b_1 \triangle_j b_2 \leq 0$. Here we have denoted $\triangle_j b = b(t_{j+1}) - b(t_j)$. Both of the obtained sequences of approximations converge towards (b_1, b_2) uniformly. The corresponding solutions $(x_1^{(n)}, x_2^{(n)})$ and $(\hat{x}_1^{(n)}, \hat{x}_2^{(n)})$, respectively, satisfy

$$\left\{ \begin{array}{lcl} \triangle_j x_1^{(n)} &=& \triangle_j b_1 \\ \triangle_j x_2^{(n)} &=& x_1^{(n)}(t_j) \triangle_j b_2 + (\triangle_j b_1 \triangle_j b_2)^+ \end{array} \right.$$

$$\left\{ \begin{array}{lcl} \triangle_j \hat{x}_1^{(n)} &=& \triangle_j b_1 \\ \triangle_j \hat{x}_2^{(n)} &=& \hat{x}_1^{(n)}(t_j) \triangle_j b_2 - (\triangle_j b_1 \triangle_j b_2)^- \end{array} \right.$$

(where we have used for a real number a, the notations $a^+ = \sup(a, 0)$ and $a^- = \sup(-a, 0)$); so their difference $\xi^{(n)}$, where $\xi_i^{(n)} = x_i^{(n)} - \hat{x}_i^{(n)}$, $i = 1, 2$, satisfies:

$$\begin{array}{lcl} \triangle_j \xi_1^{(n)} &=& 0, \\ \triangle_j \xi_2^{(n)} &=& \xi_1^{(n)}(t_j) \triangle_j b_2 + | \triangle_j b_1 \triangle_j b_2| \end{array}$$

and

$$\xi_2^{(n)}(1) = \sum_j | \triangle_j b_1 \triangle_j b_2|.$$

Now it is easy to calculate that

$$E(\xi_2^{(n)}(1)) = 2/\pi$$

$$Var(\xi_2^{(n)}(1)) \to 0$$

as the step length of the divisions decreases to 0, so that the solutions $x_2^{(n)}(1)$ and $\hat{x}_2^{(n)}(1)$ can not have the same limit.

Of course problems arise here because the approximating method is permitted to depend on the path X, rather than being pre-fixed and deterministic.

A more detailed analysis of what can go wrong and how to prevent it is given in Bally [2], [3] and Ikeda-Watanabe [24]. There the approximating sequence X_n of processes of bounded variation, converging uniformly towards the path of our stochastic process X (a Brownian motion, for instance) can be chosen fairly freely, apart from having to agree with X at division points of a deterministically chosen sequence of subdivisions $(\tau_n, n \in N)$. Especially X_n is permitted to depend on the path of X between the division points. Bally and Ikeda-Watanabe now prove that solutions with respect to the processes X_n converge towards a solution of a modified Stratonovich differential equation with respect to X, provided that the sequence X_n satisfies a stability condition. (The convergence is proved in probability, or in L_p norms.) In Bally [2], [3] this stability condition is given as follows: Denote by a_n^{ij} the piecewise constant function defined by

$$a_n^{ij}(s) = \frac{1}{2}\,\frac{1}{t_{k+1} - t_k} \int_{t_k}^{t_{k+1}} \int_{t_k}^{s} \left(dX_n^i(u)dX_n^j(s) - dX_n^j(u)dX_n^i(s)\right)$$

when $s \in [t_k, t_{k+1}]$. Then we say that the sequence (X_n) satisfies the stability condition if

$$a_n^{ij} \to a^{ij}$$

uniformly as $n \to \infty$, for some integrable function a^{ij}. The modified Stratonovich differential equation is obtained by adding into the right hand side of the differential equation a stability correction term of the form:

$$\sum_{i,j} [f_i, f_j] (Y_s) a^{ij} ds,$$

thus the modified Stratonovich differential equation is

$$dY(t) = \sum_{i=1}^{n} f_i(Y_t) dX_i(t) + \sum_{i,j} [f_i, f_j] (Y_t) a^{ij} dt.$$

Here $[f_i, f_j]$ is the Lie bracket of the vector fields f_i, f_j. All the approximating sequences in the class of "reasonable approximations" above satisfy the stability condition, with $a^{ij} = 0$. In the example given above, the sequence $\hat{z}^{(n)}$ satisfies the stability condition with $a^{ij} = 1/\pi$, and the sequence $z^{(n)}$ satisfies the stability condition with $a^{ij} = -1/\pi$. Therefore, the solutions with respect to $(\hat{z}^{(n)})$ and $(z^{(n)})$ do converge, but towards solutions of different stochastic differential equations.

The sequence a_n^{ij}, the stability condition and the stability correction term have a very obvious meaning. The integral in the definition of a_n^{ij} is the area of the path (X_n^i, X_n^j) over the time interval $[t_k, t_{k+1}]$, therefore $\int_0^t a_n^{ij}(t)dt$ gives the difference between the area of (X_n^i, X_n^j) over an interval $[0, t]$ and the polygonal approximation of it along the τ_n-division points. The polygonal area converges towards the area of the Brownian motion X. (See results in Chapter 7). The areas of the paths (X_n^i, X_n^j) could diverge as $n \to \infty$; the stability condition states that these areas do converge (although not necessarily towards the area of the Brownian motion). The stability correction term added into the SDE then in fact corrects the differences in the areas between X and its approximating sequence X_n. The relation between the brackets of the vector fields in the differential equation and the "order of impulses" produced by the components of the noise is well known, and the order of impulses is closely related to the area between the components of the driving processes. (Cf. Sussmann, [53].) In the example above, a systematic error was made in the areas of the approximating sequences because of the way the interpolating functions were chosen according to the sign of $\triangle b_1 \triangle b_2$.

Even these results can not really be said to solve the problem of continuity of solutions of differential equations — we should not have to make stability corrections on the differential equation to obtain convergence. However, these results do clearly show the connection between area integrals of paths and solutions of differential equations with respect to the paths. It is obvious that if we want to obtain a truly continuous and consistent concept of solutions for differential equations with respect to a sufficiently large class of noises, then the areas of the paths need to be taken into account.

3 Requirements from a more general theory of differential equations

In view of the problems with the classical (Ito and Stratonovich) theory of stochastic differential equations, reviewed above, we would like to construct a new theory such that the following requirements are satisfied: Our new theory should be

- inclusive, that is, it should include the classical SDE theory and also the ODE theory for functions of finite variation;
- consistent, in that in case of finite variation or semimartingale paths respectively, the definition should agree with the classical ones;
- continuous, so that if we introduce a reasonable topology in the space of the driving noises, then the mapping from noises to solutions with respect to them should be continuous;
- more general; we would like to include into our new theory almost all paths of some additional stochastic processes which fall outside the scope of the Ito theory of stochastic differential equations.

Now, the problem of limited existence for the Ito and Stratonovich integrals and the Ito and Stratonovich theories of differential equations, caused by the probabilistic nature of their definition, could be solved if we could isolate the path property, possessed by almost all Brownian paths, that makes the Ito theory work.

Then, we could hopefully define a working theory in the class of deterministic functions with this path property.

One attempt to define a pathwise integration theory for functions with paths not of finite variation, such as almost all Brownian paths, is given in Follmer [17]. There, an Ito-type integral is defined for any function of finite quadratic variation as follows:

Definition 3.1 *Let X be a continuous real function, and let $\tau = (\tau_n, n \in N)$ be a collection of finite subdivisions of the interval $[0, \infty)$ such that the maximal step length on compact intervals goes to 0 as n increases. We say that X is of quadratic variation with respect to τ if the sequence of point measures*

$$\xi_n = \sum_{t_i \in \tau_n} (X_{t_{i+1}} - X_{t_i})^2 \varepsilon_{t_i}$$

on $[0, \infty)$ converges weakly to a measure ξ on $[0, \infty)$. Here ε_t denotes the unit point mass at point t. The measure ξ is then continuous, and will be denoted by $d\langle X \rangle_t$. The continuous and increasing function $(\langle X \rangle_t,\ t \geq 0)$ is called the quadratic variation function of X with respect to τ.

Certain integrals can now be defined with these kinds of functions as integrators:

Theorem 3.2 *If X is of quadratic variation with respect to some $\tau = (\tau_n, n \in N)$, then for every function $f \in C^1$, the integral $\int_0^t f(X_s) dX_s$ exists as a limit of Riemannian sums:*

$$\int_0^t f(X_s) dX_s = \lim_{n \to \infty} \sum_{t_i \in \tau_n \cup [0,t]} f(X_{t_i})(X_{t_{i+1}} - X_{t_i}).$$

Also, the Ito formula holds for every C^2-function F:

$$F(X_t) - F(X_0) = \int_0^t F'(X_s)dX_s + \frac{1}{2}\int_0^t F''(X_s)d\langle X\rangle_s.$$

The proof of this is done simply by using a second order Taylor expansion for a general C^2-function F over each division interval of τ_n. As $n \to \infty$, we get at limit the Ito formula above, and especially the Riemannian sums for the C^1-functions $f = F'$ must converge. But on the one-dimensional case, all C^1 -functions are of this form.

Similarly, Follmer defines the quadratic variation function for a general d-dimensional function X.

Definition 3.3 *Let $X = (X_1, \ldots, X_d)$ be a continuous d-dimensional function, and let $\tau = (\tau_n, n \in N)$ be a sequence of subdivisions as above. We say that X is of quadratic variation with respect to τ if the 1-dimensional functions X_i and $X_i + X_j$ are of quadratic variation with respect to τ, for all i,j. We then denote*

$$\langle X_i, X_j\rangle_t = \frac{1}{2}\left(\langle X_i + X_j\rangle_t - \langle X_i\rangle_t - \langle X_j\rangle_t\right).$$

The matrix valued function $\langle X\rangle$ with $\langle X\rangle_{ij}(t) = \langle X_i, X_j\rangle(t)$ is called the quadratic variation of the d-dimensional function X.

Now, obviously, the procedure above can be repeated to define Ito type integrals with these functions as integrators. The problem is that in this d-dimensional case, the class of functions f we can use is much more seriously limited.

Theorem 3.4 *If X is a d-dimensional function of quadratic variation with respect to $(\tau_n, n \in N)$, then for every f with $f = \nabla F$, $F \in C^2(R^d)$, the integral $\int_0^t f(X_s)dX_s$ exists as the limit of Riemannian sums; and the Ito formula holds for* *all* $F \in C^2(R^d)$*:*

$$F(X_t) - F(X_0) = \int_0^t DF(X_s)dX_s + \frac{1}{2}\sum_{i,j}\int_0^t D_iD_jF(X_s)d\langle X_i, X_j\rangle_s$$

where the integral $\int_0^t DF(X_s)dX_s$ *is defined as the limit of Riemannian sums:*

$$\int_0^t DF(X_s)dX_s = \lim_{n\to\infty}\sum_{t_i\in\tau_n\cup[0,t]}\left(\sum_{k=1}^{d}\frac{\partial F}{\partial x_k}(X(t_i))\,[X_k(t_{i+1}) - X_k(t_i)]\right).$$

Note that this implies the existence of integrals

$$\int_0^t (X_i(s)dX_j(s) + X_j(s)dX_i(s)), \quad i,j = 1\ldots d$$

when X is a d-dimensional function of quadratic variation, but not the existence of the individual integrals

$$\int_0^t X_i(s)dX_j(s), \quad \int_0^t X_j(s)dX_i(s)), \quad i,j = 1\ldots d.$$

For these to exist as well, we would have to assume the existence of the area integrals

$$\int_0^t (X_i(s)dX_j(s) - X_j(s)dX_i(s)), \quad i,j = 1\ldots d.$$

A Stratonovich type integral could be defined from this Ito integral in the obvious manner; the Stratonovich integral

$$\int_0^t f(X_s)\circ dX_s = \int_0^t f(X_s)dX_s + \frac{1}{2}\int_0^t f'(X_s)d\langle X\rangle_s$$

for instance would exist, and be equal to the limit of the corresponding (symmetrical) Riemannian sums, for any $f \in C^2$. Some further generalizations to the above theorems are possible; for instance we could use the Tanaka formula instead of the Ito formula to define the integrals, in case X has an occupation density local time with respect to its quadratic variation function.

Now, by Result 2.6, almost all Brownian paths would be included into the above set of functions of quadratic variation with respect to any deterministically chosen sequence of subdivisions. For a given semimartingale, almost all paths would be in this set with respect to certain subdivisions (namely, for a subsequence of any deterministically chosen sequence of subdivisions.) In both cases, the quadratic variation function would almost surely equal the quadratic variation process of a semimartingale, and the obtained integral would almost surely equal the corresponding stochastic Ito integral. (Especially, it is seen that we can for each semimartingale X choose a "universal" set $\mathcal{N} \subset \Omega$ of zero probability, such that for every path $X(\omega)$, $\omega \in \mathcal{N}^c$, the Ito formula holds simultaneously for every C^2-function.)

So, this integral does give a pathwise version of stochastic integrals, and extends the Ito integral to a fairly large class of integrators. However it is obvious that the scope of integrands is far too limited to be useful for our purposes, i.e. to lead to a pathwise theory of stochastic differential equations. Also, the quadratic variation property of the paths X is not enough to give an integral which would be continuous with respect to the integrator and the integrand, thus Picard iterations would not work. Indeed, it seems very difficult to define a pathwise integration theory that would be good enough and general enough to enable us to do all the integrations that would be required to construct solutions to differential equations via Picard iterations. (Cf. Lyons [36] ,where a counter-example is given to show that the class of paths $\mathcal{C} \subseteq C([0,1], R)$ for which the Stratonovich or Ito integrals $\int_0^1 \eta \circ d\xi$ are defined for all pairs η, $\xi \in \mathcal{C}$, has Wiener measure zero.)

Accordingly, we will not try to define our theory of differential equations in the traditional way, as a differentiated form of integral equations. Instead, we will look at ways of defining solutions to differential equations directly, without an underlying complete integration theory. An approach of this kind was given for instance in Sussmann [53] and in Doss [14], but only in limited cases (either with scalar noise, or with commuting vector fields). In these cases, the situation is simple: the solution is a continuous functional of the path itself. More generally, this does not hold — but it is true that the solution is a functional of all the iterated path functionals, at least in some cases. (See for instance Yamato [63].)

From the remarks in the previous chapter, it is obvious that to get a continuous and consistent theory of differential equations and and their solution, the definition of noises X should be extended so as to also include information about their area integrals, etc. Our approach will be based on the well known idea of (formal) logarithmic series expansions, that give solutions of differential equations explicitely, in terms of iterated path integrals. We will have to extend the concept of noise to include iterated path integrals, but it will not be necessary to prove existence of more general integrals to construct our solution since the solution is constructed as a limit of operators rather than by the traditional method of Picard type iterations.

In the next chapter, we will go through some results which give an explicit expression for the solution of a differential equation in terms of iterated path integrals. The one given in Section 4.2 is possible to be generalized sufficiently for us to base our new theory of differential equations and their solutions on it. The new concepts of 'noise' and 'solution', and some convergence results for them are given in Chapter 5.

4 Exponential series expansions for solutions of differential equations

Our new approach to differential equations will be closely related to the well known idea of expressing the solutions of differential equations (ordinary or stochastic) in terms of iterated path integrals of the driving noise. Thus, for a differential equation with smooth or real analytical coefficients, the solution is expressed directly by a series expansion in terms of brackets or products of the vector fields in the differential equation, and iterated path integrals. To make the description of these kinds of results easier, we will use throughout this chapter and the rest of this thesis the following notations:

MULTI-INDICES:

A vector $J = (j_1, \ldots, j_k)$ with $j_i \in \{1 \ldots n\},\ i = 1 \ldots k$ is a multi-index of length k (over the base set $\{1 \ldots n\}$). We write $|J| = k$, for the length of the multi-index J; we also sometimes write $J \in \{1 \ldots n\}^k$.

BRACKETS OF VECTOR FIELDS:

For C^1-vector fields $Z_1,\ Z_2$ (on R^d, or more generally on a smooth manifold), $[\, Z_1, Z_2\,] = Z_1\, Z_2 - Z_2\, Z_1$ denotes the Lie bracket of Z_1 and Z_2. Here the product of two vector fields on R^d is defined as follows:

$$X\,Y = X \cdot \frac{\partial Y}{\partial x} = \sum_{i=1}^{d} X_i \frac{\partial}{\partial x_i} Y.$$

For a multi-index J, $|J| = k$, Z_J denotes the k-fold Lie bracket of vector fields:

$$Z_J = [\, Z_{j_1}, [\ldots, [\, Z_{j_{k-1}}, Z_{j_k} \,] \ldots]], \quad J = (j_1, \ldots, j_k).$$

We denote by $L(Z_1, Z_2, \ldots, Z_k)$ the Lie subalgebra of $C_\infty(R^d, R^d)$ generated by C^∞ vector fields $Z_1, \ldots, Z_k$.

ITERATED INTEGRALS:

For a multi-index $J = (j_1, \ldots j_k)$, $S_J(t)$ denotes the iterated ordinary or Stratonovich integral of a path X (of finite variation or of a semimartingale):

$$S_J(t) = \int_{\{0 \le t_1 < \cdots < t_k \le t\}} \circ dX_{j_1}(t_1) \cdots \circ dX_{j_k}(t_k)$$

THE CAMPBELL-HAUSDORFF-BAKER-DYNKIN FORMULA:

The Campbell-Hausdorff-Baker-Dynkin formula is the identity:

$$\exp(x)\,\exp(y) = \exp(H(x, y))$$

where $H(x, y)$ has a formal series expression in terms of iterated Lie brackets of the elements x and y. The first few terms are as follows:

$$H(x, y) = x + y + \frac{1}{2}[x, y] + \frac{1}{12}[[x, y], y] - \frac{1}{12}[[x, y], x] + \ldots$$

The Campbell-Hausdorff-Baker-Dynkin formula for n exponentials is correspondingly defined as the identity

$$\exp(x_1)\exp(x_2)\cdots\exp(x_n) = \exp(H(x_1, \ldots, x_n))$$

where again H has a formal series expansion in terms of the iterated brackets of the x_i. The exact definition is as follows:

$$H(x_1, \dots, x_n) = \sum_{k \geq 1} \frac{(-1)^{k-1}}{k} \sum_{P \in B_k} \frac{1}{\hat{P} P!} x_P$$

where

$$B_k = \left\{ (p_i^j)_{1 \leq i \leq n, 1 \leq j \leq k},\ p_i^j \in N,\ \forall 1 \leq j \leq k\ \sum_{i=1}^{n} p_i^j > 0 \right\}$$

and

$$\hat{P} = \sum_{i,j} p_i^j, \quad P! = \prod_{i,j} p_i^j!.$$

The Campbell-Hausdorff-Baker-Dynkin formula can of course also be seen as a formula of multiplication in a Lie group, in terms of the logarithmic coordinates. (See Hausner-Schwartz [22], Bourbaki [7], Strichartz [57], Jacobson [26].)

4.1 Series expansions for solutions of ODEs and SDEs

We will give here a review of some results connecting iterated path integrals and solutions of differential equations. In the deterministic case, where we look at solutions of ODE's, we will only mention the results by Strichartz [57], where the solution of a general ODE is expressed in terms of a solution to an autonomous ODE. Suppose that u is the solution to the ODE

$$u'(t) = A(t)\,(u(t)),$$

with initial value $u(0) = a$, where A is a smooth vector field. Then u can be written as

$$u(t) = \exp(z(t))(a),$$

where

$$z(t) = \sum_{m=1}^{\infty} \sum_{\sigma \in \sigma_m} \frac{(-1)^{e(\sigma)}}{m^2 \begin{pmatrix} m-1 \\ e(\sigma) \end{pmatrix}}$$

$$\cdot \int_{0<s_1<s_2<\cdots<s_m<t} [[\ldots[A(s_{\sigma(1)}), A(s_{\sigma(2)})], \ldots], A(s_{\sigma(m)})]\, ds.$$

Here we sum up over all m-permutations. For a permutation $\sigma \in \sigma_m$, $J \circ \sigma = (j_{\sigma(1)}, \ldots, j_{\sigma(m)})$ and $e(\sigma)$ denotes the number of errors in the ordering $\sigma(1), \ldots, \sigma(m)$: $e(\sigma)$ is the number of indices j: $j < m$ such that $\sigma(j) > \sigma(j+1)$. For a vector field z, we have denoted by $\exp(z)(a)$ the solution at time $s = 1$ to the time-homogeneous ODE:

$$v'(s) = z(v(s)), \quad v(0) = a.$$

This expression for the solution of a general ODE can be viewed as a generalized form of the Campbell-Hausdorff-Baker-Dynkin formula.

As for the stochastic case, Doss [14] and Sussmann [53] proved that the solution of an SDE driven with a Brownian motion is a continuous functional of the Brownian motion, if the noise is one-dimensional, or if the vector fields in the SDE commute. More generally, under various conditions on the Lie algebra generated by the vector fields in the SDE, Yamato [63] and others have proved that the solution is a functional of iterated integrals of the Brownian path. The simplest representation is by a stochastic Taylor series:

Let $X = (X_1, \ldots X_n)$ be an n-dimensional Brownian motion, and look at the

(Stratonovich) stochastic differential equation:

$$dY_t = \sum_{i=1}^{n} f_i(Y_t) \circ dX_i(t), \tag{5}$$

where f_i are smooth vector fields on R^d, with bounded derivatives of all orders. Then the solution of this SDE, with initial value $Y(0) = y_0$, can (at least locally) be expressed as a stochastic Taylor series:

$$Y_t = \sum_{k=1}^{p} \sum_{|J|=k} S_J(t)\,(f_{j_1} \cdots f_{j_k})(y_0) \; + \; R_p(t),$$

where for a multi-index $J = (j_1, \ldots, j_k)$,

- $S_J(t)$ is the iterated Stratonovich integral of $X_{j_1}, \ldots, X_{j_k}$,
- $(f_{j_1} \cdots f_{j_k})$ is defined as the product of vector fields, by:

$$(f_{j_1} \cdot f_{j_2} \cdots f_{j_k}) = \sum_{i=1}^{d} f_{j1}^{i} \frac{\partial}{\partial x^i} (f_{j_2} \cdots f_{j_k}),$$

  and so on.

(See, for instance, Azencott [1], Kloeden-Platen [27], Ben Arous [6].)

Another way of expressing the solution is as an exponential of a Lie series, that is, as the flow at time 1 of an ODE. (See for instance Kunita [31], [32], Castell [8], Ben Arous [6], and a slightly different way, Sussmann [54], [55].) According to Ben Arous [6], if $Lie(f_1, \ldots, f_n)$ is finite dimensional and if the vector fields $f_1, \ldots, f_n$ are complete, then the solution to the SDE at time t can, at least locally, be written in the following form:

$$Y(t) = \exp(\zeta(t))(y_0),$$

$$\zeta(t) = \sum_{k=1}^{\infty} \sum_{|J|=k} S_J(t)\, \beta_J,$$

where now β_J is a linear combination of k-fold brackets of the vector fields f_i; namely, β_J is the term which is k-homogeneous of degree 1 in the Campbell-Hausdorff series $H(f_{j_1}, \ldots f_{j_k})$. (The equation holds up to a P-almost surely strictly positive random time.) Alternatively, Castell [8] gives the solution to the SDE (5) in the following form:

$$Y(t) = \exp(\zeta(t))(y_0),$$

$$\zeta(t) = \sum_{k=1}^{\infty} \sum_{|J|=k} C^J(t)\, f_J$$

where

- f_J is the Lie bracket of the vector fields $f_{j_1}, \ldots, f_{j_k}$
- $C^J(t)$ is a linear combination of iterated Stratonovich integrals:

$$C^J(t) = \sum_{\sigma \in \sigma_m} \frac{(-1)^{e(\sigma)}}{m^2 \begin{pmatrix} m-1 \\ e(\sigma) \end{pmatrix}} S_{J \circ \sigma^{-1}}(t), \quad m = |J|.$$

  Here we sum up over all m-permutations. For a permutation $\sigma \in \sigma_m$, $J \circ \sigma = (j_{\sigma(1)}, \ldots, j_{\sigma(m)})$ and $e(\sigma)$ denotes the number of errors in the ordering $\sigma(1), \ldots, \sigma(m)$: $e(\sigma)$ is the number of indices j: $j < m$ such that $\sigma(j) > \sigma(j+1)$.

(This formulation of the result for Brownian motion noise follows fairly directly from the result by Strichartz in [57] for the ODE case, by discretization of the Brownian motion).

If the vector fields f_i in the SDE are simply assumed to be in C_∞ (and complete), then for instance the result of Castell gives:

$$Y(t) = \exp(\zeta^p(t))(y_0) \quad + \quad t^{p/2} R_p(t),$$

$$\zeta^p(t) = \sum_{k=1}^{p-1} \sum_{|J|=k} C^J(t)\, f_J$$

where R_p is bounded in probability when $t \to 0$. And, of course, if $Lie(f_1, \dots, f_n)$ is r-nilpotent, then

$$\zeta(t) = \sum_{k=1}^{r} \sum_{|J|=k} C^J(t)\, f_J.$$

By algebraic calculations it is possible to prove that the two above expressions for the solution are identical. The first few terms in these series are as follows: The series of Castell [8] gives

$$\zeta(t) = \sum_{i=1}^{n} f_i\,(X_i(t) - X_i(0)) \quad + \quad \sum_{i,j=1}^{n} \frac{1}{4}(S_{ij}(t) - S_{ji}(t))[\,f_i, f_j\,]$$

$$+ \quad \sum_{i,j,k=1}^{n} \frac{1}{9}[\,f_i, [\,f_j, f_k\,]\,]\,\{S_{ijk}(t) + S_{kji}(t)$$

$$- \frac{1}{2}\,(S_{ikj}(t) + S_{jki}(t) + S_{jik}(t) + S_{kij}(t)) \Big\} \quad + \dots$$

and the series by Ben Arous [6] gives:

$$\zeta(t) = \sum_{i=1}^{n} f_i\,(X_i(t) - X_i(0)) \quad + \quad \sum_{i,j=1}^{n} \frac{1}{2} S_{ij}(t)[\,f_i, f_j\,]$$

$$+ \quad \sum_{i,j,k=1}^{n} \frac{1}{6}\,([\,f_i, [\,f_j, f_k\,]\,] + [\,f_k, [\,f_j, f_i\,]\,])\, S_{ijk}(t) \quad + \dots$$

These expressions of the solution as an exponential of a Lie series will easily generalize into the case of SDEs on manifolds. Also, they could be considered more 'efficient' than the ones given by stochastic Taylor series. For instance, they lead to better approximations for the solution than the stochastic Taylor series if we replace the infinite series involved by a truncated one, in the sense that the approximation of the solution will preserve certain geometrical properties of the original solution. This is because the approximative solution is still obtained by following the integral curves of the vector fields f_i and their Lie brackets. So, if the true solution of the DE evolves in a certain submanifold then so does the approximation. (Cf. Sussmann [55].)

Obviously the way of representing the solution as an exponential of a Lie series can not be unique. Both the series given above involve all possible iterated integrals, but it is well known that some iterated integrals can be expressed as polynomials of other integrals. (See for instance Sussmann, [55]). Also, the convergence of the infinite sums in these representations is problematic; the proof that the sum converges when X is a Brownian motion uses the fact that a solution is known to exist and known to be analytically dependent on parameters in the vector fields. (Cf. Azencott [1].) Direct proof of the convergence, based on the sizes of the terms in the sum, seems to be very difficult.

It should be possible to choose a basis for the set of iterated integrals, and rewrite the series using only the ones in the basis, thus getting an "optimal" series expansion with no redundant integrals. For example, the second order terms in both of the series given above can be written as

$$\frac{1}{2} \sum_{i<j} [\, f_i, f_j \,] \, [\, X_i, X_j \,](t)$$

where $[X_i, X_j](t)$ denotes the skew symmetric area integral:

$$[X_i, X_j](t) = \int_0^t (X_i(s) - X_i(0))\, dX_j(s) - (X_j(s) - X_j(0))\, dX_i(s).$$

In the next section, we will look at a result by Lyons [37] which clarifies the relation between path integrals and solutions to differential equations, and what is more, can be extended to very general paths.

4.2 About the algebraic structure of iterated integrals

An alternative way of writing a formal logarithmic expansion for the solution of differential equations, using a minimal number of iterated integrals, is suggested by the results in Lyons [37]. There the theory is based on the algebraic structure of iterated path integrals of smooth paths, and can be formally extended in a generalized form to include non-smooth paths. If X is a smooth path in a vector space V, then the sequence of all iterated integrals of X can be represented as an element $g(X)$ in the tensor algebra $\mathcal{T}$ over V,

$$\mathcal{T} = \bigoplus_{i=0}^{\infty} \bigotimes_{1}^{i} V.$$

We write

$$g(X)|_s^t = X_0|_s^t + X_1|_s^t + \dots \quad \in \mathcal{T},$$

where

$$X_n|_s^t = \int \cdot\cdot \int_{s<t_1<\cdot\cdot<t_n<t} dX(t_1) \otimes \cdots \otimes dX(t_n)$$

gives all the n-th order iterated integrals. The $g(X)$ can now be interpreted as elements of a Lie group G, embedded multiplicatively in $\mathcal{T}$. (Cf. also Chen [10],

[11].) Now the *log* map is well defined :

$$\log : \mathcal{T} \to \mathcal{T}, \ \ \log(A) = (A-1) - (A-1)^2/2 + \dots$$

Especially $\beta(X) = \log g(X)$ is well defined, as an element of the free Lie algebra $\mathcal{G}$ generated by V in $\mathcal{T}$; the correspondence between $\beta(X)$ and $g(X)$ is one-to-one because G can be considered to be nilpotent. The multiplication rules for the group elements $g(X)|_s^t$, i.e. the relations between the various iterated path integrals of X, lead by the Campbell-Hausdorff-Baker-Dynkin formula to a set of consistency equations for the $\beta(X)$s:

$$\begin{cases} \beta_i^1(s,u) = \beta_i^1(s,t) + \beta_i^1(t,u) \ \forall i, \ 0 < s < t < u < 1 \\ \beta_{ij}^2(s,u) = \beta_{ij}^2(s,t) + \beta_{ij}^2(t,u) + \frac{1}{2}\left(\beta_i^1(s,t)\beta_j^1(t,u) - \beta_j^1(s,t)\beta_i^1(t,u)\right) \\ \qquad \forall i,j, \ 0 < s < t < u < 1 \\ \qquad \vdots \end{cases} \tag{6}$$

The expressions in these equations get increasingly more complicated, but the expression for $\beta^k(s,u) - \beta^k(s,t) - \beta^k(t,u)$ always involves only sums and Lie brackets of lower order terms $\beta^m(s,t), \beta^m(t,u), \ \ m = 1 \dots k-1$. This representation of the path X via the $g(X)$ and the $\beta(X)$ can formally be used to express solutions of differential equations as functionals of the iterated integrals: In a differential equation

$$dY = f(Y)dX$$

on a manifold M, f can be considered as a linear map: $V \to$ vector fields on M. This map can now be extended into a Lie algebra map: $f^* : \mathcal{G} \to$ vector fields on M, and the solution for the differential equation can be written using the iterated integrals $\beta(X) \in \mathcal{G}$ as:

$$\textit{flow of solution } = \exp f^*(\beta(X)).$$

β is an element of the Lie algebra $\mathcal{G}$ and its components are formal brackets. Let us now choose $V = R^n$ and fix a basis for $\mathcal{G}$, formally denoted by $\{\mathcal{J}(k), k = 1, 2, \ldots\}$, where $\mathcal{J}(k) \subseteq \{1, \ldots, n\}^k$ gives the k-th order brackets in the basis by identifying them with multi-indices. Then we can formally write the flow of the solution to the differential equation

$$dY_t = \sum_{i=1}^{n} f_i(Y_t) dX_i(t)$$

over time [0,1] as $\exp(z(0,1))$ where

$$z(0,1) = \left(\sum_{n=1}^{\infty} \sum_{J \in \mathcal{J}(n)} \beta_J^n(0,1)\, C_J^{(n)} \right),$$

and the C_J^k are uniquely defined vector fields, and we sum up over the chosen basis. For instance, we could choose the Hall basis which would for example give the basis for L(x,y) as $\{x, y, [x,y], [x,[x,y]], [y,[x,y]], \ldots\}$. See Sussmann [55], Bourbaki [7]. An alternative choice could be e.g. the Lyndon basis (see Melancon-Reutenauer [45], Devlin [13]).

Since we will come accross these notations later on, we will state them as a definition.

Definition 4.1 *We agree on the following notation:*

- $\{\mathcal{J}(k), k = 1, 2, \ldots\}$ *represents the Hall basis (or any other chosen basis) of the Lie algebra $\mathcal{G}$; multi-indices in $\mathcal{J}(k) \subseteq \{1, \ldots, n\}^k$ identify the elements of the basis which are of k-th degree, i.e. formed from k-fold formal brackets;*

- $C_J^{(n)}$ *are vector fields, uniquely defined from the identity above after the basis above has been fixed. For a multi-index* $J \in \mathcal{J}(k)$, $C_J^{(k)}$ *consists of k-th order brackets of the vector fields* f_i.

If we choose the Hall basis, then

$$C_i^{(1)} = f_i, \quad C_{ij}^{(2)} = [\, f_i, f_j \,], \quad \ldots$$

Now, since we are dealing with elements in Lie groups and algebras, these results for smooth paths X and their iterated path integrals can be generalised using this algebraic structure, to any collection β for which the same algebraic relations hold. Thus for any given path, the representation above is still valid in some sense if $\beta^1 = X$ and $\beta^2, \beta^3 \ldots$ are two-parameter functions for which the consistency equations (6) above hold. The $\beta^2, \beta^3, \ldots$ act then as generalized path integrals of the path X.

Definition 4.2 *(Modified from Lyons [37])*
An enhanced path of degree k is a collection of two-parameter functions $(\beta^1(t,s), \ldots \beta^k(t,s))$ *satisfying the algebraic relations (6). If* $\beta^1(s,t) = X(t) - X(s)$ *then* β *is said to be an enhancement of* X*, to degree k.*

Lyons [37] now proceeds to give a very useful result stating that if a sufficient number of initial generalized iterated integrals exist, then all the other ones can be constructed from them. The required degree k of the initial enhancement of X depends on the continuity of the functions $\beta^1 = X,\ \beta^2, \ldots, \beta^k$.

Result 4.3 *Let* $(\beta^1, \ldots, \beta^n)$ *be an enhanced path of degree n and suppose that for* *all* $k = 1, \ldots, n$*,*

$$|\beta^k(s,t)| \leq C\,|t-s|^{\alpha k}, \quad \forall\, 0 < s < t < 1,$$

where $\alpha(n+1) > 1$*. Then* β *can be extended into an enhanced path of any degree, and all* β^k*,* $k = 1, 2, ..$ *will have the continuity property given above.*

We are now going to use these results about the algebraic structure of the iterated integrals and the way they can be extended to more general paths, to justify our new definitions of noise paths and solutions to differential equations.

5 Construction of a new DE theory

In the previous chapter we referred to several results which give locally an explicit expression for the flow of an ordinary differential equation (or a stochastic differential equation)

$$dY_t = \sum_{i=1}^{n} f_i(Y_t) dX_i(t), \tag{7}$$

in some special cases — e.g. for DE with C^∞ coefficients f_i, when $Lie(f_1 \dots f_n)$ is finite dimensional, for some well known noises X. (That is, X is smooth, or a Brownian motion in a given filtered probability space — in which case the obtained result holds almost surely and perhaps only over a random time interval.) The flow of the solution can according to these results be at least locally written using a logarithmic series expansion, i.e. as a flow at time 1 of an ODE along a time invariant vector field. This vector field is defined by an infinite series built from iterated path integrals of the noise function X, and products or brackets of the vector fields f_i. More generally, if f_i are general smooth vector fields, then this series gives an asymptotic expansion for the solution near time $t = 0$. Of course, for the noises considered in these results iterated integrals exist and are well behaved.

However, the discussion in Section 4.2 suggests that even for some more general paths X, it might still be possible to formally write a "solution" to the differential equation (7) as the flow given by

$$\theta[s,t] = \exp(z[s,t]),$$

$$z[s,t] = \sum_{n=1}^{\infty} \sum_{J \in \mathcal{J}(n)} \beta_J^n(s,t)\, C_J^{(n)}$$

where

- β_J are 2-parameter functions, playing the role of generalized iterated path integrals of X;
- $C_J^{(n)}$ are well defined vector fields, and consist of n-fold brackets of the vector fields f_i;
- the series is summed over sets of indices $\{\mathcal{J}(n),\ n = 1, 2, \ldots\}$; this relates to the choice of a basis for the β_J interpreted as formal brackets.

(See Definition 4.1 in last chapter.) This remains a purely formal notation, of course, until something can be said about convergence of the series, and so on.

Here, we will use an approach suggested by this formal series to attempt to construct a solution to the differential equation (7) with a more general driving path X and with less stringent conditions on the vector fields f_i. Informally speaking, this is what we will do: Suppose that the r first brackets of the vector fields f_i exist, and suppose that for the path X we can supply "iterated path integrals" β_J^k, up to r-th order, however these may be defined. Then, we could still define for any $[s,t]$ a time homogeneous vector field by the truncated series

$$\hat{z}[s,t] = \sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta_J^k(s,t)\, C_J^{(k)}$$

and the corresponding flow over the time interval $[s,t]$:

$$\hat{\theta}[s,t] = \exp(\hat{z}[s,t]).$$

Now, if the same interrelations hold for the β_J^ks, as hold for the iterated path integrals of smooth paths, then it might be expected that $\hat{\theta}[s,t]$ would be a good approximation to a "solution" of the differential equation (7), over short intervals $[s,t]$ at least. So, for a fixed time interval [0,1], we will construct a series of flows $\hat{\theta}^{(n)}[0,1]$ as follows: We will choose a sequence of finite subdivisions of [0,1] with $[0,1] = \bigcup I_k^{(n)}$ such that $\delta(n) \to 0$ as $n \to \infty$ for $\delta(n) =$ maximal length of intervals $I_k^{(n)}$. Over each of these subintervals $I_k^{(n)}$ we will then use the truncated series to define $\hat{\theta}^{(n)}(I_k^{(n)}) = \exp(\hat{z}(I_k^{(n)}))$. And then, over the whole interval, we take

$$\hat{\theta}^{(n)}[0,1] = \prod_k \hat{\theta}^{(n)}(I_k^{(n)}).$$

If now we could prove that $\lim_{n\to\infty} \hat{\theta}^{(n)}[0,1]$ exists then we could call it "a solution to (7) at time 1". To be able to prove convergence, we will of course also need to control the sizes of the $\beta_J^k(s,t)$s. More precise definitions are given in the following section.

5.1 Notations and assumptions

The set-up we have in mind is as follows:

THE PATH X

Assume from the path X the following:

— $X = (X_1, \dots X_n)$ where $X_i : [0,1] \to R$ are Holder(α)-continuous functions, $i = 1 \dots n$, for some $\alpha \geq 0$.

— In addition to X itself, we are given its "iterated integrals" up to r-th order, that is, real-valued 2-parameter functions $\beta^2, \dots, \beta^r$,

$$\beta^k = \left\{ \beta_J^k(s,t) \mid J \in \mathcal{J}(k) \subseteq \{1, \dots, n\}^k,\ [s,t] \subseteq [0,1] \right\}$$

such that when we define

$$\beta_i^1(s,t) = X_i(t) - X_i(s),\ i = 1 \ldots n,$$

then $\left(\beta^1, \ldots, \beta^r\right)$ forms an (α, r)-system (see the following definition). We will denote this system by $\beta = \left(\beta^1, \ldots, \beta^r\right)$ and say that the path X has been extended into the generalized path β. (Cf Lyons [37].)

Definition 5.1 *A collection of 2-parameter functions* $\left(\beta^1, \ldots, \beta^r\right)$*, where*

$$\beta^k = \left\{\beta_J^k(s,t) \mid J \in \mathcal{J}(k) \subseteq \{1, \ldots, n\}^k,\ [s,t] \subseteq [0,1]\right\}$$

forms an (α, r)*-system of generalized iterated integrals over [0,1], if:*

1. *The consistency relations of Section 4.2 hold for* $\beta^1, \ldots, \beta^r$*:*

$$\begin{cases} \beta_i^1(s,u) & = \beta_i^1(s,t) + \beta_i^1(t,u)\ \forall i,\ 0 < s < t < u < 1 \\ \beta_{ij}^2(s,u) & = \beta_{ij}^2(s,t) + \beta_{ij}^2(t,u) + \frac{1}{2}\left(\beta_i^1(s,t)\beta_j^1(t,u) - \beta_j^1(s,t)\beta_i^1(t,u)\right) \\ & \quad \forall i,j,\ 0 < s < t < u < 1 \\ & \vdots \\ \beta_{j_1 \ldots j_r}^r(s,u) & = \ldots \end{cases}$$

2. *The system* $\left(\beta^1, \ldots, \beta^r\right)$ *is* α*-continuous:* $\forall\ [s,t] \subseteq [0,1]$, $\forall\ k = 1, \ldots, r$, $J \in \mathcal{J}(k) \subseteq \{1, \ldots, n\}^k$,

$$\left|\beta_J^k(s,t)\right| \leq C\,|t-s|^{k\alpha}.$$

Remarks:

— There are generally more than one way of choosing $\beta^2, \ldots, \beta^r$ for a given path

X, and thus extending X into a generalized path β. However, the continuity and consistency requirements do limit the choices to some extent. Thus for instance if $\alpha > 1/k$, then β^k is uniquely defined once $\beta^1, \ldots, \beta^{k-1}$ are fixed. This follows from the fact that once $\beta^1, \ldots, \beta^{k-1}$ are given, it follows from the consistency conditions that β^k is uniquely defined, up to an additive term: If β^k and $\hat{\beta}^k$ both satisfy the consistency conditions with the same given $\beta^1, \ldots, \beta^{k-1}$ then

$$\beta^k(s,t) - \hat{\beta}^k(s,t) = \gamma(s,t)$$

with

$$\gamma(s,t) + \gamma(t,u) = \gamma(s,u).$$

But γ also has to be αk-continuous, therefore it must be zero when $\alpha > 1/k$.

— We will sometimes denote $\beta^2_{ij} = \frac{1}{2}[X_i, X_j]$. The consistency relation

$$\beta^2_{ij}(s,u) = \beta^2_{ij}(s,t) + \beta^2_{ij}(t,u) + \frac{1}{2}\left(\beta^1_i(s,t)\beta^1_j(t,u) - \beta^1_j(s,t)\beta^1_i(t,u)\right)$$

suggests that β^2_{ij} is a generalised area of the curve (X_i, X_j). (See Definition 7.2 in Section 7.1.)

THE DIFFERENTIAL EQUATION

We will look at a differential equation with respect to the path X, written as:

$$dY_t = \sum_{i=1}^{n} f_i(Y_t)dX^i_t \tag{8}$$

where $t \in [0,1]$, $Y_t \in R^d$, and f_i are vector fields on R^d. Note that (8) is so far only a purely formal notation; it cannot here be interpreted as a differentiated form of a corresponding integral equation, since X is an arbitrary continuous function and therefore, an integration theory with respect to it obviously does not necessarily exist.

We will assume from the vector fields f_i so far only that all brackets of them up to the r-th level exist, and are Lipschitz continuous. By brackets up to r-th level, we mean the vector fields

$$[f_i, f_j] \;\; \forall\, i, j$$

$$[f_i, [f_j, f_k]] \;\; \forall\;\; i, j, k$$

$$[f_{i_1}, [f_{i_2}, [\ldots, [f_{i_{r-1}}, f_{i_r}] \ldots]]] \;\; \forall\, i_1, \ldots, i_r.$$

CONSTRUCTION OF SOLUTION

Take $(\tau_n, n \in N)$ to be the sequence of dyadic subdivisions of interval [0,1]; i.e. $\tau_n = \{k2^{-n},\; k = 0 \ldots 2^n\}$, which gives

$$[0,1] = \bigcup_{k=1}^{2^n} \left[S_{k-1}^{(n)}, S_k^{(n)}\right]$$

where $\left[S_{k-1}^{(n)}, S_k^{(n)}\right] = [(k-1)2^{-n}, k2^{-n}]$. We define now a sequence of operators $\theta^{(n)}[0,1] :\; R^d \to R^d,\; n \in N$ as follows: For a fixed n, for all subintervals [s,t] of (τ_n)-subdivision over [0,1], we will define

$$\theta^{(n)}[s,t] = \exp(z[s,t])$$

where

$$z[s,t] = \sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta_J^k(s,t)\, C_J^{(k)}.$$

Here,

- β_J^k are the "iterated integrals" of X in the chosen path generalization $\beta = (\beta^1, \ldots, \beta^r)$,
- the vector fields $C_J^{(k)}$ are as defined in Definition 4.1,
- summing up is over the Hall basis as defined in Definition 4.1,

and for a vector field z on R^d, $\exp(z)$ is defined as the flow at time 1 of the (time-homogeneous) ODE with respect to z: $\exp(z)(x) = y(1)$ for y with $\dot{y} = z(y)$, $y(0) = x$.

This gives, for instance, for $r = 1$:

$$z[s,t] = \sum_{i=1}^{n} (X_i(t) - X_i(s))\, f_i,$$

and for $r = 2$,

$$z[s,t] = \sum_{i=1}^{n} (X_i(t) - X_i(s))\, f_i \ + \ \frac{1}{2} \sum_{i<j} [X_i, X_j](s,t)\, [\, f_i, f_j\,].$$

The operators $\theta^{(n)}[s,t]$ are well defined, since by assumptions on the f_is, the vector fields $z[s,t]$ are Lipschitz continuous.

Define, finally, $\theta^{(n)}$ over the whole interval [0,1] by taking

$$\theta^{(n)}[0,1] = \theta^{(n)}[S^{(n)}_{2^n-1}, S^{(n)}_{2^n}] \cdots \theta^{(n)}[S^{(n)}_0, S^{(n)}_1].$$

Remarks:

- Each $\theta^{(n)}[s,t]$ is a flow of homeomorphisms $R^d \to R^d$, so $\theta^{(n)}[0,1]$ is a homeomorphism $R^d \to R^d$ for each n.

– If the f_i are linear vector fields, then $\theta^{(n)}[s,t]$ and $\theta^{(n)}[0,1]$ are linear operators $R^d \to R^d$. Even in the general case we can always interpret them as linear operators, via action on functions in some function space F:

$$\theta^{(n)}[s,t]: F \to F, \quad f \to \theta^{(n)}[s,t](f)$$

where

$$\theta^{(n)}[s,t](f)(x) = f\left(\theta^{(n)}[s,t](x)\right), \quad \forall\, x \in R^d.$$

– Finally, we can interpret the operators $\theta^{(n)}[s,t]$ and $\theta^{(n)}[0,1]$ as elements of a group of operators on R^d, namely the group of homeomorphisms.

Now, assuming that we can interpret the operators $\theta^{(n)}[0,1]$ for all n as elements of some suitable metric space, we can define the solution as follows:

Definition 5.2 *If the sequence $\left(\theta^{(n)}[0,1]\right)_{n\in N}$ converges, i.e. if the limit*

$$\theta^{(\infty)}[0,1] = \lim_{n\to\infty} \theta^{(n)}[0,1]$$

exists, then we will call $\theta^{(\infty)}[0,1]$ a solution to the differential equation (8) at time 1, with respect to the generalized path β. (That is, $\theta^{(\infty)}[0,1]$ is the flow of the solution at time 1; and $Y(1) = \theta^{(\infty)}[0,1](y_0)$ gives a solution at time 1 of the differential equation with initial value $y_0 \in R^d$.)

Note that here we look at the limit $\theta^{(\infty)}[0,1]$ which only gives the solution at time 1. We could also look at convergence towards $\theta^{(\infty)}$ defined as a flow of the solution over the whole interval [0,1].

In the following sections we will look more closely at the sequence $\theta^{(n)}[0,1]$, and its convergence. That is, we will look at what happens under various assumption on the values of α and r, and on the vector fields f_i. The next section is mainly technical; in Section 5.3 we will look at a special case, namely linear differential equations, and in Section 5.4 we look briefly at the general case. In Section 5.5 we will investigate the convergence of our sequence of operators towards a given flow.

5.2 The sequence $\theta^{(n)}$ and its convergence

In this section we give some general results about the sequence $\theta^{(n)}$ and establish some new technical notations, also we discuss the various options of choosing a metric for it.

Let $(\tau_n, n \in N)$ be the sequence of dyadic subdivisions of [0,1]. We will denote the division points of τ_n by $S_i^{(n)}$, $i = 1 \dots 2^n$; thus

$$[S_{i-1}^{(n)}, S_i^{(n)}] = [(i-1)2^{-n}, i2^{-n}]$$

is the i-th subinterval of the τ_n-subdivision. For simplicity, we will sometimes use the notation "$\Delta \in \tau_n$" to mean that the interval Δ is a subinterval of the τ_n-subdivision, i.e. that it is of the form $\Delta = [S_{i-1}^{(n)}, S_i^{(n)}]$ for some $i \in \{1, \dots, 2^n\}$. As defined in the previous section, each $\theta^{(n)}$ corresponds to a fixed subdivision τ_n. The operator $\theta^{(n)}$ is defined by giving first its value over all subintervals $[S_{i-1}^{(n)}, S_i^{(n)}]$

of the τ_n-subdivision as:

$$\theta^{(n)}[S_{i-1}^{(n)}, S_i^{(n)}] = \exp(z(S_{i-1}^{(n)}, S_i^{(n)})), \quad i = 1 \dots 2^n$$

where for $[s,t] \subseteq [0,1]$,

$$z(s,t) = \sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta_J^k(s,t) C_J,$$

and by then defining

$$\theta^{(n)}[0,1] = \theta_{2^n}^{(n)} \cdot \theta_{2^n-1}^{(n)} \cdots \theta_1^{(n)},$$

or more generally:

$$\theta^{(n)}[S_{i-1}^{(n)}, S_j^{(n)}] = \theta_j^{(n)} \cdot \theta_{j-1}^{(n)} \cdots \theta_{i+1}^{(n)} \cdot \theta_i^{(n)}$$

for arbitrary division points $S_i^{(n)}, S_j^{(n)}$ of τ_n. Here we have used notation

$$\theta_i^{(n)} = \theta^{(n)}[S_{i-1}^{(n)}, S_i^{(n)}].$$

To look at the convergence of the sequence $\theta^{(n)}[0,1]$ of operators, we have to establish which metric we will use. If the vector fields f_i are C^∞-continuous, and if $Lie(f_1, \dots f_n)$ is finite dimensional, then using the multiplicative notation and the Campbell-Hausdorff-Baker-Dynkin formula, we could look at the convergence of the operators $\theta^{(n)}[0,1]$ in logarithmic coordinates, with respect to a metric for smooth vector fields on R^d. That is, we can write

$$\theta^{(n)}[0,1] = \exp(a^{(n)})$$

and look at the convergence of the sequence $(a^{(n)},\ n \in N)$ of smooth vector fields on R^d.

More generally, we can choose to interpret the $\theta^{(n)}[0,1]$s as elements in a Banach space of linear operators, with a norm $\|\cdot\|$ (through action on smooth functions if necessary). To prove convergence we prove that $\theta^{(n)}[0,1]$ is a Cauchy sequence of linear operators, for which it suffices to prove that

$$\sum_{k=1}^{\infty} \| \Delta^{(k+n)} \| \leq C(n)$$

where $C(n) \to 0$ as $n \to \infty$, for

$$\Delta^{(k)} = \theta^{(k)}[0,1] - \theta^{(k+1)}[0,1]$$

We will start by comparing $\theta^{(n)}$ and $\theta^{(n+1)}$. Over an interval $[S_{i-1}^{(n)}, S_i^{(n)}]$ of the subdivision τ_n,

$$\theta_i^{(n)} = \theta^{(n)}[S_{i-1}^{(n)}, S_i^{(n)}] = \exp(z(S_{i-1}^{(n)}, S_i^{(n)})).$$

The value of $\theta^{(n+1)}$ over this interval is obtained by dividing $[S_{i-1}^{(n)}, S_i^{(n)}]$ into two τ_{n+1}-intervals:

$$[S_{i-1}^{(n)}, S_i^{(n)}] = [S_{2i-2}^{(n+1)}, S_{2i-1}^{(n+1)}] \ \cup\ [S_{2i-1}^{(n+1)}, S_{2i}^{(n+1)}]$$

and defining then

$$\theta^{(n+1)}[S_{i-1}^{(n)}, S_i^{(n)}] = \theta^{(n+1)}[S_{2i-2}^{(n+1)}, S_{2i-1}^{(n+1)}] \cdot \theta^{(n+1)}[S_{2i-1}^{(n+1)}, S_{2i}^{(n+1)}]$$

$$= \exp(z(S_{2i-2}^{(n+1)}, S_{2i-1}^{(n+1)}))\exp(z(S_{2i-1}^{(n+1)}, S_{2i}^{(n+1)})).$$

We will denote

$$\theta_{(i)}^{(n+1)} = \theta^{(n+1)}[S_{i-1}^{(n)}, S_i^{(n)}] = \theta_{2i-1}^{(n+1)} \cdot \theta_{2i}^{(n+1)}$$

which is the value of $\theta^{(n+1)}$ over the τ_n-interval $[S_{i-1}^{(n)}, S_i^{(n)}]$. Then we can write

$$\theta^{(n)}[0,1] = \theta_{2^n}^{(n)} \cdots \theta_1^{(n)},$$

$$\theta^{(n+1)}[0,1] = \theta_{(2^n)}^{(n+1)} \cdots \theta_{(1)}^{(n+1)}.$$

The crucial fact in establishing the convergence is that these one step differences are small, because of the way that the operators $\theta^{(n)}$ are constructed: $\theta_i^{(n+1)}$ and $\theta_{(i)}^{(n+1)}$ agree "up to r-th order terms"; what exactly is meant by this will become clear later on.

The sequence in multiplicative notation:

We will first mention one special case, where the multiplicative notation proves very useful.

Lemma 5.3 *For fixed n, define*

$$\varepsilon_i^{(n)} = \theta_i^{(n)} \left(\theta_{(i)}^{(n+1)}\right)^{-1}, \quad i = 1 \dots 2^n.$$

Assume that for each n, all the operators $\varepsilon_i^{(n)}$ commute with all the operators $\theta_j^{(n)}$ (i.e. $\theta_j^{(n)}\varepsilon_i^{(n)} = \varepsilon_i^{(n)}\theta_j^{(n)}$ $\forall\ i,j$.) Then

$$\theta^{(n)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} = \overrightarrow{\prod_{1\le k\le 2^n}} \varepsilon_k^{(n)},$$

$$\theta^{(0)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} = \overleftarrow{\prod_{0\le m\le n}} \overrightarrow{\prod_{1\le k\le 2^m}} \varepsilon_k^{(m)}.$$

Proof.

$$\begin{aligned}
& \theta^{(n)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} \\
= \; & \theta^{(n)}_{2^n}\,\theta^{(n)}_{2^n-1}\cdots\theta^{(n)}_2\,\theta^{(n)}_1\,\left(\theta^{(n+1)}_{(1)}\right)^{-1}\cdots\left(\theta^{(n+1)}_{(2^n)}\right)^{-1} \\
= \; & \theta^{(n)}_{2^n}\cdots\theta^{(n)}_2\,\varepsilon^{(n)}_1\,\left(\theta^{(n+1)}_{(2)}\right)^{-1}\cdots\left(\theta^{(n+1)}_{(2^n)}\right)^{-1} \\
= \; & \varepsilon^{(n)}_1\,\theta^{(n)}_{2^n}\cdots\theta^{(n)}_2\,\left(\theta^{(n+1)}_{(2)}\right)^{-1}\cdots\left(\theta^{(n+1)}_{(2^n)}\right)^{-1} \\
\vdots \; & \\
= \; & \varepsilon^{(n)}_1\,\varepsilon^{(n)}_2\cdots\varepsilon^{(n)}_{2^n} \\
= \; & \overrightarrow{\prod_{1\le k\le 2^n}} \varepsilon_k^{(n)}
\end{aligned}$$

and repeating this we get

$$\begin{aligned}
\theta^{(n-1)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} &= \theta^{(n-1)}[0,1]\left(\theta^{(n)}[0,1]\right)^{-1}\,\theta^{(n)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} \\
&= \overrightarrow{\prod_{1\le k\le 2^{n-1}}} \varepsilon_k^{(n-1)} \overrightarrow{\prod_{1\le k\le 2^n}} \varepsilon_k^{(n)}
\end{aligned}$$

and finally

$$\theta^{(0)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} = \overleftarrow{\prod_{0\le m\le n}} \overrightarrow{\prod_{1\le k\le 2^m}} \varepsilon_k^{(m)}.$$

□

Note that we have here denoted the operator product $\varepsilon_1\cdots\varepsilon_n$ by $\overrightarrow{\prod}_{i=1..n}\varepsilon_i$. By the definitions of the $\theta^{(n)}$, all the $\varepsilon_i^{(n)}$ are of the form

$$\varepsilon_i^{(n)} = \exp(z(\triangle_1\cup\triangle_2))\,\exp(-z(\triangle_1))\exp(-z(\triangle_2)),$$

where $\triangle_1 \cup \triangle_2 \in \tau_n$, and $\triangle_1,\ \triangle_2\ \in \tau_{n+1}$.

This case includes the "structural" differential equations satisfied by the iterated path integrals $\beta(s,t)$. For instance, if we write out the differential equation satisfied by β^{r+1}, then the ε_is commute with the $\theta_j^{(n)}$s since the $\theta_j^{(n)}$ are defined using an r-th order method. This is obvious from the algebraic structures of the β^ks. Also, the vector fields in this differential equation are nilpotent, so that the Campbell-Hausdorff-Baker-Dynkin formula can be applied and convergence can be proved in logarithmic coordinates — this is how Result 4.3 is proved in Lyons [37]. (Cf. also Gaveau [19], Gaveau-Vauthier [20], Gaveau-Greimer-Vauthier [21], Chaleyat-Maurel and LeGall [9] about group structure of Brownian motion and its area, interpreted as diffusions on Heisenberg groups.)

In the general case, the multiplicative notation gives

$$
\begin{aligned}
& \theta^{(n)}[0,1]\left(\theta^{(n+1)}[0,1]\right)^{-1} \\
= \;& \theta_{2^n}^{(n)}\,\theta_{2^n-1}^{(n)}\cdots\theta_2^{(n)}\,\theta_1^{(n)}\left(\theta_{(1)}^{(n+1)}\right)^{-1}\cdots\left(\theta_{(2^n)}^{(n+1)}\right)^{-1} \\
= \;& \theta_{2^n}^{(n)}\cdots\theta_2^{(n)}\,\varepsilon_1^{(n)}\left(\theta_{(2)}^{(n+1)}\right)^{-1}\cdots\left(\theta_{(2^n)}^{(n+1)}\right)^{-1} \\
= \;& \varepsilon_{2^n}^{(n)}\cdot Ad\,\theta_{(2^n)}^{(n+1)}(\varepsilon_{2^n-1}^{(n)}\cdot Ad\,\theta_{(2^n-1)}^{(n+1)}(\cdots Ad\,\theta_2^{(n+1)}(\varepsilon_1^{(n)}))\cdots)
\end{aligned}
$$

where $Ad\,x(y) = x\,y\,x^{-1}$.

In additive notation:

To compare $\theta^{(n)}$, $\theta^{(n+1)}$ interpreted as linear operators, we can use the following lemma.

Lemma 5.4 *Assume that*

$$\theta = \theta_N \cdots \theta_1, \quad \hat{\theta} = \hat{\theta_N} \cdots \hat{\theta}_1$$

where θ_i, $\hat{\theta}_i$ are linear operators. Then

$$\hat{\theta} - \theta = \sum_{i=1}^{N} \hat{\theta}_N \cdots \hat{\theta}_{i+1}\, e_i\, \theta_{i-1} \ldots \theta_1$$

where $e_i = \hat{\theta}_i - \theta_i$.

Proof. Denote

$$r(m) = \hat{\theta}_m \cdots \hat{\theta}_1 - \theta_m \cdots \theta_1,$$

for $1 \leq m \leq N$. Then the $r(j)$ obviously satisfy the equation:

$$r(j+1) = \varepsilon_{j+1} + \hat{\theta}_{j+1} r(j)$$

where

$$\varepsilon_{j+1} = e_{j+1}(\theta_j \cdots \theta_1),$$

with initial values $r_1 = \varepsilon_1 = e_1$. By recursion, it follows that for all $1 \leq m \leq N$,

$$\begin{aligned} r(m) &= e_m \theta_{m-1} \theta_{m-2} \cdots \theta_1 \\ &+ \hat{\theta}_m e_{m-1} \theta_{m-2} \cdots \theta_1 \\ &+ \cdots \\ &+ \hat{\theta}_m \hat{\theta}_{m-1} \cdots \hat{\theta}_2 e_1. \end{aligned}$$

□

This gives

$$\theta^{(n)}[0,1] - \theta^{(n+1)}[0,1] = \sum_{i=1}^{2^n} \theta^{(n)}[S_i^{(n)}, 1]\, e_i^{(n)}\, \theta^{(n+1)}[0, S_{i-1}^{(n)}] \tag{9}$$

where

$$e_i^{(n)} = \theta^{(n)}[S_{i-1}^{(n)}, S_i^{(n)}] - \theta^{(n+1)}[S_{i-1}^{(n)}, S_i^{(n)}]$$

is the 1-step difference of $\theta^{(n)}$ and $\theta^{(n+1)}$, over the i-th interval of the τ_n-subdivision, and $\theta^{(n)}[S_i^{(n)}, 1]$ and $\theta^{(n+1)}[0, S_{i-1}^{(n)}]$ act as "error propagation" terms.

So, we need to look at:

- The one-step errors $e_i^{(n)}$, $\forall\, i, n$, which can be written as

  $$e_i^{(n)} = \exp(z(\Delta_1 \cup \Delta_2)) - \exp(z(\Delta_1))\exp(z(\Delta_2))$$

  where $\Delta_1 \cup \Delta_2 = [S_{i-1}^{(n)}, S_i^{(n)}] \in \tau_n$, $\Delta_1, \Delta_2 \in \tau_{n+1}$ (i.e. $\Delta_1 \cup \Delta_2$ is an interval of length 2^{-n} and Δ_1, Δ_2 are intervals of length $2^{-(n+1)}$).

- Error propagation terms, in general of the form

  $$\theta^{(n)}[S_{i-1}^{(n)}, S_j^{(n)}], \quad 1 \leq i \leq j \leq 2^n$$

  which we can write as

  $$\exp(z(\Delta_i)) \cdots \exp(z(\Delta_j))$$

  for $\Delta_k = [S_{k-1}^{(n)}, S_k^{(n)}] \in \tau_n$ (interval of length 2^{-n}).

If we now can prove that

$$\sup_{1 \leq i \leq 2^n} \|e_i^{(n)}\| \leq C|\delta_n|^A \quad \forall n,$$

$$\sup_{1 \leq i < j \leq 2^n} \|\,\theta^{(n)}[S_{i-1}^{(n)}, S_j^{(n)}]\,\| \leq C|\delta_n|^B \quad \forall n$$

where $\delta_n = 2^{-n}$ is the length of subintervals in the τ_n-subdivision, then

$$\| \triangle^{(n)} \| = \|\theta^{(n)}[0,1] - \theta^{(n+1)}[0,1]\|$$

$$\leq \sum_{j=1}^{2^n} \|\theta^{(n)}[S_{j-1}^{(n)}, 1]\| \cdot \|e_j^{(n)}\| \cdot \|\theta^{(n+1)}[0, S_j^{(n)}]\|$$

$$\leq 2^n \sup_{1\leq j\leq 2^n} \|e_j^{(n)}\| \cdot \sup_{1\leq j\leq 2^n} \|\theta^{(n)}[S_{j-1}^{(n)}, 1]\| \cdot \|\theta^{(n+1)}[0, S_j^{(n)}]\|$$

$$\leq C\, 2^{-n(A+2B-1)}.$$

So in this case,

$$\sum_{k=1}^{\infty} \| \triangle^{(n+k)} \| \leq C\, 2^{-n(A+2B-1)} \sum_{k=1}^{\infty} \left(2^{(A+2B-1)}\right)^{-k}.$$

Therefore for $\theta^{(n)}[0,1]$ to be a Cauchy sequence a sufficient condition will be that

$$A + 2B - 1 > 0.$$

We will write this observation as a Lemma for future reference.

Lemma 5.5 *Assume that*

$$\sup_{1\leq i\leq 2^n} \|e_i^{(n)}\| \leq C|\delta_n|^A \quad \forall n,$$

$$\sup_{1\leq i<j\leq 2^n} \| \theta^{(n)}[S_{i-1}^{(n)}, S_j^{(n)}] \| \leq C|\delta_n|^B \quad \forall n.$$

Then $\theta^{(n)}(0,1)$ *is a Cauchy sequence, and thus converges, if* $A + 2B - 1 > 0$.

We will now establish some notations that will clarify the situation. Note that each $z(s,t)$ is of the form

$$z(s,t) = \sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta_J^k(s,t)\, C_J^{(k)}$$

where the $C_J^{(k)}$ are fixed vector fields and the $\beta_J^k(s,t)$ are real valued coefficients which depend on the time intervals [s,t], and the sizes of which are controlled by the assumed α-continuity of the system $(\beta^1, \ldots, \beta^r)$. Thus, we need to look at the continuity properties, near $x = 0$, of the map

$$x \to T(x) = exp(\langle x, C \rangle)$$

where:

- x is a real vector of finite dimension:

$$x = \left((x_i^{(1)})_{1 \le u \le n},\ (x_{ij}^{(2)})_{1 \le i < j \le n}, \ldots \right);$$

  or more exactly,

$$x = \left((x_{I_k}^{(k)}),\ I_k \in \mathcal{J}(k) \subseteq \{1, \ldots, n\}^k,\ \ k = 1 \ldots r \right),$$

- C is similarly a finite array of vector fields:

$$C = \left((C_{I_k}^{(k)}),\ I_k \in \mathcal{J}(k) \subseteq \{1, \ldots, n\}^k,\ \ k = 1 \ldots r \right),$$

- $\langle x, C\rangle$ denotes the vector field

$$\langle x, C\rangle = \sum_{k=1}^{r} \sum_{I_k \in \mathcal{J}(k)} x_{I_k}^{(k)} \, C_{I_k}^{(k)}$$

(i.e. we sum over all $k = 1 \ldots r$ and over all multi-indices I_k in $\mathcal{J}(k) \subseteq \{1, \ldots, n\}^k$: $I_k = (i_1, \ldots, i_k)$, $i_j \in \{1, \ldots, n\}$.)

- Again, $\exp(\langle x, C\rangle)$ is the flow at time 1 of the ODE on R^d along the vector field $\langle x, C\rangle$.

Note that here $T = T(C)$, and dimensions of x and C depend on r.

Using this notation, we can write

$$\theta^{(n)}[S_{i-1}^{(n)}, S_i^{(n)}] = \exp(z(S_{i-1}^{(n)}, S_i^{(n)})) = T(\beta(S_{i-1}^{(n)}, S_i^{(n)}))$$

where $\beta(s,t) = ((\beta_{I_k}^k(s,t)),\ I_k \in \mathcal{J}(k),\ k = 1 \ldots r)$; which means that

- the one-step errors are of the form

$$T(\beta(\triangle_1 \cup \triangle_2)) - T(\beta(\triangle_1))\, T(\beta(\triangle_2))$$

$$\triangle_i \in \tau_{n+1}, \quad \triangle_1 \cup \triangle_2 \in \tau_n;$$

- the error propagation terms are of the form

$$T(\beta(\triangle_i)) \cdots T(\beta(\triangle_j))$$

$$\triangle_k \in \tau_n, \ k = 1, 2, \ \ 1 \le i \le j \le 2^n.$$

5.3 Case of linear DE

Let us now assume that the vector fields f_i are linear. Then $Lie(f_1, \ldots, f_n)$ is finite-dimensional, and we could use multiplicative notation and do all the calculations in logarithmic coordinates. We choose here to use the additive notation, so we interpret all f_i as bounded linear operators $R^d \to R^d$, with a norm given for instance by the operator norm $\|\cdot\|$ with

$$\|f_i\| = \sup_{\|u\| \neq 0} \frac{\|f_i(u)\|}{\|u\|} < \infty.$$

Now, their brackets of all orders exist and are bounded linear operators; so all components of C and all $z(s,t)$, $0 \leq s,t \leq 1$ are bounded linear operators. Furthermore, the operators $\exp(z(s,t))$ can be defined as bounded linear operators, given by the series

$$\exp(z) = \sum_{n=0}^{\infty} \frac{1}{n!} z^n,$$

and this definition agrees with the definition of $\exp(z)$ as a flow at time 1 of an ODE with respect to the vector field z. So, in this case, $T(x) = \exp(\langle x, C\rangle)$ is a bounded linear operator, and the series

$$\exp(\langle x, C\rangle) = \sum_{n=0}^{\infty} \frac{1}{n!} \langle x, C\rangle^n$$

gives a Taylor expansion of $T(x)$ with respect to the x-components at all points, where the series can be truncated at any level, giving a controlled remainder.

Thus, for each N, we can write:

$$T(x) = I + \sum_{M=1}^{N} \frac{1}{M!} \langle x, C\rangle^M + r_{(N+1)}(x)$$

where

$$\langle x, C\rangle^M = \prod_{m=1}^{M} \sum_{k(m)=1}^{r} \sum_{I_{k(m)} \in \mathcal{J}(k(m))} x_{I_{k(m)}}^{(k(m))} C_{I_{k(m)}}^{(k(m))}$$

and

$$\left\| r_{(N+1)}(x) \right\| \leq C \left\| \langle x, C\rangle \right\|^{N+1}.$$

Above, the Taylor series was "homogeneous", i.e. we took along terms up to N-th degree in all components of x, and the remainder term contains all other terms. However, we are interested in the case $x = \beta$, where the sizes of the components will vary. Thus, the components

$$\beta_i^1(s,t), \quad \beta_{ij}^2(s,t), \quad \beta_{ijk}^3(s,t) \ldots$$

are going to be of the order

$$|t-s|^{\alpha}, \quad |t-s|^{2\alpha}, \quad |t-s|^{3\alpha} \ldots$$

respectively. We will determine which terms to take along in our series expansion as follows: We will scale components of the vector x with

$$(\varepsilon x_i^{(1)}, \ \varepsilon^2 x_{ij}^{(2)}, \ \varepsilon^3 x_{ijk}^{(3)}, \ldots),$$

and take along terms where degree in ε would be at most N.

So, we are taking along only terms involving products

$$\prod_{m=1}^{M} x_{I_{k(m)}}^{(k(m))} C_{I_{k(m)}}^{(k(m))}$$

with $\sum_{m=1}^{M} k(m) \leq N$. This gives then

$$T(x) = I + \sum_{M=1}^{N} \frac{1}{M!} \prod_{m=1}^{M} \sum_{k(1),\dots,k(M):\sum_{m=1}^{M} k(m) \leq N} \sum_{I_{k(m)} \in \mathcal{J}(k(n))} x_{I_{k(m)}}^{(k(m))} C_{I_{k(m)}}^{(k(m))}$$

$$+ R_{(N+1)}(x)$$

where $R_{(N+1)}(x) = r_{(N+1)}(x) + L_{(N+1)}(x)$. The new remainder term $L_{(N+1)}(x)$ here contains the product with

$$\sum_{m=1}^{M} k(m) > N, \quad M \leq N :$$

that is,

$$L_{(N+1)}(x) = \sum_{M=1}^{N} \frac{1}{M!} \prod_{m=1}^{M} \sum_{k(1),\dots,k(M):\sum_{m=1}^{M} k(m) > N} \sum_{I_{k(m)} \in \mathcal{J}(k(n))} x_{I_{k(m)}}^{(k(m))} C_{I_{k(m)}}^{(k(m))}.$$

For instance, for $r = 2$,

$$\langle x, C \rangle = \sum_{i=1}^{n} x_i^{(1)} C_i^{(1)} + \sum_{i<j} x_{ij}^{(2)} C_{ij}^{(2)}$$

so that the expansion for T when we choose $N = 2$ is as follows:

$$T(x) = I + \sum_{i=1}^{n} x_i^{(1)} C_i^{(1)} + \sum_{i<j} x_{ij}^{(2)} C_{ij}^{(2)} + \frac{1}{2} \sum_{i,j} x_i^{(1)} x_j^{(1)} C_i^{(1)} C_j^{(1)} + R_3(x).$$

If we now choose $x = \beta(s,t)$, that is, for each $k = 1 \dots r$, $I(k) \in \{1, \dots, n\}^k$,

$$x_{I_k}^{(k)} = \beta_{I(k)}^{k}(s,t)$$

then the continuity assuption given in Definition 5.1 gives for all components of $\beta(s,t)$:

$$|\beta^k_{I(k)}(s,t)| \leq D\,|t-s|^{\alpha k}$$

for all $k = 1 \ldots r$, and all $I(k) \in \mathcal{J}(k)$ for a constant D. From this it follows that

$$\|\langle \beta(s,t), C\rangle\| \leq D\,|t-s|^{\alpha} \tag{10}$$

(if $\alpha > 0$, $|t-s| < 1$) where, for instance,

$$D = \sum_{k=1}^{r} \sum_{I(k)} \|C^{(k)}_{I_k}\|.$$

Lemma 5.6 *If β is an (α, r)-system, then for any N, the remainder term is bounded by*

$$\|R_{N+1}(\beta(s,t))\| \leq D\,|t-s|^{\alpha(N+1)}$$

for a constant $D = D(N, r, C)$.

Proof. As defined above, $R_{N+1}(\beta(s,t)) = r_{N+1}(\beta(s,t)) + L_{N+1}(\beta(s,t))$. Here $r_{N+1}(\beta(s,t))$ is the remainder term in the truncated exponential series, and therefore bounded by

$$\|r_{N+1}(\beta(s,t))\| \leq D\,\|\langle \beta(s,t), C\rangle\|^{N+1} \leq D_1\,|t-s|^{\alpha(N+1)}.$$

As for the remainder term $L_{N+1}(\beta(s,t))$, it is a sum of terms involving products

$$\prod_{m=1}^{M} \beta^{k(m)}_{I_{k(m)}}(s,t)\, C^{(k(m))}_{I_{k(m)}}$$

with

$$M \leq N, \quad \sum_{m=1}^{M} k(m) > N;$$

so for these products

$$\| \prod_{m=1}^{M} \beta_{I_{k(m)}}^{k(m)}(s,t) \| \leq D|t-s|^{\alpha(N+1)}$$

and

$$\max \left\{ \prod_{m=1}^{M} \|C_{I_{k(m)}}^{(k(m))}\|, \quad k(m) = 1 \ldots r, \right.$$

$$\left. I_{k(m)} \in \mathcal{J}(k(m)), \quad M = 1 \ldots N, \quad \sum_{m=1}^{M} k(m) > N \right\} < \infty.$$

□

So, this gives us the following expression:

$$T(\beta(\triangle)) = I$$

$$+ \sum_{M=1}^{N} \frac{1}{M!} \prod_{m=1}^{M} \sum_{k(1),\ldots,k(M):\sum_{m=1}^{M} k(m) \leq N} \sum_{I_{k(m)} \in \mathcal{J}(k(m))} \beta_{I_{k(m)}}^{k(m)}(\triangle)\, C_{I_{k(m)}}^{(k(m))}$$

$$+ R_{(N+1)}(\beta(\triangle))$$

with

$$\|R_{(N+1)}(\beta(\triangle))\| \leq C|\triangle|^{\alpha(N+1)}$$

for general intervals $\triangle = [s,t] \subseteq [0,1]$. We will use this expression on to get estimates for the sizes of the one step errors

$$T(\beta(\triangle_1 \cup \triangle_2)) - T(\beta(\triangle_1))\, T(\beta(\triangle_2)),$$

$$\triangle_1, \triangle_2 \in \tau_{n+1}, \quad \triangle_1 \cup \triangle_2 \in \tau_n,$$

and error propagation terms

$$T(\beta(\triangle_i)) \cdots T(\beta(\triangle_j)),$$

$$\triangle_l \in \tau_n, \;\; 1 \leq i \leq j \leq 2^n.$$

Theorem 5.7 *For the one step errors,*

$$\|T(\beta(\triangle_1 \cup \triangle_2)) - T(\beta(\triangle_1))\, T(\beta(\triangle_2))\| \leq D \mid \triangle \mid^{\alpha(r+1)},$$

where $\mid \triangle \mid$ is the length of the subintervals $\triangle_1$ and $\triangle_2$.

Proof. By the Campbell-Hausdorff-Baker Dynkin formula, we can write

$$\exp(z(\triangle_1)) \exp(z(\triangle_2)) = \exp(H(z(\triangle_1), z(\triangle_2)).$$

Then $z(\triangle_1 \cup \triangle_2)$ and $H(z(\triangle_1), z(\triangle_2))$ are identical, up to r-th order terms. This immediately follows from the definition of the $\beta(\triangle)$s and the Cs. (They obey the same formal algebraic rules as the corresponding quantities for smooth paths in the infinite series

$$\zeta(s,t) = \sum_{k=1}^{\infty} \sum_{I_k \in \mathcal{J}(k)} \beta^k_{I_k}(s,t)\, C^{(k)}_{I_k};$$

but $\exp(\zeta(s,t))$ has the flow property, which means that

$$H(\zeta(\triangle_1), \zeta(\triangle_2)) = \zeta(\triangle_1 \cup \triangle_2).)$$

Therefore, the series expansions for $\exp(z(\triangle_1 \cup \triangle_2))$ and

$$\exp(z(\triangle_1)) \exp(z(\triangle_2)) = \exp(H(z(\triangle_1), z(\triangle_2)))$$

are also the same up to r-th order terms, and their difference is a remainder term bounded according to Lemma 5.6 by

$$D \mid \triangle \mid^{\alpha(R+1)}.$$

□

Note that above, "k-th order terms" are defined as follows: we will scale the components of β as follows:

$$\left(\varepsilon\beta_i^1, \varepsilon^2\beta_{ij}^2, \ldots\right),$$

then k-th order terms are the coefficients of ε^k.

Of course, we can also prove the result by using the previously obtained Taylor expansions, with $N = r$, for all the exponentials, and by then comparing the terms. We will then use the consistency results between the $\beta^k(s,t)$ and the algebraic relations that hold for the vector fields C. These algebraic relations were of course initially based on the requirement that $z(\triangle_1 \cup \triangle_2)$ must equal $H(z(\triangle_1), z(\triangle_2))$, up to r-th order terms. We will do here the calculations in case r=2 to see what exactly happens:

Here

$$\langle x, C \rangle = \sum_{i=1}^{n} x_i^{(1)} C_i^{(1)} + \sum_{i<j} x_{ij}^{(2)} C_{ij}^{(2)}.$$

We use the expansion

$$T(x) = I + \sum_{i=1}^{n} x_i^{(1)} C_i^{(1)} + \sum_{i<j} x_{ij}^{(2)} C_{ij}^{(2)}$$

$$+ \frac{1}{2} \sum_{i,j} x_i^{(1)} x_j^{(1)} C_i^{(1)} C_j^{(1)} \quad + \quad R^3(x).$$

We will write

$$T(\beta(\Delta)) = \Sigma(\Delta) + R_3(\Delta)$$

where $R_3(\Delta) \leq D \mid \Delta \mid^{3\alpha}$; then

$$T(\beta(\Delta_1 \cup \Delta_2)) = \Sigma(\Delta_1 \cup \Delta_2) + R_3(\Delta_1 \cup \Delta_2),$$

$$T(\beta(\Delta_1))\, T(\beta(\Delta_2)) = \Sigma(\Delta_1)\, \Sigma(\Delta_2) + R_3(\Delta_1)\, \Sigma(\Delta_2)$$

$$+ \quad R_3(\Delta_2)\, \Sigma(\Delta_1) + R_3(\Delta_1)\, R_3(\Delta_2)$$

from which it follows that

$$T(\beta(\Delta_1 \cup \Delta_2)) - T(\beta(\Delta_1))\, T(\beta(\Delta_2))$$

$$= \quad \Sigma(\Delta_1 \cup \Delta_2) - \Sigma(\Delta_1)\, \Sigma(\Delta_2) \quad + \quad \delta_1(\Delta)$$

where $\|\delta_1(\Delta)\| \leq D \mid \Delta \mid^{3\alpha}$, and $\Delta = \Delta_1$ or Δ_2.

Claim: $\Sigma(\Delta_1 \cup \Delta_2) - \Sigma(\Delta_1)\, \Sigma(\Delta_2) = \delta_2(\Delta)$ where $\|\delta_2(\Delta)\| \leq D \mid \Delta \mid^{3\alpha}$.

Proof of claim:

The algebraic relations that hold for the vector fields and the iterated integrals are now,

$$\left[C_i^{(1)}, C_j^{(1)}\right] = C_{ij}^{(2)},$$

$$\beta_{ij}^2(\Delta_1 \cup \Delta_2) = \beta_{ij}^2(\Delta_1) + \beta_{ij}^2(\Delta_2)$$

$$+ \ \frac{1}{2}\left(\beta_i^1(\Delta_1)\beta_j^1(\Delta_2) - \beta_j^1(\Delta_1)\beta_i^1(\Delta_2)\right).$$

In $\Sigma(\Delta_1 \cup \Delta_2) - \Sigma(\Delta_1)\,\Sigma(\Delta_2)$, the first order terms are:

$$\sum_{i=1}^{n}\left(\beta_i^1(\Delta_1 \cup \Delta_2) - \beta_i^1(\Delta_1) - \beta_i^1(\Delta_2)\right)\ C_i^{(1)} = 0$$

due to the first consistency conditions; the second order terms are

$$\sum_{i<j}\beta_{ij}^2(\Delta_1 \cup \Delta_2)C_{ij}^{(2)} - \sum_{i<j}\beta_{ij}^2(\Delta_1)C_{ij}^{(2)} - \sum_{i<j}\beta_{ij}^2(\Delta_2)C_{ij}^{(2)}$$

$$+ \ \frac{1}{2}\sum_{i,j}\beta_i^1(\Delta_1 \cup \Delta_2)\beta_j^1(\Delta_1 \cup \Delta_2)C_i^{(1)}C_j^{(1)}$$

$$- \ \frac{1}{2}\sum_{i,j}\beta_i^1(\Delta_1)\beta_j^1(\Delta_1)C_i^{(1)}C_j^{(1)}$$

$$- \ \frac{1}{2}\sum_{i,j}\beta_i^1(\Delta_2)\beta_j^1(\Delta_2)C_i^{(1)}C_j^{(1)}$$

$$- \ \sum_{i,j}\beta_i^1(\Delta_1)\beta_j^1(\Delta_2)C_i^{(1)}C_j^{(1)}.$$

We will denote here the sum of the last 4 lines by B. To prove that the second order terms all vanish, it suffices to prove that

$$B = \frac{1}{2}\sum_{i<j}C_{ij}^{(2)}\left[\beta_i^1(\Delta_2)\beta_j^1(\Delta_1) - \beta_i^1(\Delta_1)\beta_j^1(\Delta_2)\right].$$

By writing in B,

$$\begin{aligned}\beta_i^1(\triangle_1 \cup \triangle_2) &= \beta_i^1(\triangle_1) + \beta_i^1(\triangle_2),\\ \beta_i^1(\triangle_1 \cup \triangle_2) &= \beta_i^1(\triangle_1) + \beta_i^1(\triangle_2)\end{aligned}$$

we get

$$B = \frac{1}{2}\sum_{i,j} C_i^{(1)} C_j^{(1)} \left[\beta_i^1(\triangle_2)\beta_j^1(\triangle_1) - \beta_i^1(\triangle_1)\beta_j^1(\triangle_2)\right].$$

But since the term inside the brackets is antisymmetric, this equals

$$\begin{aligned}&\frac{1}{2}\sum_{i<j}(C_i^{(1)}C_j^{(1)} - C_j^{(1)}C_i^{(1)})[\beta_i^1(\triangle_2)\beta_j^1(\triangle_1) - \beta_i^1(\triangle_1)\beta_j^1(\triangle_2)]\\ &= \frac{1}{2}\sum_{i<j} C_{ij}^{(2)}[\beta_i^1(\triangle_2)\beta_j^1(\triangle_1) - \beta_i^1(\triangle_1)\beta_j^1(\triangle_2)].\end{aligned}$$

The error propagation term is more problematic. Of course, according to Lemma 5.5 we have the following:

Theorem 5.8 *If the error propagation terms are uniformly bounded:*

$$\sup_{i,j} \|T(\beta(\triangle_i)) \cdots T(\beta(\triangle_j))\| < C$$

for all consequent subintervals of any subdivision τ_n, then the series $\theta^{(n)}[0,1]$ converges if

$$\alpha(r+1) > 1$$

If $\alpha \geq 1$, then we can use a direct estimate: For instance,

$$\|T(\beta(\triangle_1)) \cdots T(\beta(\triangle_{2^n}))\|$$

$$\leq \prod_{k=1}^{2^n} \|T(\beta(\triangle_k))\| \leq \prod_{k=1}^{2^n} \exp \|\langle \beta(\triangle_k), C\rangle\|$$

$$= \exp \left(\sum_{k=i}^{2^n} \|\langle \beta(\triangle_k), C\rangle\| \right)$$

$$\leq \exp\left(D 2^n \mid \triangle \mid^{\alpha}\right) = \exp(D\,(2^n)^{1-\alpha}) \leq D_1,$$

according to (10). So, in this case, according to Lemma 5.5 in last section, we get the following result:

Theorem 5.9 *If* $\alpha \geq 1$ *and* $r \geq 1$ *then* $\theta^{(n)}(0,1)$ *converges.*

For $\alpha < 1$, obtaining general estimates for the error propagation term is more difficult. In this linear case we can see fairly clearly what happens to the error propagation terms. We know from Lyons [37] that in case $\alpha(r+1) > 1$ there exists a unique procedure for obtaining for each $p > r$, approximations for β^p over the interval

$$\triangle = \bigcup_{k=1}^{2^n} \triangle_k$$

by using the values of $\beta^i(\triangle_k)$, $i = 1 \ldots r$, $k = 1 \ldots 2^n$. Denote this approximation by $\beta^p(n)$. From the convergence result that holds for these approximations (cf. Lyons [37], see also Result 4.3) we know that for each p,

$$\beta^p(n) \to \beta^p, \quad n \to \infty,$$

and β^p is αp-continuous. Now, using this and the Campbell-Hausdorff-Baker-Dynkin formula, we can actually write

$$\prod_{k=1}^{2^n} T(\beta(\Delta_k)) = \exp(\langle \beta(n)(\Delta), C \rangle)$$

where

$$\Delta = \bigcup_{k=1}^{2^n} \Delta_k.$$

Here $\beta(n) = (\beta^1(n), \beta^2(n), \ldots)$ is an infinite sequence and C is the complete infinite array of vector fields $C_J^{(k)}$, $k = 1, 2, ..$ as in Definition 4.1; so $\langle \beta(n)(\Delta), C \rangle$ is an infinite series. Particularly then, $\beta^p(n) = \beta^p$ when $p = 1 \ldots r$ and for any fixed $p > r$, the sequence $\beta^p(n)$ converges towards a αp-continuous two parameter function β^p. Now, if we could prove that $\langle \beta(n)(\Delta), C \rangle$ converges towards a bounded linear operator, then obviously boundedness of the error propagation terms for our construction of the solution would follow and we could have a result stating that the solution exists, i.e. the sequence $\theta^{(n)}[0,1]$ converges, if $\alpha(r+1) > 1$. For this it is sufficient to prove that

$$\langle \beta(n)(\Delta), C \rangle \rightarrow \langle \beta(\Delta), C \rangle = \sum_{k=1}^{\infty} \sum_{J \in \mathcal{J}(k)} \beta_J^k(\Delta) C_J^k$$

and this series converges. The convergence of the sequence here of course immediately follows if we could prove that the convergence

$$\beta^p(n) \rightarrow \beta^p$$

is uniform over all p simultaneously. Unfortunately, the question of convergence of this infinite series and the rate of convergence of the $\beta^p(n)$s is still open.

Of course, if the conditions of Lemma 5.3 hold, i.e. if the one step errors (in multiplicative notation) commute with all the operators $\theta_i^{(n)}$, then the error propagation ceases to be a problem and we get directly the following result: (all the calculations can easily be done in logarithmic coordinates)

Theorem 5.10 *Assume that the vector fields f_i are such that for each n, all the operators $\varepsilon_i^{(n)}$ commute with all the operators $\theta_j^{(n)}$ (i.e. $\theta_j^{(n)}\varepsilon_i^{(n)} = \varepsilon_i^{(n)}\theta_j^{(n)} \ \forall\, i,j.$) Then the series $\theta^{(n)}[0,1]$ converges if $\alpha(r+1) > 1$. Here, $\varepsilon_i^{(n)} = \theta_i^{(n)} \left(\theta_{(i)}^{(n+1)}\right)^{-1}$; $i = 1 \dots 2^n$.*

5.4 About the general case

We will look in this section briefly at the general case, i.e. the differential equations with C^r continuous vector fields. We can still interpret the operators $\theta^{(i)}$ as linear operators, defined in a function space if necessary, and a norm can be accordingly defined for the operators involved — however in this case the choice of the norm is complicated. The operators T will be well defined, as homeomorphisms or as linear operators via action on smooth functions. If the map

$$x \rightarrow T(x)$$

is $r+1$ times continuously differentiable, for which the condition of C^r-continuity for the vector fields is enough, then we can use a Taylor expansion to obtain again a series expression for the $T(x)$ as above, with a controlled remainder term, assuming that we have fixed a good enough norm in the operator space. For the one-step errors, the Campbell-Hausdorff-Baker-Dynkin formula is now not available, but we can still compare the terms in the expressions for the one step errors to obtain the following result:

Theorem 5.11 *If the vector fields f_i are in C^r, then the one step errors are bounded by*

$$\|T(\beta(\triangle_1 \cup \triangle_2)) - T(\beta(\triangle_1))\,T(\beta(\triangle_2))\| \leq D \mid \triangle \mid^{\alpha(r+1)},$$

where $\mid \triangle \mid$ is the length of the subintervals $\triangle_1$ and $\triangle_2$.

Again this means that if we could prove that the error propagation term is bounded, then convergence would follow if $\alpha(r+1) > 1$. But direct control of the error propagation here is even more difficult than in the linear case.

Of course we have again the special cases:

Theorem 5.12 *If $Lie(f_1 \dots f_n)$ is finite dimensional, and the conditions of Lemma 5.3 hold, then the series $\theta^{(n)}(0,1)$ converges if $\alpha(r+1) > 1$.*

5.5 Convergence towards a given flow

The next theorem gives sufficient conditions for establishing the convergence of the sequence $\theta^{(n)}$ towards a given flow $\hat{\theta}$. This will make it easy, for instance, to compare the solution in the sense of Definition 5.2 with other concepts of solution. We denote again

$$z[s,t] = \sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta_J^k(s,t)\, C_J^{(k)}.$$

Theorem 5.13 *Let $\hat{\theta} = (\hat{\theta}(s,t),\ 0 \leq s,t \leq 1)$ be a flow of homeomorphisms over $[0,1]$, with the boundedness condition*

$$\|\hat{\theta}(s,t)\| \leq C, \quad \forall\ [s,t] \subseteq [0,1].$$

If

$$\frac{\|\hat{\theta}(s,t) - \exp(z[s,t])\|}{|t-s|} \to 0$$

as $|t-s| \to 0$ *for all* $[s,t] \subseteq [0,1]$, *then*

$$\|\hat{\theta}(0,1) - \theta^{(n)}(0,1)\| \to 0, \quad n \to \infty.$$

Especially, it suffices that for all $[s,t] \subseteq [0.1]$

$$\frac{\|\hat{\theta}(s,t) - \exp(z[s,t])\|}{|t-s|^{\alpha(r+1)}}$$

is bounded as $|t-s| \to 0$, *with* α, r *such that* $\alpha(r+1) > 1$.

Thus, to guarantee convergence it is sufficient that the flow $\hat{\theta}$ is asymptotically close to our way of constructing the solution. The continuity requirement means that the two operators above agree 'up to terms of degree αr'.

We will use the following simple modification of Lemma 5.4 to compare two operators:

Lemma 5.14 *If*

$$r(m) = \hat{\theta}_m \cdots \hat{\theta}_1 - \theta_m \cdots \theta_1$$

then for $e_i = \hat{\theta}_i - \theta_i$

$$r(m) = \sum_{i=1}^{m} \hat{\theta}_m \cdots \hat{\theta}_{i+1}\, e_i\, \hat{\theta}_{i-1} \cdots \hat{\theta}_1 - \sum_{i=1}^{m} \hat{\theta}_m \cdots \hat{\theta}_{i+1}\, e_i\, r(i-1).$$

From this, it follows that

$$\sup_{m \leq M} \|r(m)\| \leq M \left(\sup_{1 \leq j \leq i \leq M} \|\hat{\theta}_i \cdots \hat{\theta}_j\| \right)^2 \sup_{1 \leq i \leq M} \|e_i\|$$

$$+ M \sup_{1\le j\le i\le M} \|\hat{\theta}_i \cdots \hat{\theta}_j\| \sup_{1\le i\le M} \|e_i\| \sup_{m\le M} \|r(m)\|$$

and therefore,

$$\sup_{m\le M} \|r(m)\| \le \frac{1}{1 - M \sup_{1\le j\le i\le M} \|\hat{\theta}_i \cdots \hat{\theta}_j\| \sup_{1\le i\le M} \|e_i\|}.$$

$$M \left(\sup_{1\le j\le i\le M} \|\hat{\theta}_i \cdots \hat{\theta}_j\| \right)^2 \sup_{1\le i\le M} \|e_i\|.$$

Proof of Theorem:. The above Lemma gives:

$$\|\hat{\theta}(0,1) - \theta^{(n)}(0,1)\| \le \frac{1}{1 - 2^n \sup_{i,j} \|\hat{\theta}(\triangle_i) \cdots \hat{\theta}(\triangle_j)\| \sup_i \|\hat{\theta}(\triangle_i) - \theta^{(n)}(\triangle_i)\|}.$$

$$\cdot 2^n \sup_{i,j} \|\hat{\theta}(\triangle_i) \cdots \hat{\theta}(\triangle_j)\|^2 \sup_i \|\hat{\theta}(\triangle_i) - \theta^{(n)}(\triangle_i)\|$$

where $\triangle_i \in \tau_n$. By assumption, $\hat{\theta}$ has the flow property, so that

$$\hat{\theta}(\triangle_i) \cdots \hat{\theta}(\triangle_j) = \hat{\theta}(\triangle_i \cup \ldots \cup \triangle_j).$$

Now, since we assumed

$$\sup_{i,j} \|\hat{\theta}(\triangle_i) \cdots \hat{\theta}(\triangle_j)\| \le C,$$

we get

$$\|\hat{\theta}(0,1) - \theta^{(n)}(0,1)\| \le \frac{1}{1 - 2^n C \sup_i \|\hat{\theta}(\triangle_i) - \theta^{(n)}(\triangle_i)\|} C^2 2^n \sup_i \|\hat{\theta}(\triangle_i) - \theta^{(n)}(\triangle_i)\|,$$

from which the claim immediately follows when we let $n \to \infty$. □

Note that the result above holds for either the linear or the general differential equation. The norm of the operators has to be fixed of course; this is simple in the linear case but will again be more complicated in the general case.

5.6 Notes and open questions

PROPERTIES OF THE SOLUTION

The solution here will in general depend not only on the path X, but also on the chosen "iterated integrals" $(\beta^2, \ldots, \beta^r)$, i.e. on the whole generalized path β. Thus the solution is a map

$\beta \quad \rightarrow \quad$ solution of (8) with respect to generalized noise β.

Therefore, this concept of solution for differential equations would hopefully have the continuity properties we wished for: $\beta \approx \hat{\beta}$ should imply that [solution with respect to β] $\approx$ [solution with respect to $\hat{\beta}$], since by definition β involves not only the path X but some of its iterated path integrals. Thus, the map from generalized paths to solutions should be continuous with respect to some suitably chosen metric for the generalized noises β. We will not here look into this matter except to point out that a suitable metric could be

$$\|\beta\|_\alpha = \max_{1 \leq k \leq r} \sup_{0<s,t<1} \frac{|\beta^k(s,t)|}{|t-s|^{\alpha k}}$$

(cf. Lyons [37].)

ABOUT HIGHER ORDER METHODS

If the r first "iterated integrals" $(\beta^1, \ldots, \beta^r)$ are given and form an (α, r)-system, and $\alpha(r+1) > 1$, then according to the results in Lyons [37] (see Result 4.3 above), it is possible to construct any amount of higher order iterated integrals, such that the corresponding continuity and consistency requirements still hold. Now, if the vector fields in the differential equation (8) are sufficiently smooth, then we might consider using a "higher order" method to construct a solution to the differential equation. That is, for any fixed $p > r$, we could construct from the given (α, r)-system $(\beta^1, \ldots, \beta^r)$ an (α, p)-system $(\beta^1, \ldots, \beta^p)$ which is then also a generalization of the path X. Now using a construction identical to the one given in Section 5.1, but with r replaced by p, would give a sequence of approximate solutions based now on the iterated integrals $(\beta^1, \ldots, \beta^p)$. Denote this sequence by $\theta^{(n)}_{(p)}(0,1)$, $n \in N$ – accordingly, the $\theta^{(n)}(0,1)$-sequence in the previous could be denoted by $\theta^{(n)}_{(r)}(0,1), n \in N$. The next theorem guarantees that these alternative methods (in the first one, constructing a solution directly from $(\beta^1, \ldots, \beta^r)$; in the second, constructing first $\beta^{r+1}, \ldots, \beta^p$ from $(\beta^1, \ldots, \beta^r)$ and then constructing a solution from $(\beta^1, \ldots, \beta^p)$) both give the same answer when they do converge towards a solution.

In the following, similarly to the notation used at the end of Section 5.3, we will denote by $\beta^q_{(n,r)}, q = r+1, \ldots, p$, the approximations for β^q obtained by using the values $\beta^k(\Delta_i)$, $k = 1 \ldots r$, $i = 1 \ldots 2^n$; the result in Lyons [37] gives for each q, the convergence result that

$$\beta^q_{(\infty,r)} = \lim_{n \to \infty} \beta^q_{(n,r)}$$

exists.

Theorem 5.15 *Let $(\beta^1, \ldots, \beta^r)$ be an (α, r)-system, with $\alpha(r+1) > 1$. Fix $p > r$, and assume that the vector fields f_i in the differential equation (8) are C^{p+1}-continuous. Let $\theta^{(n)}_{(r)}(0,1)$, and $\theta^{(n)}_{(p)}(0,1)$ be sequences of operators formed with a "r-th order method" and a "p-th order method", respectively; that is, define*

$$\theta^{(n)}_{(r)}(0,1) = \prod_{i=1}^{2^n} \exp\left(\sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta^k_J((i-1)2^{-n}, i2^{-n})\, C^{(k)}_J \right),$$

and

$$\theta^{(n)}_{(p)}(0,1) = \prod_{i=1}^{2^n} \exp\left(\sum_{k=1}^{p} \sum_{J \in \mathcal{J}(k)} \beta^k_J((i-1)2^{-n}, i2^{-n})\, C^{(k)}_J \right),$$

where $\beta^{r+1}, \ldots, \beta^p$ are constructed from $(\beta^1, \ldots, \beta^r)$ by the limiting procedure given in Lyons [37] (i.e. $\beta^q = \beta^q_{(\infty,r)}$, $q = r+1, \ldots, p$). If the sequence $\theta^{(n)}_{(r)}(0,1)$ converges, then the sequence $\theta^{(n)}_{(p)}(0,1)$ converges also, and towards the same value.

Proof. Due to Theorem 5.13, it suffices to prove that

$$\|\theta^{(\infty)}_{(r)}(\Delta) - \exp\left(\sum_{k=1}^{p} \sum_{J \in \mathcal{J}(k)} \beta^k_J(\Delta)\, C^{(k)}_J \right) \| \leq C|\,\Delta\,|^{\alpha(p+1)}$$

for all sufficiently small intervals $\Delta \subseteq [0,1]$. So fix a subinterval $\Delta \in \tau_n$; then

$$\theta^{(n)}_{(p)}(\Delta) = \exp\left(\sum_{k=1}^{p} \sum_{J \in \mathcal{J}(k)} \beta^k_J(\Delta)\, C^{(k)}_J \right),$$

$$\theta^{(n)}_{(r)}(\Delta) = \exp\left(\sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta^k_J(\Delta)\, C^{(k)}_J \right).$$

For any m, we can write

$$\|\theta^{(\infty)}_{(r)}(\Delta) - \theta^{(n)}_{(p)}(\Delta)\|$$

$$= \|\theta^{(\infty)}_{(r)}(\Delta) - \theta^{(n+m)}_{(r)}(\Delta)\| + \|\theta^{(n+m)}_{(r)}(\Delta) - S^{(p)}_{(m,r)}(\Delta)\|$$

$$+ \|S^{(p)}_{(m,r)}(\Delta) - S^{(p)}(\Delta)\| + \|S^{(p)}(\Delta) - \theta^{(n)}_{(p)}(\Delta)\|$$

where:

- $$\theta^{(n+m)}_{(r)}(\Delta) = \prod_{i=1}^{2^m} \theta^{(n)}_{(r)}(\Delta_i) = \prod_{i=1}^{2^m} \exp\left(\sum_{k=1}^{r} \sum_{J \in \mathcal{J}(k)} \beta^k_J(\Delta_i)\, C^{(k)}_J \right)$$

  (here, $\Delta = \cup_{i=1}^{2^m} \Delta_i$);

- $S^{(p)}_{(m,r)}(\triangle)$ denotes the sum up to p-th order terms in the Taylor expansion for the product $\theta^{(n+m)}_{(r)}(\triangle)$, with respect to the $\beta^k(\triangle_i)$s; (here, which terms are to be taken along is again determined by the scaling rules described in Section 5.3);

- $S^{(p)}(\triangle)$ is the sum up to p-th order terms in the Taylor expansion for $\theta^{(n)}_{(p)}(\triangle)$.

We assumed that $\theta^{(n)}_{(r)}$ converges, so that

$$\|\theta^{(\infty)}_{(r)}(\triangle) - \theta^{(n+m)}_{(r)}(\triangle)\| \to 0$$

as $m \to \infty$. By Lemma 5.6,

$$\|S^{(p)}(\triangle) - \theta^{(n)}_{(p)}(\triangle)\| \leq C|\triangle|^{\alpha(p+1)}.$$

Also, by applying this Lemma to the components of the product in $\theta^{(n+m)}_{(r)}$, we get:

$$\|\theta^{(n+m)}_{(r)}(\triangle) - S^{(p)}_{(m,r)}(\triangle)\| \leq M(m)|\triangle|^{\alpha(p+1)}$$

where $M(m) \to 0$ as $m \to \infty$.

As for the difference of the p-th order Taylor expansions $S^{(p)}_{(m,r)}(\triangle)$ and $S^{(p)}(\triangle)$, note first that we can obviously write

$$S^{(p)}(\triangle) = I + \sum_{q=1}^{p} h^q_p(\beta^1(\triangle), \ldots, \beta^r(\triangle), \beta^{r+1}(\triangle), \ldots, \beta^q(\triangle))$$

for some continuous functions h^q_p, $q = 1 \ldots p$. Now it is seen, by looking at the way the sequence $\theta^{(n+m)}_{(r)}(\triangle)$ over the whole interval is constructed from its values over subintervals, and the way the new higher order iterated integrals are constructed from lower-order ones (cf. proof of Theorem 5.7), that similarly,

$$S^{(p)}_{(m,r)}(\triangle) = I \; + \; \sum_{q=1}^{p} h^q_p(\beta^1(\triangle), \ldots, \beta^r(\triangle), \beta^{r+1}_{(m+n,r)}(\triangle), \ldots, \beta^q_{(m+n,r)}(\triangle)).$$

Therefore, it follows that

$$S^{(p)}_{(m,r)}(\triangle) \to S^{(p)}(\triangle), \quad m \to \infty.$$

The claim of the theorem now follows when we let $m \to \infty$. □

Note that this means that when we try to solve the differential equation by the method outlined in this thesis, starting from a given (α, r)-system $(\beta^1, \ldots, \beta^r)$ with $\alpha(r+1) > 1$, then nothing can be gained by first extending it to a $(\beta^1, \ldots, \beta^p)$-system with $p > r$, and then using a p-th order method, since one then ends up with repeating essentially the same limiting procedure twice; both methods give the same answer, but a higher order method requires more smoothness from the vector fields in the differential equation.

REMAINING QUESTIONS

The question of convergence of our solution is still unsolved in the general case. In the linear case the convergence was proved to be directly linked to the convergence of the formal Lie series in Section 4.2 — proving the convergence of this might also shed light on the convergence of the corresponding Lie series in results by Ben Arous [6] and Castell [8] in Section 4.1, which have in general so far only been proved in a fairly roundabout way in both these articles. However, in many applications of our results the result in Section 5.5, which proves the convergence towards an existing solution, is quite interesting in itself.

In the general case we have not tackled in a very detailed manner the question of choosing a norm for the operators, interpreted as linear operators on function spaces.

Finally, there are some further special cases which would be worth looking into: E.g. the cases where $Lie(f_i \dots f_n)$ is nilpotent, or solvable.

6 Special cases and examples

Let $X = (X_t, t \in [0,1])$ be a continuous path on R^n, and look at a "differential equation"

$$dY_t = \sum_{i=1}^{n} f_i(Y_t) dX_t^i \tag{11}$$

formally defined with respect to it. (See Chapter 5.) According to the remarks in the previous chapter, an important question when trying to interpret what we mean by this equation is whether we can extend X into an (α, r) -system, with $\alpha(r+1) > 1$. (Cf. Definition 5.1 in Chapter 5.) If this can be done, then, using the procedure given in Chapter 5, it may be possible to give explicit meaning to the "differential equation", for a large class of vector fields (f_i), together with a way of constructing a "solution" of it. Thus, if we want to use this method for a given path X, we need to:

- define plausible generalized "area integrals" and other "iterated integrals" $\beta^k, k = 2 \ldots r$, which obey the consistency rules in Definition 5.1;
- prove that these together with X form an α -continuous system:

$$|\beta^k(s,t)| \le c\,|t-s|^{\alpha k} \quad \forall [s,t];$$

- and do this with values α, r such that $\alpha(r+1) > 1$.

Of course, to start with, the path X must be Holder(α)-continuous. The condition $\alpha(r+1) > 1$ means that:

- $r = 0$ is sufficient with $\alpha > 1$. (This is trivial of course, since a Holder(α)-continuous function with $\alpha > 1$ is a constant function, and thus the flow of the solution is the identity map.)
- $r = 1$ is sufficient with $\alpha \in (1/2, 1]$. (Here β^2 in the (α, r)-system will be unique, if it exists.)
- $r = 2$ is sufficient with $\alpha \in (1/3, 1/2]$ (in which case β^2 will not be unique).

If $\alpha = 1$, i.e. X is Lipschitz-continuous, then there is of course also available the Lebesque-Stieltjes theory of integration and the resulting theory of ordinary differential equations. β^2 and the other iterated integrals are now equal to the corresponding Lebesque-Stieltjes iterated integrals, and the solution obtained by using the results in Chapter 5 coincides with the solution of the corresponding ODE.

The cases of real interest are when $\alpha < 1$, the ODE theory is not available, and the β^ks do not have an obvious definition. For certain stochastic processes X, suitable β^ks may be defined through probabilistic reasoning, even when almost all paths of X have less than Holder(1/2) -continuity. It is this stochastic setting that will be our main concern here.

6.1 Case of stochastic processes

Let X be a stochastic process, defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. To adapt our result to handle paths of stochastic processes, we obviously need to prove that P-almost all paths $X(\omega)$ can be extended into an (α, r)-system. That is, we must define for almost every $\omega \in \Omega$, 2-parameter functions $\beta^2(\omega), ..., \beta^r(\omega)$ which together with $X(\omega)$ form a consistent system; and we must prove that almost surely, $X(\omega)$ and $\beta^2(\omega), ..., \beta^r(\omega)$ have the required α-continuities.

The differential equation (11) and the solution of it have then pathwise meaning, when defined through the procedure in Chapter 5. The path $X(\omega)$ has been extended into a generalised path $\beta(\omega)$. The obtained flow of solution $\theta = \theta(\omega)$, when it exists, provides a map

$$\beta(\omega) \rightarrow Y(\omega, y_0)$$

for P-almost every $\omega \in \Omega$, where $Y(\omega, y_0)$ is the solution in the sense of Chapter 5 of the differential equation (11), with starting point y_0:

$$Y(\omega, y_0) = \theta(\omega) y_0.$$

Also, importantly, since the procedure will be completely pathwise, the concept of filtrations loses its meaning here.

The cases where almost all paths $X(\omega)$ are Lipschitz-continuous can be passed without a mention. In addition to these cases, however, the stochastic nature of some processes will enable us to complete them with "iterated integrals", defined by probabilistic means and as 2-parameter stochastic processes. Before going into some special cases, we will give here a result which will prove useful in establishing the required α-continuity for a system $(\beta^1, \ldots, \beta^r)$ of continuous stochastic processes.

6.2 A continuity result for stochastic processes

All the stochastic processes in this section are defined in the same probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $(\beta^k, k = 1..r)$ form a"consistent r-system" of 2-parameter stochastic processes over $[0,1]$, assumed to be continuous in both parameters; i.e.

$$\beta^k = \left(\beta^k_J(s,t) \mid k = 1 \ldots r,\ J \in \{1, \ldots, n\}^k,\ [s,t] \subseteq [0,1]\right)$$

and the β^k's are linked together by the consistency equations in Definition 5.1 now taken to hold almost surely. The following result reduces the almost sure pathwise Holder-continuity of the β^k's to conditions on the moments of these processes. (This result is, of course, closely related to classical proofs of pathwise continuity of Brownian motion — see Ito-McKean [25], Knight [29], and others.)

Theorem 6.1 *Fix $\gamma > 0$. If for all $[s,t] \subseteq [0,1],\ k = 1 \ldots r, J \in \{1, \ldots, n\}^k$,*

$$E|\beta^k_J(s,t)|^{(p/k)*} \leq c(p)|t-s|^{p\gamma}$$

for some $p \in N, c(p) < \infty$, then for any $\alpha > 0$ with $(\gamma - \alpha)p > 1$, for almost every $\omega \in \Omega$, there exists a $C = C(\omega) < \infty$ such that

$$|\beta^k_J(s,t)| \leq C|t-s|^{\alpha k}.$$

(Here we denote $a* = min\{n \in N : n \geq a\}$.)

The proof is given here in detail only for cases $r = 1, r = 2$. For more general r the proof is similar, but notations get increasingly more complicated.

For $r = 1, r = 2$ to simplify the situation we will write

$$\begin{cases} \beta_i^1(s,t) &= X_i(t) - X_i(s), \ i = 1,2 \\ \beta_{1,2}^2(s,t) &= \beta(s,t) \end{cases}$$

where X_1, X_2 and β are all continuous, and linked together by the consistency condition:

$$\beta(s,u) = \beta(s,t) + \beta(t,u) \tag{12}$$

$$+\frac{1}{2}\left[(X_1(t) - X_1(s))(X_2(u) - X_2(t)) - (X_1(u) - X_1(t))(X_2(t) - X_2(s))\right]$$

$$\forall s \leq t \leq u.$$

Here the last term is the area of the triangle on R^2 with vertices

$$(X_1, X_2)(s), \ (X_1, X_2)(t), \ (X_1, X_2)(u);$$

this geometrical interpretation will be utilized in the proof for cases $r = 1, r = 2$.

Theorem 6.2 *(This is a special case of Theorem 6.1, with $r = 1, r = 2$.)*

Fix $\gamma > 0$.

(i) If

$$E(|X_i(t) - X_i(s)|^p) \leq C(p)|t-s|^{p\gamma}, \ i = 1,2, \ \forall [s,t] \subseteq [0,1] \tag{13}$$

for a $p \in N$ and a $C(p) < \infty$, then for any α with $(\gamma - \alpha)p > 1$, for almost every $\omega \in \Omega$, there exists a constant $C = C(\omega) < \infty$ such that

$$|X_i(t) - X_i(s)| \leq C|t-s|^{\alpha} \quad \forall [s,t] \subseteq [0,1], i = 1,2. \tag{14}$$

(ii) If furthermore

$$E|\beta(s,t)|^{(p/2)*} \leq C(p)|t-s|^{p\gamma} \quad \forall [s,t] \subseteq [0,1], \tag{15}$$

then also P -almost surely

$$|\beta(s,t)| \leq C|t-s|^{2\alpha} \quad \forall [s,t] \subseteq [0,1] \tag{16}$$

(Here $a* = min\{n \in N : n \geq a\}$.)

Proof of Theorem 6.2.

For both (i) and (ii) the proofs will be done in the following steps:

Step 1. First, it will be proved by a Borell-Cantelli type argument that for any fixed $K \in N$, the inequality ((14) or (16) respectively) holds almost surely simultaneously over all intervals of the form $[s,t] = [i2^{-n}, j2^{-n}]$ where $|j-i| \leq K$ and n is sufficiently large: $n \geq n(K,\omega)$.

Step 2. To extend this to arbitrary time intervals $[s,t]$ (sufficiently small: $|t-s| \leq \delta(K,\omega)$), we will write

$$[s,t] = [s, i2^{-n}] \cup [i2^{-n}, j2^{-n}] \cup [j2^{-n}, t],$$

where $|j-i| \leq K$. These three intervals are either dyadic or can be written as unions of dyadic intervals, so that the continuities over these intervals follow from step 1.

Step 3. Finally, the consistency assumption (1) tells us how to reduce the continuity over $[s,t]$ into continuities over these subintervals. Of course, in case (ii), this step involves the continuities of both β and X_1, X_2.

Proof of (i). Let K be fixed. We'll prove that P -almost surely the inequality

$$|X(t) - X(s)| \leq C|t-s|^{\alpha}, \quad X = X_1 \textit{ or } X_2$$

holds simultaneously over all intervals of the form $[s,t] = [i2^{-n}, j2^{-n}]$ with $|j-i| \leq K$, when $n \geq n(K,\omega)$. By the Borell-Cantelli Lemma, it suffices to prove that

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

where $A_n = \{\omega \in \Omega |$ inequality

$$|X(t) - X(s)| > C|t-s|^{\alpha} \tag{17}$$

holds for at least one subinterval $[s,t] = [i2^{-n}, j2^{-n}]$ with $|j-i| \leq K\}$. Here

$$\begin{aligned} P(A_n) &= P\left(\bigcup_{j-i \leq K, i<j} \{\omega \in \Omega \,|\, (17) \textit{ holds over } [i2^{-n}, j2^{-n}]\} \right) \\ &\leq \sum_{m \leq K} \sum_{|j-i|=m} P\{\omega \in \Omega \,|\, (17) \textit{ holds over } [i2^{-n}, j2^{-n}]\} \end{aligned}$$

and Chebyshev's inequality and the assumed sizes of the p^{th} moments give

$$\begin{aligned} & P\left\{(17) \textit{ holds over } [i2^{-n}, j2^{-n}]\right\} \\ = \; & P\left\{|X(j2^{-n}) - X(i2^{-n})| > C|j2^{-n} - i2^{-n}|^{\alpha}\right\} \\ \leq \; & \frac{C(p)}{C^p}|m2^{-n}|^{(\gamma-\alpha)p} \\ = \; & \frac{C(p)}{C^p} m^{(\gamma-\alpha)p} 2^{-n(\gamma-\alpha)p} \end{aligned}$$

when $[i2^{-n}, j2^{-n}]$ is an interval of length $(j-i)2^{-n} = m2^{-n}$.

Also,

$$\#\{(i,j) \mid 0 \leq i < j \leq 2^n, |j-i| = m\} = 2^n - m,$$

so that

$$\begin{aligned} P(A_n) &\leq \frac{C(p)}{C^p} 2^{-n(\gamma-\alpha)p} \sum_{1 \leq m \leq n} (2^n - m) m^{(\gamma-\alpha)p} \\ &\leq \frac{C(p)}{C^p} 2^{-n(\gamma-\alpha)p} 2^n K K^{(\gamma-\alpha)p}. \end{aligned}$$

Therefore

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

if $(\gamma - \alpha)p > 1$.

To extend the required almost sure continuity to hold simultaneously over all intervals $[s,t]$ with $|t-s| \leq K2^{-n(K)}$, note that for each such interval there exists $n \geq n(K)$ and i, j with $|j-i| \leq K$ such that

$$(i-1)2^{-n} < s \leq i2^{-n} < j2^{-n} \leq t < (j+1)2^{-n}.$$

Now we can write

$$X(t) - X(s) = (X(i2^{-n}) - X(s))$$

$$+(X(j2^{-n}) - X(i2^{-n})) + (X(t) - X(j2^{-n}))$$

where

$$|X(j2^{-n}) - X(i2^{-n})|$$

$$\leq C|j2^{-n} - i2^{-n}|^\alpha \leq C|t-s|^\alpha$$

by the result above, since $n \geq n(K)$.

To prove that

$$\begin{aligned} |X(i2^{-n}) - X(s)| &\leq C|s - i2^{-n}|^\alpha \\ &\leq C|t - s|^\alpha \\ |X(t) - X(j2^{-n})| &\leq C|t - j2^{-n}|^\alpha \\ &\leq C|t - s|^\alpha, \end{aligned}$$

we will write the corresponding intervals as unions of dyadic subintervals. Due to symmetry, it suffices to prove that

$$|X(r) - X(s)| \leq C|r - s|^\alpha$$

when $r = i2^{-n}, r > s, s$ is arbitrary, and $|r - s| < 2^{-n}$. It is possible to construct a sequence of disjoint dyadic intervals, the union of which is the interval $[r, s]$. Define p_i such that

$$s = r - \left(\sum_{i=1}^{\infty} 2^{-p_i} \varepsilon_i\right),$$

where $n < p_1 < p_2 < \ldots$, $\varepsilon_i = 0$ or 1 and $\varepsilon_1 = 1$. Then

$$[s, r] = \bigcup_{i=1}^{\infty} [d_i, d_{i-1}]$$

with dyadic division points d_k obviously defined by

$$d_0 = r, \; d_k = r - \sum_{i=1}^{k} 2^{-p_i} \varepsilon_i.$$

Here in fact,

$$\begin{aligned} |r-s| &\geq 2^{-p_1} \\ |r-s| &< 2^{-(p_1-1)}. \end{aligned}$$

Now, the inequalities

$$\begin{aligned} |X(d_{i-1}) - X(d_i)| &\leq C|d_i - d_{i-1}|^\alpha \\ &= C|2^{-p_i}|^\alpha, i \geq 1 \end{aligned}$$

hold because each of these intervals is in fact

$$[d_{i-1}, d_i] = [k2^{-p_i}, (k+1)2^{-p_i}]$$

and $p_i \geq n \geq n(K)$ $\forall i$. Therefore, since X is continuous,

$$|X(r) - X(s)| \leq \sum_{i=1}^{\infty} |X(d_{i-1}) - X(d_i)|$$

$$\leq \sum_{i=1}^{\infty} C|d_i - d_{i-1}|^\alpha \leq \sum_{i=1}^{\infty} C|2^{-p_i}|^\alpha$$

$$\leq \sum_{i=p_1}^{\infty} C|2^{-i}|^\alpha = C(\alpha)2^{-p_1\alpha} \leq C(\alpha)|s-r|^\alpha.$$

It now remains to extend the obtained P-almost sure pathwise continuity

$$|X(t) - X(s)| \leq C|t-s|^\alpha,$$

over all $[s,t]$ with $|t-s| \leq K2^{-n(k)} = \delta(K)$ for a fixed K, to all intervals $[s,t] \subseteq [0,1]$. This is easily done by taking as a new constant

$$\tilde{C} = \max(C, \frac{M(K)}{\delta(K)})$$

where

$$M(K) = \sup\{\beta(s,t)| : |t-s| > \delta(K)\} < \infty.$$

Proof of (ii).

We need to prove here that if

$$E|\beta(s,t)|^q \leq C(q)|t-s|^{2q\gamma} \tag{18}$$

for a q such that $(\gamma - \alpha)2q > 1$, then P-almost surely

$$|\beta(s,t)| \leq C|t-s|^{2\alpha}, \ \forall [s.t].$$

Again, Chebyshev's inequality and the assumed moment sizes (18) give for a fixed K, an upper bound for $P(A_n)$ with $A_n = \{\omega \in \Omega :$ inequality $|\beta(s,t)| > |t-s|^{2\alpha}$ holds for at least one subinterval $[i2^{-n}, j2^{-n}]$ with$|j-i| \leq K\}$. We get

$$P(A_n) = P(\bigcup_{i,j:j-i \leq K} \{\beta(i2^{-n}, j2^{-n}) > C|j2^{-n} - i2^{-n}|^{2\alpha}\})$$

$$\leq \sum_{m=1}^{K} \sum_{i,j:|j-i|=\ldots} P\{\beta(i2^{-n}, j2^{-n}) > C|j2^{-n} - i2^{-n}|^{2\alpha}\}$$

$$\leq \frac{C(q)}{C^q} 2^{-n(\gamma-\alpha)2q} \ 2^n K^{(\gamma-\alpha)2q+1}$$

so that

$$\sum_{n=1}^{\infty} P(A_n)$$

converges if $(\gamma-\alpha)2q > 1$. Therefore, Borell-Cantelli Lemma gives P-almost surely, $|\beta(s,t)| \leq C|t-s|^{2\alpha}$ for all $[s,t] = [i2^{-n}, j2^{-n}], |j-i| \leq K$, when $n \geq n(K)$.

To extend this to general $[s,t]$, we will again choose $n \geq n(K)$ and i,j with $|j-i| \leq K$ such that

$$(i-1)2^{-n} < s \leq i2^{-n} < j2^{-n} \leq t < (j+1)2^{-n}.$$

Now the consistency rule (12) gives (cf. Chapter 7)

$$\beta(s,t) = \beta(s, i2^{-n}) + \beta(i2^{-n}, j2^{-n})$$

$$+ \ \beta(j2^{-n}, t) \ + \ (\textit{area of polygon}\mathcal{P}_1),$$

where $\mathcal{P}_1$ is the polygon on R^2 with vertices given by the points

$$(X_1, X_2)(s), \ (X_1, X_2)(i2^{-n}), \ (X_1, X_2)(j2^{-n}), \ (X_1, X_2)(t).$$

Again, result

$$|\beta(i2^{-n}, j2^{-n})| \leq C|j2^{-n} - i2^{-n}|^{2\alpha} \leq C|t-s|^{2\alpha}$$

has already been proved. For $\beta(s, i2^{-n})$ and $\beta(j2^{-n}, t)$ we will once again write these intervals as unions of dyadic intervals. To prove that, almost surely,

$$|\beta(s,r)| \leq C|r-s|^{2\alpha}$$

if $r = i2^{-n}$, $s \in R$, and $s < r$, $r - s < 2^{-n}$, we'll thus write

$$[s,r] = \bigcup_{i=1}^{\infty}[d_i, d_{i-1}]$$

where

$$d_0 = r, \quad d_k = r - \sum_{i=1}^{k} 2^{-p_i}\varepsilon_i$$

$$n < p_1 < p_2 < \dots, \varepsilon \; are \; 0 \; or \; 1, \varepsilon_1 = 1$$

and

$$r = d_0 > d_1 > \dots \to s.$$

Over each interval $[d_{i-i}, d_i]$, the results already established give

$$|\beta(d_i, d_{i-1})| \leq C|d_i - d_{i-1}|^{2\alpha}$$

$$= C|2^{-p_i}|^{2\alpha}.$$

By the continuity of β, the consistency rule and some geometrical arguments (cf Chapter 7) we get

$$\beta(s,r) = \sum_{i=1}^{\infty} \beta(d_{i-1}, d_i) + (area \; of \; polygon \; \mathcal{P}_2),$$

where $\mathcal{P}_2$ is the polygon on R^2, with vertices (possibly infinitely many) given by

$$(X_1, X_2)(d_0), \; (X_1, X_2)(d_1), \dots, (X_1, X_2)(s).$$

So,

$$|\beta(s,r)| \leq \sum_{i=1}^{\infty} |\beta(d_{i-1}, d_i)| + |area \; of \; polygon \; \mathcal{P}_2|$$

where

$$\sum_{i=1}^{\infty} |\beta(d_{i-1}, d_i)| \leq \sum_{i=1}^{\infty} C(2^{-p_i})^{2\alpha}$$

$$\leq \sum_{i=p_1}^{\infty} C2^{-i2\alpha} \ \leq \ C_1|s-r|^{2\alpha}.$$

For the area of the polygon $\mathcal{P}_2$, we need the continuity results for X_1, X_2 proved in (i). These give, a.s., for the edges of the polygon $\mathcal{P}_2$,

$$\|(X_1, X_2)(d_i) - (X_1, X_2)(d_{i-1})\|$$

$$\leq \ \sqrt{2}C|d_i - d_{i-1}|^{\alpha}.$$

Some simple geometrical observations, and the fact that

$$\sum_{i=1}^{\infty} |d_i - d_{i-1}|^{\alpha} \leq \sum_{i=p_1}^{\infty} 2^{-i\alpha} \leq C_2|r-s|^{\alpha}$$

produce an estimate:

$$|\mathit{area\ of\ polygon}\ \mathcal{P}_2| \leq C_3|r-s|^{2\alpha}.$$

It now remains to look at the area of polygon $\mathcal{P}_1$; that this is bounded by

$$|\mathit{area\ of\ polygon}\ \mathcal{P}_1| \leq C_4|t-s|^{2\alpha}$$

follows again directly from the continuity results for X_1 and X_2 obtained in (i). Extending the result of a.s. pathwise continuity from $[s,t]$ with $|t-s| \leq K2^{-n(K)}$ to any $[s,t] \subseteq [0,1]$ is again straightforward. □

Proof of Theorem 1, with general r would proceed similarly, proving the required continuity of β^k first for $k = 1$, then for $k = 2$, etc., up to $k = r$. For each k, we need to use the consistency assumptions between the β^ks, to write the value of β^k over [s,t] in terms of $\beta^1, \dots, \beta^k$ over the (dyadic) subintervals; and then we will use the already established continuity results for $\beta^1, \dots, \beta^{k-1}$. (Note that proving the required continuity for a β^k, expressed in terms of $\beta^1, \dots, \beta^{k-1}$ over subintervals, is identical to proving the corresponding continuity results given in Result 4.3, in Lyons [37].)

Note: Of course this result, based only on moments of the processes, can not give a sharp estimate for pathwise continuity. Similarly it would be pointless to look for an optimal value of the coefficient C. However, these estimates will be quite sufficient for our purposes.

Note: Here p, the number of moments that need to be controlled, depends on γ - i.e., the assumed continuity of the moments - and the desired pathwise continuity given by α. In results of Chapter 5 and Theorem 6.1 here, there is a tradeoff between α, r, p and γ as follows:

- For purposes of Chapter 5 we need an α -continuous system of iterated integrals up to r-th order, with $\alpha(r+1) > 1$. Thus assuming less continuity means having to take r higher, i.e. more iterated integrals will need to be initially provided.

- And then, using the Theorem 6.1 above to prove the pathwise α-continuity, we need γ-continuity for p-th order moments of the r iterated integrals, with $(\gamma - \alpha)p > 1$.

(More accurately, we need

$$E(|\beta^k(s,t)|^{(p/k)*}) \leq C|t-s|^{\gamma p}, k = 1\dots r). \tag{19}$$

In practice there will usually exist a maximal γ-value for which the inequality (19) for the moments can be expected to hold. To keep p small, one should choose an α, with $\alpha < \gamma$, as small as possible. Conversely, to obtain pathwise continuity with an α close to the optimal value γ, more and more moments would need to be controlled.

For instance, if $\gamma = 1/2$, then we could hope to prove pathwise α-continuity with any $\alpha < 1/2$, if sufficiently many moments of the processes β^k are controlled by (19). However, if $\alpha < 1/2$ then $r \geq 2$ will be needed anyway; i.e. the path X must be completed with an "area integral" β^2. But then, any $\alpha \in (1/3, 1/2)$ will work as well with $r = 2$, and for smaller α, a smaller value of p will be required. Thus it would be quite sufficient to prove, for X and β^2, almost sure pathwise continuity with $\alpha = 1/3 + \epsilon$, with any $\epsilon > 0$. To establish this, according to Theorem 6.1 it suffices to assume that

$$\begin{aligned} E|X_i(t) - X_i(s)|^8 &\leq C|t-s|^4, \\ E|\beta_{ij}^2(s,t)|^4 &\leq C|t-s|^4 \end{aligned}$$

for all i, j, $[s,t] \subseteq [0,1]$. From these it follows that (X, β^2) -system is α -continuous with any $\alpha < 3/8$.

6.3 Brownian motion

A non-trivial example where our results will apply is the case where X is an n-dimensional Brownian motion, defined on some filtered probability space $(\Omega, P, \mathcal{F}, (\mathcal{F}_t, t \geq 0))$. It is well known that the exact pathwise continuity of X is given by Levy's modulus of continuity (see for example Knight [29], McKean [25]):

Result 6.3 *For a one-dimensional Brownian motion, P-almost surely*

$$\lim_{|t-s|\to 0} \sup_{0<s,t<1} \frac{|X(t)-X(s)|}{\sqrt{2|t-s|\log|t-s|}} = 1. \tag{20}$$

This means that a.a. paths $X(\omega)$ are Holder(α) -continuous, if and only if $\alpha < 1/2$.

To utilize our results, we have to provide almost every path $X(\omega)$ with an "area integral" $\beta^2(\omega)$ of suitable continuity. The β^2 will not be uniquely defined, of course, since $\alpha < 1/2$.

Now, there exists an obvious candidate for β^2, namely P. Levy's stochastic area of Brownian motion. There are several ways of constructing this area, all of which rely on probabilistic arguments to prove its existence. (See Levy, [33], [34], [35].) Here we define the stochastic area via stochastic integrals; see Section 7.1 for more details.

Definition 6.4 *The stochastic area of Brownian motion is defined to be the 2-parameter continuous stochastic process*

$$\beta^2_{ij}(s,t) = \frac{1}{2}\int_s^t \int_s^u (dX_i(v)dX_j(u) - dX_j(v)dX_i(u))$$

$$for\ i,j = 1\ldots n,\ \ [s,t] \subseteq [0,1],$$

where the iterated integrals are either Ito- or Stratonovich integrals.

That almost every $\beta^2(\omega)$ is an area, i.e. obeys the consistency laws (12) is obvious. Proving that it has the required pathwise continuity properties is easy, using martingale properties of the stochastic integrals.

Lemma 6.5 *For all* $[s,t] \subseteq [0,1]$, $i,j = 1 \dots n$.

$$E|\beta^2_{i,j}(s,t)|^4 \leq C|t-s|^4$$

Proof. If X and Y are (independent) Brownian motions, then basic results from the theory of martingales and stochastic integration given in Chapter 2 give

$$\begin{aligned}
E|\int_s^t (X_u - X_s)dY_u|^4 &\leq C_1\ E|\langle\int_s^\cdot (X_u - X_s)dY_u\rangle_t|^2 \\
&= C_1 E|\int_s^t (X_u - X_s)^2 du|^2 \\
&= C_1 \left(\int_s^t \int_s^t E\left[(X_u - X_s)(X_v - X_s)\right] dudv\right)^2 \\
&= C_2 (\int_s^t (u-s)du)^2 \\
&\leq C_3 |t-s|^4,
\end{aligned}$$

And, of course,

$$E|\beta^2_{ij}(s,t)|^4 \leq C_4(|I_1|^4 + |I_2|^4)$$

where

$$\begin{cases} I_1 = \int_s^t (X_u - X_s)dY_u \\ I_2 = \int_s^t (Y_u - Y_s)dX_u \end{cases}$$

□

Now, from Theorem 6.2 follows immediately

Corollary 6.6 *For any $\alpha < 3/8$, P-almost surely,*

$$|\beta_{ij}^2(s,t)| \leq C|t-s|^{2\alpha}$$

for all $[s,t] \subseteq [0,1]$, and all $i,j = 1 \ldots n$.

So now, for almost all $\omega \in \Omega$, a Brownian path $X(\omega)$ together with its stochastic area $\beta^2(\omega)$ forms an (α, r) -system in the sense of Chapter 5, with $\alpha(r+1) > 1$, and the method given there for constructing path by path a solution to the differential equation can be attempted. We will next prove that for almost every $\omega \in \Omega$, this method will work, i.e. the sequence $\theta^{(n)}(0,1)(\omega)$ will converge. Here, the convergence can be easily proved by using the results in Chapter 5.5, and the fact that for X a Brownian motion, we already have the theory of stochastic differential equations at our disposal. (We are once again looking at the differential equation of the form

$$dY_t = \sum_{i=1}^{n} f_i(Y_t) dX_t^i.$$

For simplicity we will assume in the next that $f_i \in C_b^3$, i.e. that the vector fields and their first three derivatives are not only continuous but also bounded. However, a similar but "local" version of the result holds for differential equations with coefficients in C^3.)

Theorem 6.7 *Let, for each $\omega \in \Omega$, $(X(\omega), \beta(\omega))$ consist of a Brownian motion enhanced with its Levy area (both assumed to be continuous functions). Look at the differential equation above, where the coefficients (f_i) are C_b^3-continuous vector fields on R^d. Let $\theta^{(n)}(0,1)(\omega)$ be the sequence of operators defined as in Section*

5.1. Then, for almost every $\omega \in \Omega$,

$$\sup_{y \in R^d} \|\theta^{(n)}(0,1)(\omega)(y) - \hat{\theta}(0,1)(\omega)(y)\| \to 0,$$

where $\hat{\theta} = (\hat{\theta}(s,t),\ [s,t] \subseteq [0,1])$ is the stochastic flow corresponding to the solution of the differential equation interpreted as a Stratonovich differential equation.

Proof. According to Theorem 5.13, it is sufficient to prove that for almost every $\omega \in \Omega$,

$$\|\hat{\theta}(s,t)(\omega)(y) - \exp(z[s,t])(y)\| \leq C\,|t-s|^{3\alpha}, \quad \forall y \in R^d,$$

for some $\alpha < 3/8$. (Here, of course,

$$z[s,t] = \sum_i (X_i(t) - X_i(s))\, f_i + \sum_{i<j} \beta_{ij}(s,t)[f_i, f_j].)$$

This can easily be established by comparing simple approximations for the solutions of the SDE and ODE corresponding to the operators involved here. Firstly, for $\hat{\theta}$, the solution of the SDE, we will get by using a straightforward second order Taylor expansion with respect to the initial point y that almost surely, for all $[s,t]$ and all y, for $Y_t = \hat{\theta}(s,t)(y)$: the solution of the differential equation at time t, starting from point y at time s,

$$Y_t - y = \sum_i (X_i(t) - X_i(s))\, f_i(y) + \sum_{i<j} \beta_{ij}(s,t)[f_i, f_j](y)$$

$$+ \frac{1}{2} \sum_{i,j} (X_i(t) - X_i(s))(X_j(t) - X_j(s))\, (f_j\, Df_i)(y) \ + \ r_1(s,t;y).$$

Here $r_1(s,t;y)$ consists of iterated Stratonovich integrals, and its size can be estimated, using arguments from the theory of martingales and stochastic integration to give:

$$\forall y, \quad \|r_1(s,t;y)\| \leq K(\omega)|t-s|^{\varepsilon}$$

for any $\varepsilon < 3/2$.

Secondly, for a $y \in R^d$, if $y(1) = \exp(z[s,t])(y)$ is the solution at time 1 of the ODE

$$dy_u = z[s,t](y_u)du, \quad y(0) = y$$

then $y(1)$ equals $x(t-s)$, the solution at time $t-s$ of the ODE

$$dx_u = \frac{1}{t-s} z[s,t](x_u)du, \quad x(0) = y.$$

And for the solution of this ODE, a second order Taylor expansion gives

$$x(t-s) - y = \sum_i (X_i(t) - X_i(s))\, f_i(y) + \sum_{i<j} \beta_{ij}(s,t)[f_i, f_j](y)$$

$$+ \frac{1}{2} \sum_{i,j} (X_i(t) - X_i(s))(X_j(t) - X_j(s))\, (f_j\, Df_i)(y) \;+\; r_2(s,t;y),$$

where again, given the assumed continuities of the X_is and the β_{ij}s, the remainder term can easily be proved to be bounded by

$$\forall y, \quad \|r_2(s,t;y)\| \leq K(\omega)|t-s|^{\varepsilon}$$

for any $\varepsilon < 3/2$. □

Thus, using the construction of Chapter 5 for almost all Brownian paths, with the area integrals chosen to be the paths of the stochastic area integrals, gives a solution which agrees almost surely with the "classical" Stratonovich solution. Here, though, the construction of the solution is completely pathwise. (Note that we have established that the path property, possessed by almost every Brownian path, that makes the theory of differential equations possible with respect to them, is the fact that the stochastic area for Brownian motions exists, and has sufficient continuity properties.)

Also, in our approach the method of Picard iterations to construct a solution has been replaced by an alternative method. Note also that this leads to an alternative discretization scheme to obtain numerical approximations for solutions of SDEs. The result in Section 5.5 proves the convergence of these numerical methods towards the solution of the SDE. (Cf. the traditional numerical methods, e.g. in Rumelin [52] — it is a well recognized fact that including iterated path integrals into the discretization scheme leads to improved convergence of numerical methods. Related work is also the results in Sussmann [54], [55], where an expression for the solution of an SDE is given as an infinite product of exponentials of the vector fields multiplied by iterated integrals, which also leads to alternative discretization schemes.)

6.4 Other stochastic processes

As well as Brownian paths, obviously also almost all paths of more general martingales could easily be incorporated into our theory. The area can for them again be defined using the stochastic integrals. To prove the pathwise continuity of the areas we again use Theorem 6.2; since the stochastic integrals are martingales, the necessary moment conditions for them are easy to obtain, even if the calculations are a bit more complicated since for a general martingale, the quadratic variation process is itself a stochastic process.

One very interesting special case where our results would be extremely useful is the theory of Dirichlet processes. (See e.g. Bertoin [4],[5], Fukushima [18], Lyons-Zheng [38], [39], [40].) A Dirichlet process can be written as a sum of a backwards and a forwards martingale, and some stochastic Stratonovich type integrals with respect to them can be defined as differences of backwards and forwards driven stochastic Ito integrals, so their behaviour is well controlled by the martingale theory. However, the fact that their martingale representation involves two filtrations, one backwards in time and the other forwards in time, means that the Ito theory of stochastic differential equations can not be applied to them.

However, the results in Chapter 5 can easily be applied, since the area integral happens to be one of the integrals which are well defined for the Dirichlet process. We only have to establish that almost all paths of the Dirichlet process and its area integral have the appropriate path continuities; and for this we can use the moment result established above. The moments of the integrals with respect to Dirichlet processes can be reduced to moments of martingale integrals, which are of course easily controlled.

7 About areas

7.1 Classical and generalised areas

Let $\Gamma = ((X_t, Y_t), t \in [0,1])$ be a continuous curve in R^2. By the area of the curve Γ over the time interval $[0,1]$, denoted by $\beta[0,1]$, we mean in this thesis the area included by the curve Γ and its chord from (X_0, Y_0) to (X_1, Y_1). The convention here is that counter clockwise loops around a region are counted as positive, and clockwise ones as negative.

The area can formally be represented by integral notation

$$\beta(0,1) = 1/2 \int_0^1 \int_0^t [dX_s dY_t - dY_s dX_t]. \tag{21}$$

For well-behaved paths this, interpreted as a difference of two Lebesque-Stieltjes integrals, defines the area. For a more general definition, denote:

$$\begin{aligned} \mathcal{T} \quad = \quad & \text{The set of all sequences } (\tau_n, n \in N) \text{ of partitions of } [0,1], \\ & \text{with } \tau_0 = \{0,1\} \subset \tau_1 \subset \cdots \\ & \text{and } \lim_{n\to\infty} \|\tau_n\| = 0 \end{aligned}$$

where

$$\|\tau_n\| = 2 \max_{0<t<1} \min_{s\in\tau_n} |t-s|$$

is the maximal step length of the τ_n-partition. For $(\tau_n, n \in N) \in \mathcal{T}$, define Γ_n

to be the piecewise linear approximation of Γ along partition τ_n. That is, Γ_n coincides with Γ at all partition points of τ_n, and is linear between these points. The area of Γ_n over [0,1], denoted by $\beta_n[0,1]$, is well defined.

Definition 7.1 *(Cf. Levy [33].) If for all $(\tau_n, n \in N) \in \mathcal{T}$, the limit*

$$\beta(0,1) = \lim_{n\to\infty} \beta_n(0,1)$$

exists and does not depend on the choice of $(\tau_n, n \in N)$, then we say that the area of Γ over [0,1] exists in the classical sense and is given by $\beta(0,1)$; in this case the integral (21) is well defined.

Note that in this case, it is also possible to write the area as

$$\beta(0,1) = \sum p_n\, s_n,$$

where

- s_n is the area of a region, circled by the curve Γ and its chord;
- p_n is a number, positive, negative, or zero, which tells how many times Γ goes around the area s_n, and in which way (counter-clockwise or clockwise).

If the area exists in the classical sense, then this series is absolutely convergent.

If a curve Γ has an area over [0,1] in the classical sense, then the definition above extends trivially to define the area of the curve Γ over any interval $[s,t] \subseteq [0,1]$. (By this is meant of course the area included by the curve section $(X_u, Y_u), u \in [s,t])$ and its chord from (X_s, Y_s) to (X_t, Y_t).) In this case we have again a well-defined integral representation

$$\beta(s,t) = \int_s^t \int_s^u (dX_v dY_u - dY_v dX_u). \tag{22}$$

The summation rule

$$\beta(s,u) = \beta(s,t) + \beta(t,u) \tag{23}$$

$$+ \frac{1}{2}((X_t - X_s)(Y_u - Y_t) - (Y_t - Y_s)(X_u - X_t))$$

$$\forall\, 0 < s < t < u < 1$$

follows immediately; note that the last term here is the area of the triangle with vertices $(X_s, Y_s), (X_t, Y_t), (X_u, Y_u)$. More generally, from (23) it follows that

$$\beta(t_0, t_n) = \beta(t_0, t_1) + \beta(t_1, t_2)$$

$$+ \cdots + \ \beta(t_{n-1}, t_n) \ + \ area(\mathcal{P})$$

$$\forall\, 0 < t_0 < t_1 < \cdots < t_n < 1$$

where $\mathcal{P}$ is the polygon in R^2 with vertices $(X_{t_0}, Y_{t_0}), \ldots, (X_{t_n}, Y_{t_n})$.

Of course there are many continuous paths for which this definition of the area does not apply. The interpretation of the area as a two-parameter function leads to a more general concept of area.

Definition 7.2 *A generalized area of the curve* $\Gamma = ((X_t, Y_t), t \in [0,1])$ *is any two-parameter real-valued function* $(\beta(s,t), 0 < s, t < 1)$ *for which (23) holds.*

Note that a generalized area of a curve is not unique:

- If β is a generalized area of a curve Γ then so is $\hat{\beta}$ where $\hat{\beta}(s,t) = \beta(s,t) + \gamma(t) - \gamma(s)$ and γ is an arbitrary real valued function;
- If β and $\hat{\beta}$ are both generalized areas of Γ, then $\beta(s,t) - \hat{\beta}(s,t) = \gamma(t) - \gamma(s)$ for some function γ.

A typical case where an area exists in the generalized sense but not in the classical sense is when approximating Γ with different sequences of polygons can lead to different values at limit for the area.

Definition 7.3 *If the limit*

$$\beta_\tau(0,1) = \lim_{n\to\infty} \beta_n(0,1)$$

exists for a given $\tau = (\tau_n, n \in N)$, then we will call it the area of Γ over [0,1] with respect to τ. Definition of $\beta_\tau(s,t), [s,t] \subseteq [0,1]$ follows trivially, and (23) holds.

Again in this case, the integrals (22) are well defined, this time as a limit of Riemannian sums with respect to the τ_n-partitions.

7.2 Stochastic areas; P. Levy's area for Brownian motion

In the stochastic set-up, with a given filtered probability space $(\Omega, P, \mathcal{F}, (\mathcal{F}_t, t \geq 0))$, we will look at a curve Γ given by a two-dimensional stochastic process:

$$\Gamma(\omega) = (X_t(\omega), Y_t(\omega)),\ t \in [0,1],\ \omega \in \Omega.$$

A typical example is the two-dimensional Brownian motion. It is easy to see that for P-almost every Brownian path, the definition of area in the classical sense does not apply: For a.e. path, it is possible to choose subdivisions $\tau \in \mathcal{T}$ such that the polygonal areas do not converge. Of course the problem here arises because the choice of $\tau = \tau(\omega)$ is permitted to depend on the sample path Γ.

For a better definition of area in the stochastic case, we will define (cf. Levy [33]) $\tau = \tau(\omega)$ to be a random sequence of subdivisions over [0,1] as follows: Choose the division points of τ by defining $t_0 = 0$, $t_1 = 1$, and then generating infinitely many division points $t_2, t_3, \ldots$ independently of each other, each with uniform distribution over [0,1]. Take now

$$\begin{aligned} \tau_n &= (\{t_0, \ldots, t_n\} \text{ in increasing order}), \\ \Gamma_n &= \text{the polygonal approximation of } \Gamma \text{ along } \tau_n, \\ \beta_n(0,1) &= \text{area of } \Gamma_n \text{ over [0,1] (a random variable).} \end{aligned}$$

Definition 7.4 *If*

$$\beta_n(0,1) \to \beta(0,1)$$

with convergence in probability, in square mean, or almost surely, then we call $\beta(0,1)$ the stochastic area of Γ over [0,1] (in probability, in square mean, or almost surely, respectively).

For Brownian motion, the following holds: (See Levy [33],[34])

Theorem 7.5 *Assume that the curve Γ is given by a two-dimensional Brownian motion over [0,1]. For any fixed $\tau \in \mathcal{T}$, for all $[s,t] \subseteq [0,1]$, the sequence $\beta_n(s,t)$ converges in square mean and almost surely as $n \to \infty$. Also, the limit, denoted by $\beta(s,t)$, does not depend on the choice of τ: If β is the limit along $\tau \in \mathcal{T}$ and $\hat{\beta}$ is the limit along $\hat{\tau} \in \mathcal{T}$ then almost surely $\beta = \hat{\beta}$.*

Here, for almost every $\omega \in \Omega$, $\beta(s,t)(\omega)$ can be extended to a continuous function of s, t. Especially then, if we fix any $\tau = (\tau_n, n \in N) \in \mathcal{T}$ then for almost every path of a two-dimensional Brownian motion the area with respect to τ does exist in the sense of Definition 7.3; and the outcome is a.s. independent of which (deterministic) sequence of subdivisions $\tau \in \mathcal{T}$ we choose.

Here is a brief outline of the proof of the theorem above:
We will look at the case of dyadic subdivisions (the general case is similar). So, we denote by Γ_n the polygonal approximation of Γ along division points ($k/2^n$, $k = 0 \dots 2^n$). To prove convergence of the corresponding areas β_n, we will write them as

$$\beta_n = \sum_{k=1}^{n-1} A_k$$

where A_k is the area between the successive polygonal approximations Γ_k and Γ_{k+1}. Then each A_n is the sum the areas of 2^n triangles; all these areas of triangles are random variables, independent of each other and with expected value 0 and variance 2^{-2n-3}.

(Since Γ_{k+1} is obtained from Γ_k by adding a new point between each pair of subsequent partition points, the areas of the triangles obtained are independent of each other. As for the moments of them, note that the basis of each triangle is the vector between two subsequent points on Γ_k, thus it is the the increment of the 2-dimensional Brownian motion during a time interval of length 2^{-k} — so,

$$E[(length\ of\ basis)^2] = 2 \cdot 2^{-k}.$$

As for the height of the triangle, the new point added equals: the midpoint in R^2 of the two old points, plus an increment $(\triangle x, \triangle y)$ where $\triangle x$ and $\triangle y$ are independent identically distributed random variables with law $N(0, 2^{-n}/4)$ — thus the distance of the new point from the basis of the triangle is also $N(0, 2^{-n}/4)$-distributed.)

Summing up these areas gives

$$E(A_n) = 0, \qquad E(A_n^2) = 1/2^{n+3}$$

which is enough to get convergence of the series β_n in quadratic mean. Chebychev's inequality gives now

$$P\left(|A_n| > \frac{1}{2^{(n+6)/4}}\right) \leq \frac{1}{2^{n/2}}$$

from which, by Borel-Cantelli lemma, it can be seen that β_n converges even almost surely.

There are alternative probabilistic definitions for the stochastic area of Brownian motion; all these versions of the area agree almost surely. In a Fourier series approach (see Levy [35]) we write the two independent Brownian motions as

$$\begin{cases} X(t) &= \frac{\alpha' t}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \frac{1}{n\sqrt{\pi}} \left[\alpha_n(\cos nt - 1) + \alpha_n'(\sin nt)\right] \\ Y(t) &= \frac{\beta' t}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \frac{1}{n\sqrt{\pi}} \left[\beta_n(\cos nt - 1) + \beta_n'(\sin nt)\right] \end{cases}$$

where α',α_n,α_n', β',β_n,β_n' are all N(0,1) -distributed random variables, and are all independent of each other. Then the stochastic area of $(X(t), Y(t))$ over $[0, 2\pi]$ can be directly defined by the Fourier series as

$$\beta(0, 2\pi) = \sum_{n=1}^{\infty} \frac{1}{n} \left[\alpha_n(\beta_n' - \beta'\sqrt{2}) - \beta_n(\alpha_n' - \alpha'\sqrt{2})\right]$$

where the Fourier series converges almost surely.

Finally, the area of Brownian motion can be simply written by using the theory of stochastic integration: P-almost surely,

$$\beta(s,t) = \frac{1}{2}\int_s^t \int_s^u (dX_v dY_u - dY_v dX_u),$$

where the iterated integrals are either Ito or Stratonovich integrals; they give the same result in this case, since the quadratic covariation terms cancel out:

$$\begin{aligned} & \int_0^t (X_u - X_s) \circ dY_u - \int_0^t (Y_u - Y_s) \circ dX_s \\ = & \left(\int_0^t (X_u - X_s) \cdot dY_s - \frac{1}{2}\langle X, Y, \rangle_t\right) \\ - & \left(\int_0^t (Y_u - Y_s) \cdot dX_s - \frac{1}{2}\langle X, Y, \rangle_t\right) \\ = & \int_0^t (X_u - X_s) \cdot dY_u - \int_0^t (Y_u - Y_s) \cdot dX_s. \end{aligned}$$

(This integral representation for the area reflects of course the fact that for any fixed $\tau \in \mathcal{T}$, stochastic integrals with respect to Brownian motions equal almost surely the limit of the appropriate Riemannian sums along the τ_n-partitions.)

The law of P. Levy's stochastic area is well known, and its scaling properties and its first two moments are easy to obtain:

$$E[\beta(s,t)] = 0, \qquad E[\beta(s,t)]^2 = \frac{t-s}{4}.$$

(See for example Levy [33], [34], [35], and Gaveau [19] for details.) We obtain some further upper bounds for higher order moments of it, and prove a result about almost sure pathwise continuity of the area $\beta(s,t)$ in Chapter 6.

References

[1] AZENCOTT R.
Formule de Taylor stochastique et developpements asymptotiques d'integrales de Feynman, in *Seminaire de Probabilites XVI*, LNM 921, Springer-Verlag, 1980-81, pp. 237-284

[2] BALLY V.
Approximation for the solutions of stochastic differential equations I: L^p-convergence, *Stochastics and Stochastics reports*, vol. 28, pp. 209-246 (1989)

[3] BALLY V.
Approximation for the solutions of stochastic differential equations II: Strong convergence, *Stochastics and Stochastics reports*, vol. 28, pp. 357-385 (1989)

[4] BERTOIN J.
Les processus de Dirichlet et tant qu'espace de Banach, *Stochastics*, vol. 18, pp. 155-168 (1986)

[5] BERTOIN J.
Applications des processus de Dirichlet aux temps locaux et temps d'intersection d'un mouvement Brownien, *Probab. Th. Rel. Fields* 80, pp. 433-460 (1989)

[6] BEN AROUS G.
Flots et series de Taylor stochastiques, *Probab. Th. Rel. Fields* 81, pp. 29-77 (1989)

[7] BOURBAKI N.
Elements of Mathematics, Lie groups and Lie algebras, Part I, Hermann, Paris 1975

[8] CASTELL F.
Developpement asymptotique de flots stochastiques, *Preprint*, 1991

[9] CHALEYAT-MAUREL M.,LE GALL J.-F.
Green function,capacity and sample path properties for a class of hypoelliptic diffusion processes, *Probability theory and related fields* 83, pp. 219-264 (1989)

[10] CHEN K.-T.
Integration of paths, geometric invariants and a generalised Baker- Hausdorff formula, *Annals of Mathematics* 65, pp. 163-178 (1957)

[11] CHEN K.-T.
Integration of paths-A faithful representation of paths by noncommutative formal power series,*Trans. Ann. Math. Soc.*89, pp. 395-407 (1958)

[12] DELLACHERIE C. AND MEYER P.A.
Probabilités et Potentiel Chapters V-VIII, Herman, Paris 1980

[13] DEVLIN J.
Word problems related to periodic solutions of a non-autonomous system, *Math. Proc. Camb. Phil. Soc.* 108, pp. 127-151 (1990)

[14] DOSS H.
Liens entre equations differentielles stochastiques et ordinaires, *Ann. Inst. H. Poincare* 13, pp. 99-125 (1977)

[15] DURRETT R.
Brownian motions and martingales in analysis, Wadsworth, Belmont, California 1984

[16] EMERY M.
Equations differentielles stochastiques lipschitziennes: Etude de la stabilite, in: *Seminaire de Probabilites XIII*, LNM 721, Springer-Verlag, New York 1979, pp. 281-293

[17] FOLLMER H.
Calcul d'Ito sans probabilites, in: *Seminaire de Probabilites XV*, LNM 850, Springer, Berlin 1981, pp. 143-150

[18] FUKUSHIMA M.
Dirichlet forms and Markov processes, Kodansha, Tokyo 1980

[19] GAVEAU B.
Principle de moindre action, propagation de la chaleur et estimees sous-elliptiques sur certains groupes nilpotents, *Acta mathematica* 139, pp. 96-153 (1977)

[20] GAVEAU B.,VAUTHIER J.
The Dirichlet problem for the subelliptic Laplacian on the Heisenberg group II, *Can. J. Math.*, vol. 37, no. 4, pp. 760-766 (1985)

[21] GAVEAU B.,GREIMER P.,VAUTHIER J.
Integrals de Fourier quadratiques et calcul symbolique exact sur le groupe d'Heisenberg, *Journal of Functional Analysis* 68, pp. 248-272 (1986)

[22] HAUSNER M., SCHWARTZ J.T.
Lie groups — Lie algebras, Notes on Mathematics and its Applications, Nelson, 1968

[23] IKEDA N., NAKAO S., YAMATO Y.
A class of approximations of Brownian motion, Publ. RIMS, Kyoto Univ. 13 1977, pp. 285-300

[24] IKEDA N.,WATANABE S.
Stochastic Differential Equations and Diffusion Processes, North-Holland – Kodansha, Amsterdam and Tokyo 1981

[25] ITO K., MCKEAN H.P.
Diffusion processes and their sample paths, Springer-Verlag, Berlin 1964

[26] JACOBSON N.
Lie Algebras "Interscience", Wiley, New York 1962

[27] KLOEDEN P.E., PLATEN E.
Stratonovich and Ito stochastic Taylor expansions, *Math. Nachr.* 151, pp. 33-50 (1991)

[28] KLOEDEN P.E., PLATEN E.
Numerical Solutions of Stochastic Differential Equations, Springer-Verlag, Berlin 1992

[29] KNIGHT F.B.
Essentials of Brownian Motion and Diffusion, American Mathematical Society, Providence, Rhode Island 1981

[30] KONECNY F.
On Wong-Zakai approximation of stochastic differential equations, *Journal of multivariate analysis* 13, pp. 605-611 (1983)

[31] KUNITA H.
On the representation of solutions of stochastic differential equations, in: *Seminaire de Probabilites XIV*, LNM 784, Springer-Verlag, Berlin 1978-79, pp. 282-304

[32] KUNITA H.
On the decomposition of solutions of stochastic differential equations, in: *Stochastic Integrals* (ed. D.Williams), LNM 851, Springer-Verlag, Berlin 1981, pp. 213-255

[33] LEVY P. Le mouvement Brownien plan, *Amer. Jour. Math.*, vol. 62, pp. 487-550 (1940)

[34] LEVY P.
Processes Stochastiques et Mouvement Brownien, Gauthier-Villais, Paris 1948

[35] LEVY P.
Wiener's random function, and other Laplacian random functions, in *Proceedings of the second Berkeley Symposium of Probability and Statistics*, pp. 171-187 (1950)

[36] LYONS T.J.
On the non-existence of path integrals (preprint) 1991

[37] LYONS T.J.
The interpretation and solution of ordinary differential equations driven by rough signals (preprint) 1993

[38] LYONS T.J., ZHENG W.A.
On conditional diffusion processes, *Proc. Roy. Soc. Ed.* A, vol. 115, pp. 243-255 (1990)

[39] LYONS T.J. WEIAN Z., ZHENG W.
A crossing estimate for the canonical process on a Dirichlet space and a tightness result, *Asterisque* 157-158, pp. 249-271 (1988)

[40] LYONS T.J., ZHENG W.A.
Diffusion processes with non-smooth diffusion coefficients and their density functions, *Proc. Roy. Soc. Ed.* A, vol. 115, pp. 231-242 (1990)

[41] MARCUS S.I.
Modelling and approximation of stochastic differential equations driven by semimartingales, *Stochastics*, vol. 4, pp. 223-245 (1981)

[42] MCKEAN H.P.
Stochastic Integrals, Academic Press, New York 1969

[43] MCSHANE E.J.
Stochastic differential equations and models of random processes, in: *Proc. of Sixth Berkeley Symposium of Math. Stat. and Prob. III*, pp. 263-294 (1972)

[44] MCSHANE E.J.
Stochastic Differential Equations,*Journal of Multivariate Analysis* 5, pp. 121-177 (1975)

[45] MELANCON G., REUTENAUER C.
Lyndon words, free algebras and shuffles, *Can. J. Math.*, vol. XLI, no. 4, pp. 577-591 (1989)

[46] MEYER P.A.
Un cours sur les integrales stochastiques, in: *Seminaire de Probabilites X*, LNM 511, Springer-Verlag, Berlin, pp. 245-400 (1976)

[47] NAKAO S. AND YAMATO Y.
Approximation theorem on stochastic differential equations, in: *Proc. of the International Symp. on Stochastic Differential Equations* (ed. K.Ito), New York, Wiley 1978, pp. 283-296

[48] PROTTER P.
On the existence,uniqueness, convergence and explosions of solutions of systems of stochastic integral equations, *Ann. Prob.* 5, pp. 243-261 (1977)

[49] PROTTER P.
Hp Stability of solutions of stochastic differential equations,*Z. fur W. Verw. Gebiete* 44, pp. 337-352 (1978)

[50] PROTTER P.
Approximations of solutions of stochastic differential equations driven by semimartingales, *The Annals of Probability*, vol. 13, no. 3, pp. 716-743 (1985)

[51] ROGERS L.C.G., WILLIAMS D.
Diffusions, Markov Processes and Martingales, vol.2:Ito Calculus, John Wiley & sons, New York 1987

[52] RUMELIN W.
Numerical treatment of stochastic differential equations, *SIAM J. Anal.*, vol. 19, no. 3, pp. 604-613 (1982)

[53] SUSSMAN H.J.
On the gap between deterministic and stochastic ordinary differential equations, *The Annals of Probability*, vol. 6, no. 1, pp. 19-41 (1978)

[54] SUSSMAN H.J.
A product expansion for the Chen series, in: *Theory and Applications of Nonlinear Control Systems* (ed: C.I.Byrnes, A.Linquist), North-Holland 1986, pp. 323-335

[55] SUSSMAN H.J.
Product expansions of exponential Lie series and the discretization of stochastic differential equations, in: *Stochastic Differential Systems, Stochastic Control Theory, and Applications* (ed. W.Fleming, J.Lions), Springer IMA Series, vol. 10, Springer-Verlag,Berlin, 1988, pp. 563-582

[56] STRATONOVICH R.L.
A new form of representation of stochastic integrals and equations, *SIAM J. Control*, vol. 4, no. 2, pp. 362-371 (1966)

[57] STRICHARTZ R.S.
The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations, *Journal of Functional Analysis* 72, pp. 320-345 (1987)

[58] WILLIAMS D.
Diffusions, Markov Processes, and Martingales, vol.1: Foundations, John Wiley & sons, Chichester, New York 1979

[59] WILLIAMS D.(ed)
Stochastic Integrals, Proceedings, LMS Durham Symposium, LNM 851, Springer-Verlag, Berlin 1981

[60] WONG E., ZAKAI M.
On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Stat.* 36, pp. 1560-64 (1965)

[61] WONG E., ZAKAI M.
On the relation between ordinary and stochastic differential equations, *Intern. J. Eng. Sci.* 3, pp. 213-229 (1965)

[62] WONG E., ZAKAI M.
Riemann-Stielties approximations of stochastic integrals, *Z. fur W. Verw. Gebiete* 12, pp. 87-97 (1969)

[63] YAMATO Y.
Stochastic differential equations and nilpotent Lie algebras, *Z. fur W. Verw. Gebiete* pp. 213-229 (1979)

[64] YOUNG L.C.
Some new stochastic integrals and Stieltjes integrals Part I, in: *Advances in Probability* 2, (ed. P.Ney), M.Dekker Inc., New York 1970 pp. 161-239

[65] YOUNG L.C.
Some new stochastic integrals and Stieltjes integrals Part II: Nigh-Martingales in:*Advances in Probability* 3 (ed. P.Ney, S.Port), M.Dekker Inc., New York 1974, pp. 101-177