Revista Matemática Iberoamericana
Vol. 14, N.$^{\text{o}}$ 2, 1998

Differential equations driven by rough signals

Terry J. Lyons

1. Preliminaries.

1.1. Introduction.

1.1.1. Inhomogeneous differential equations.

Time inhomogeneous (or non-autonomous) systems of differential equations are often treated rather formally as extensions of the homogeneous (or autonomous) case by adding an extra parameter to the system; however this can be a travesty. Consider an equation of the kind

$$dy_t = \sum_i f^i(y_t)\, dx_t^i\,, \tag{1.1}$$

where the f^i are vector fields, x_t represents some (multi-dimensional) forcing or controlling term and the trajectory y_t represents some filtered effect thereof. In this case the effect of such a reduction produces an equation whose expression involves a derivative of the term x_t. In problems from control, or where noise is involved, or even in algebra (developing a path from a Lie algebra into a group) this path will rarely be smooth, so the resulting autonomous system will have a defining vector field which will frequently not be continuous; perhaps it will only exist as a distribution. In this case the classical theory does not suggest the correct approach to identifying solutions; and even in highly oscillatory but smooth situations suggests inefficient algorithms for numerical approximation to classical solutions.

1.1.2. Objectives.

This paper aims to provide a systematic approach to the treatment of differential equations of the type described by (1.1) where the driving signal x_t is a rough path. Such equations are very common and occur particularly frequently in probability where the driving signal might be a vector valued Brownian motion, semi-martingale or similar process.

However, our approach is deterministic, is totally independent of probability and permits much rougher paths than the Brownian paths usually discussed. The results here are strong enough to treat the main probabilistic examples and significantly widen the class of stochastic processes which can be used to drive stochastic differential equations. (For a simple example see [10], [1]).

We hope our results will have an influence on infinite dimensional analysis on path spaces, loop groups, etc. as well as in more applied situations. Variable step size algorithms for the numerical integration of stochastic differential equations [8] have been constructed as a consequence of these results.

1.1.3. The Itô map.

Suppose every vector field f^i in (1.1) is Lipschitz with respect to some complete metric on a manifold M and that the driving signal x_t is continuous and piece-wise smooth; then classical solutions to (1.1) exist for all time and are unique; by fixing y_0, we may regard (1.1) as defining a functional (which we will refer to as the Itô map) taking each smooth path x_t (in a certain vector space V) to a unique path based at y_0 in a manifold M. By varying the starting point y_0 and taking the induced flow, one may also regard (1.1) as defining a map taking the path x_t to a path in the group of homeomorphisms of M.

We would like to extend this Itô map to a far richer class of paths. Our intention is to identify a family of metric topologies on smooth paths for which the Itô map is uniformly continuous (and even differentiable although we cannot show this here [17], [18], [19]). A point in the completion of the smooth paths in one of these metrics corresponds to a path in V with proscribed low order integrals and having finite p-variation for some $p < \infty$. As a first application we have the theorem that the solution to a Stratonovich stochastic differential equation of the classical type is a continuous function of the driving Wiener process and Lévy area taken as a pair.

1.1.4. The fundamental problem: Lack of continuity.

Before we proceed to develop the technology required to prove the main results it is useful to consider a simple example which highlights the obstruction we must overcome.

There is in general no natural extension of the Itô map to all continuous paths x_t. The following very simple example shows that the Itô map is rarely a continuous function in the uniform topology.

Example 1.1.1. *Some of the simplest differential equations are those whose solutions can be expressed as exact integrals of the driving term x_t. The simplest nontrivial example is the second iterated integral*

$$
\begin{aligned}
X^2(0,t) &= \int_{t>u_2>0} \Big(\int_{u_2>u_1>0} dx_{u_1} \Big)\, dx_{u_2} \\
&= \iint_{t>u_2>u_1>0} dx_{u_1}\, dx_{u_2}\ .
\end{aligned}
\tag{1.2}
$$

In the one dimensional case, where x_t is real valued, we see that $X^2(0,t) = (x_t - x_0)^2/2$ and so the functional $x. \longrightarrow X^2(0,\cdot)$ clearly is continuous in the uniform topology.

The multi-dimensional case is quite different. Let $\boldsymbol{x}_t = (x_t^1, \dots, x_t^d)$ be vector valued and interpret the second iterated integral as the $d \times d$-matrix defined by

$$
(\boldsymbol{X}^2(0,t))^{ij} = \iint_{t>u_2>u_1>0} dx_{u_1}^i\, dx_{u_2}^j\ ,
$$

or better, as a 2-tensor

$$
\iint_{t>u_2>u_1>0} d\boldsymbol{x}_{u_1} \otimes d\boldsymbol{x}_{u_2}\ . \tag{1.3}
$$

Now decompose this integral into it's symmetric and anti-symmetric components $\boldsymbol{S}^{ij}$, $\boldsymbol{A}^{ij}$. We see that the symmetric part has a form differing little from the one dimensional situation

$$\boldsymbol{S}^{ij} = \frac{1}{2}\,(x_t^i - x_0^i)\,(x_t^j - x_0^j)\,, \tag{1.4}$$

in particular, it is continuous in the uniform topology. The anti-symmetric part, which only arises in dimension two and higher, has the form

$$\boldsymbol{A}^{ij} = \frac{1}{2}\iint_{t>u_2>u_1>0} dx_{u_1}^i\,dx_{u_2}^j - dx_{u_1}^j\,dx_{u_2}^i \tag{1.5}$$

and has a well known geometric interpretation. For any two distinct coordinates i, j, the projection (x_t^i, x_t^j) of the path into $\mathbb{R}^2$ is a directed planar curve. The integral $\boldsymbol{A}^{ij}$ is the area between that curve $(x_\cdot^i, x_\cdot^j)$ and the chord from (x_t^i, x_t^j) to (x_0^i, x_0^j) where multiplicity and orientation are taken into account in the calculation.

Using this obvious geometric remark, it is trivial to see that $\boldsymbol{A}(0,t)$ is not a continuous function of $\boldsymbol{x}_\bullet$ in the uniform topology. Take

$$\boldsymbol{x}_t^n = \Big(\frac{\cos{(n^2\,t)}}{n}, \frac{\sin{(n^2\,t)}}{n}\Big)\,,$$

then as n converges to infinity, the area integral converges locally uniformly to $\pi\,t$ whereas the paths $\boldsymbol{x}_t^n$ converge uniformly to the zero path.

However, closer examination of the example shows that $\boldsymbol{x}_t^n$ is converging to zero in p-variation norm for $p > 2$, and a more complicated example could be given showing that $\boldsymbol{A}$ is discontinuous even for the 2-variation norm. This and other considerations suggest that we should restrict attention to the case where $p < 2$. It is shown in [14], [15] that the Itô map extends uniquely as a continuous function to all paths of finite p-variation norm with $p < 2$ providing the vector fields f^i are smooth enough. In this case one can indeed develop a theory very similar to the classical one.

Nevertheless, there are important formal examples of equations of our basic type (1.1) in which the driving signal fails to have finite 2-variation and these have motivated several attempts to treat equations driven by rougher signals. Easily the most important and successful up to now has been the approach originating with Itô; which treats equations driven by Brownian motion or more generally by semi-martingales

(Brownian paths have finite p-variation norm for every $p > 2$ but do not have finite 2-variation norm) [12]. Although Itô's approach only constructs solutions as random variables it has lead to an enormous range of applications and must be regarded as a major achievement of 20th century mathematics.

Although Itô's approach is not path-wise, it makes it clear that any deterministic approach to interpreting (1.1) that only treats paths of finite p-variation norm with $p < 2$ is missing its target and failing to explain the richest class of examples we have.

We have just seen that iterated integrals provide the obstruction to the continuous extension of the Itô map. The remainder of the paper is dedicated to showing that they are also lead to the solution of the problem. *We will show that the solution is a continuous function of the path and its low order iterated integrals in an appropriate variation norm.* The rougher the path the more iterated integrals required and the more smoothness required of the vector fields.

1.1.5. Summary of existing approaches.

The main approaches to the solution of differential equations seem to have two key features:

- A notion of integral (Riemann, Itô, Stratonovich or Skorohod)

- An understanding of change of variable (Fundamental Theorem of Calculus, Itô's formulae, etc.)

These together allow one to use integral equations to define what one means by a solution. At this point existence can sometimes be shown via fixed point arguments, but in any case one usually wishes to add a method for *constructing* solutions (power series, Picard iteration) which will work under slightly stronger regularity conditions on the vector fields f^i and which usually gives the bonus of uniqueness of solution under these improved regularity assumptions.

Finally one needs to complete the discussion with the observation that characterisations of differential equations via integrals depend on a choice of coordinates for the underlying space where y_t takes its values. So although the equation (1.1) gives the impression of being coordinate independent, the definition of a solution may not be. The issue is a real one; in probability theory the Stratonovich equation has co-ordinate invariant solutions, while Itô equations do not.

In this paper we will concentrate on developing the co-ordinate invariant theory, the full theory is more mathematically challenging, and although we hope to return to it later we do not have a complete description at the current time.

1.1.6. History.

A number of authors have tried to develop deterministic theories of *integration* appropriate for rough paths, attempting to make sense of $\int Y\,dX$. L. C. Young [32] showed that such integrals make sense providing both paths are continuous, X has finite p-variation and Y has finite q-variation and $1/p + 1/q > 1$. For some reason he did not clinch the nonlinear question and show the existence of solutions of differential equations driven by paths with p-variation less than 2 and this was closed off in [14], [15]. Föllmer [6], [7] has written a number of interesting papers giving deterministic meaning to Itô's change of variable formula. Föllmer also made a verbal conjecture at an Oberwolfach meeting several years ago that knowing the Lévy area would be sufficient to construct solutions to SDE's. In some sense we prove his conjecture below.

The case where x_t is one dimensional or 2-dimensional and of the form $(\tilde{x}_t, t)$ is special. In this case the stochastic functional is continuous in the uniform topology - this was established by [26], [27], [28], [29].

1.1.7. Advantages to a probabilist.

A probabilist, interested in stochastic differential equations, might be tempted to believe that this article has little interest for him (except as a theoretical curiosity) because he can do everything that he wanted to do using Itô calculus. So we briefly mention a few situations where we believe that the results we develop here have consequences.

The first is conceptual, until now the probabilist's notion of a solution to an SDE has been as a function defined on path space and lying in some measure class or infinite dimensional Sobolev space. As such, the solution is only defined off an unspecified set of paths of capacity or measure zero. It is never defined at a given path. Given the results below, the solutions to all differential equations can be computed simultaneously for a path with an area satisfying certain Hölder conditions.

The set of Brownian paths with their Lévy area satisfying this condition has full measure. Therefore and with probability one, one may simultaneously solve *all* differential equations[1] over a given driving noise (the content of this remark is in the fact that there are uncountably many different differential equations).

Related consequences include:

1) Stochastic flows can be constructed simply. Changing the starting point in the differential equation is a special case of changing the differential equation. With a little more work one gets continuity, and with increasing smoothness of the vector fields, increasing smoothness of the flow.

2) It can be interesting to solve differential equations subject to boundary conditions other than initial conditions and the construction of a flow often allows one to find an initial value so that the resulting solution satisfies the boundary condition. However, in the classical framework, it is tricky to be precise about the sense in which this "solution" really is a solution. It does not satisfy the predictability condition necessary for the definition of an Itô integral to make sense; the standard approach involving changing the measure is quite deep. We have no such problems of interpretation because we use no probability, (although there will always be a problem of existence of solutions to nonlinear boundary problems - and this can be easy or difficult depending on the precise problem).

3) Stroock and Varadhan established a support theorem for solutions to stochastic differential equations. In one strong and non-trivial form it says that if we fix a smooth path in V and look at the solution to the SDE (1.1) when the driving noise is Brownian motion conditioned to be uniformly close to the smooth path; then the random solution converges in distribution to the deterministic one obtained by driving the equation with the fixed path. It is clear that all such theorems will follow if one establishes the continuity of (1.1) and that Brownian motion conditioned to be uniformly close to the smooth path converges in probability in the metric topology involving the area. Therefore our results below reduce the problem to one about Brownian motion alone.

4) Not all interesting stochastic dynamical systems are semi-martingales. It seems completely natural that there are nonlinear systems forced by random processes that may be Markov or Gaussian but are

[1] The vector fields should be Lipschitz of order greater than two.

certainly not inside the normal framework. One thinks immediately of diffusion processes associated with elliptic operators in divergence form where the coefficients are not differentiable, or of diffusions on fractals. Both of these frequently have area processes and satisfy our hypotheses although they are not semi-martingales. Since this article was written work of Bass, Hambly and Lyons has established that the class of reversible processes to which this theory applies is really much wider than the class for which semi-martingale methods can be used. The iterated Brownian motion (IBM) studied by Burdzy and Adler is another example [1].

5) Numerical algorithms for solving differential equations which adapt their step sizes can be vastly more efficient than fixed step algorithms in certain settings. However, the decisions about step size are most efficiently made on the basis of previous rough approximations to the solution, and identification of the sensitive areas where accurate solution is required (*e.g.* before the trajectory approaches a critical point to ensure it passes on the correct side). The choice of step is typically based on knowledge of the *future evolution* of the solution and is therefore not predictable and constitute illicit information. For example if non-predictable infomation is used to determine the step size in classical approaches to solving SDEs numerically then in general these schemes will converge nicely to the wrong answer. Using the ideas set out below, and ensuring approximations to the path and area of the driving noise are correct over every interval it is possible to have a genuine variable step algorithm that converges to the correct answer for any choice of the intervals of approximation as the mesh size of the dissection goes to zero [8].

6) Stochastic filtering is concerned with the estimation of the conditional law of a Markov process, given observations of some function of it. The normal formulation (due to Zakai) looks at the case where the process is of diffusion type and splits into a first part (known as the signal) and a second part, known as the observation process with values in a vector space, and whose martingale part has stationary increments independent of the signal. In this case, Zakai showed that it was possible to completely describe the conditional density of the signal given knowledge of the observation process. In fact, the density evolves according to an infinite dimensional SDE of parabolic type. It is a commutative equation, and so the relationship between the observation process and the conditional density is a relatively stable one. On the other hand, it is really rather rare that real filtering problems present

themselves with the noise in the observation process being independent of the signal. And the transformation involved in making it so involves the solution of a generic SDE which will not commute. It follows that to do robust and stable filtering it is important to measure the "area" process as well as the values of the observation process.

7) Finally we hope that by solving the one dimensional differential equation without using predictability, our ideas might produce a few pointers to the correct way to treat PDE's driven by spatial noise. Of course in that situation predictability assumptions are quite inappropriate - at least in the initial assumptions and final conclusions. But at the moment this remains pure speculation.

1.2. Background.

1.2.1. Preliminaries: Groups and differential equations.

We set out some basic material and notation.

The logarithm of a flow. Throughout this paper we will make implicit use of the standard identification of autonomous differential equations, flows, and vector fields. If f is a Lipschitz vector field for some choice of complete Riemannian metric on a manifold then the autonomous differential equation

$$dy_t = f(y_t)\,dt\,, \qquad y_0 = a\,, \tag{1.6}$$

has a unique solution defined for all time. By varying the initial condition, one may associate with it a flow π_t defined by $\pi_t(y_0) = y_t$. The assumptions ensure the flow is defined for all positive and negative times and is a homeomorphism. We may use the notation $\pi_t = \exp(tf)$ to emphasise that vector fields should be regarded, at least formally, as elements of the Lie algebra of the group of homeo(diffeo)morphisms of the underlying manifold.

If a homeomorphism π can be realized by flowing along a fixed vector field ϕ so that $\pi = \exp\phi$, we say $\phi = \log\pi$. In general, it is not possible for one to construct a logarithmic vector field even for the smoothest diffeomorphisms homotopic to the identity; equally the logarithm need not be unique when it exists. If one has a time varying differential equation such as (1.6), and one looks at the flow obtained by solving it over a short time then it is useful to be able to determine if the resulting flow has a logarithm and express that logarithmic vector field directly in terms of x_s and the f^i, [3]. (*e.g.* in numerical analysis, to solve the time varying and rough equation over a short interval it would be sufficient to solve the smooth and time independent differential equation determined by the logarithmic vector field).

Determining this logarithm as a vector is in fact an analytic extension of the Dynhin-Campbell-Baker-Hausdorff formula (which in its algebraic form considers the effect of flowing for unit time along one left invariant vector field on a group and then a second, and tries to find an expression for the logarithm of the result). In this paper, we will be able to construct the logarithm of a flow driven by a rough signal for a short period under the hypotheses that the vector fields are invariant vector fields on a finite dimensional group.

Matrix groups. Recall some very basic facts about Lie groups. Suppose that a topological group G has a connected *finite dimensional* manifold structure, then it is very well known that it is a Lie group and can always be represented as a real analytic group of matrices, or a quotient thereof by a discrete group. In this representation, the tangent space to a point in the group is a linear space of matrices.

The tangent space g at the identity can be made into a Lie algebra in two equivalent ways.

If a is an element of the tangent space at the identity of a matrix group, then $t \longrightarrow \exp t\,a$ (where exp is the power series in the the matrix) defines a smooth path in the group (and hence a direction in the tangent space over the identity) starting at the identity element. Consider any other element ρ of the group. The map $t \longrightarrow \exp t\,a\,\rho$ defines a path and hence a direction in the tangent space over ρ; clearly the induced vector field a^* on the group is right invariant, depends linearly on a, and defines an isomorphism between right invariant fields and the tangent space over the identity. We may take the Lie bracket of these fields in the sense of vector fields and as this yields another right invariant vector field we define a Lie algebra structure on the tangent space. Alternatively, we can use the matrix representation and simply define $[A, B] = AB - BA$. They give the same results. The Lie algebra of a Lie group is important in many ways and we cannot recall them all here. However, we mention a couple of basic facts that will be essential. A group is abelian if $[a, b] \equiv 0$, and has nilpotency rank at most n if $[a_1, [a_2, [a_3, \ldots [a_{n-1}, a_n]] \ldots]] \equiv 0$ for all elements in the Lie algebra.

A homomorphism ν of one Lie group to another induces (by differentiation) a Lie map $d\nu$ from the Lie algebra of the first group to the Lie algebra of the second. These two maps intertwine with the exponential map (applied to the vector field or the matrix) and so

$$\nu(\exp t\, a) = \exp t\, d\nu(a)\,. \tag{1.7}$$

Conversely, to every finite dimensional Lie algebra we may associate a unique simply connected Lie group, and to every Lie algebra map from such a finite dimensional Lie algebra to the Lie algebra of a Lie group is associated a unique homomorphism whose derivative is the Lie algebra map.

Differential equations on matrix groups. Suppose we have a smooth path X_t in the Lie algebra of our matrix group, we may develop it onto the group. That is we solve the differential equation for the path ρ_t in the group which at time t is always tangential to $(dX_t/dt)^*$. The differential equation has the form

$$d\rho_t = (dX_t)^*(\rho_t) \tag{1.8}$$

and since * is a linear map from the vector space carrying X_t to vector fields on a manifold (the group) it falls into the general category of time inhomogeneous differential equations we introduced in (1.6).

Any time inhomogeneous differential equation can be regarded, at least formally, in the same way if one is prepared to consider the group of homeomorphisms (or diffeomorphisms) of the manifold. Any vector field defines a parameterised flow on the manifold (1.6) and hence a tangent vector to the identity map on the group of homeomorphisms. Consider the flow π_t on that same group defined by the inhomogeneous equation

$$dy_t = \sum_i f^i(y_t)\, dx_t^i\,, \qquad x_t \in V\,, \qquad f(y) : V \longrightarrow TM_y\,. \tag{1.9}$$

Now $f(\,)\, x_t$ is a path in the space of vector fields, and π_t its development onto the group of homeomorphisms.

Although there are very big differences between this formal infinite dimensional setting and the finite dimensional one (the vector fields will not in general be smooth enough to form a Lie algebra, etc.) the abstract picture is very helpful in the following two ways. It suggests that there might be a universal object, and also that we could learn something about the general problem by studying the simpler case of development of a rough path on a finite dimensional Lie group.

Definition 1.2.1. *A Lie algebra* $\mathfrak{g}$ *containing* V *is said to be free over* V *if it has the universal property that any linear map* f *of* V *into a Lie algebra* $\mathfrak{h}$ *extends in a unique way to a Lie algebra map* $\tilde{f}$ *of* $\mathfrak{g}$ *to* $\mathfrak{h}$*. Such a Lie algebra exists but is infinite dimensional.*

Now suppose we consider again our basic differential equation. That is, we have a path x_t in a vector space V and a linear map f of V into the Lie algebra $\mathfrak{h}$ of a Lie group H and we would like to develop a path y_t in H tangential to $(f(y_t)\,dx_t)^*$.

Pretend for a minute that we could associate a simply connected group G with the free Lie algebra $\mathfrak{g}$, and that there was a group homomorphism from it to H induced by the Lie algebra map. *It would be sufficient to develop* x_t *in the simply connected group* G *with Lie algebra* $\mathfrak{g}$ *and use the homomorphism*

$$G \xrightarrow{\tilde{f}} H$$

to produce a path in H. It follows that it would be both necessary and sufficient to solve our problem in general if we could develop rough paths from V to this Lie group G alone. However, there is a problem with this picture - there is no simple analytic object we can call the free group - but still the picture definitely points one in the correct direction.

Linear differential equations. Suppose that Y_t takes its values in a vector space W and that for each x the vector fields $y \longrightarrow f(y)\,x : W \longrightarrow W$ is linear in y, then we say the standard equation (1.1) is linear, and observe that the sum of two solutions is a solution; the flow is therefore a linear map (which by solving the equation backwards in time is invertible), and the solution flow takes its values in a matrix group.

Thus we see that to solve a time inhomogeneous *linear* equation (which are certainly not linear in the relationship between x and y) is essentially the same problem as to develop a path in a *finite dimensional* Lie algebra onto the associated *finite dimensional* Lie group using the right invariant extensions of the vector fields.

More generally, we can re-parameterise our problem and reduce it to a finite dimensional linear problem whenever the vector fields in the range of f are smooth enough that one can take Lie brackets and the resulting Lie algebra is finite dimensional. Although this is not the generic case, the equation

$$dy_t = f(y_t)\,dx_t\ , \qquad x_t \in V\ , \qquad f(y) : V \longrightarrow TM_y\ , \tag{1.10}$$

where the dimension of V is one satisfies the finite dimensionality hypotheses in a rather trivial way. In this case let $d\theta_t = f(\theta_t)\,dt$ be the flow defined by the autonomous equation. One readily sees that for smooth x_t the solution of (1.10) can easily be expressed as $y_t = \theta_{x_t}(y_0)$ and that this is uniformly continuous in the forcing term x_t. It is generally true that (1.10) is uniformly continuous in this way if and only if the Lie algebra is trivial and the vector fields commute. In Section 1.1.4 we showed that the iterated integral for the area produced a discontinuous Itô map. The associated differential equation has a Lie algebra of the simplest non-commutative type - nilpotent of rank 2.

Einstein expansions. Consider a linear differential equation. Let $x \longrightarrow A(\,)\,x : V \longmapsto \hom(W,W)$ be a bounded linear map (of Banach spaces) and consider the linear equations

$$dy_t = A(y)\,dx_t\ , \tag{1.11}$$

$$d\pi_t = A(\cdot)\,dx_t\,\pi_t\ , \tag{1.12}$$

for the trajectory and flow. If the path x_t is smooth and y_t is the classical solution, then one may construct a Taylor series expansion for it (and the operator π_t) in terms of iterated integrals of x_t.

$$\begin{aligned} y_t &= y_s + \int_s^t dy_u \\ &= y_s + \int_s^t A(y_u)\,dx_u \end{aligned} \tag{1.13}$$

$$\begin{aligned} &= y_s + A(y_s) \int\limits_{s<u<t} dx_u \\ &\quad + \iint\limits_{s<u_1<u_2<t} A(A(y_{u_1}))\,dx_{u_1}\,dx_{u_2} \end{aligned} \tag{1.14}$$

$$= \sum_{i=0}^{n} A(A(\cdots A(y_s))) \int \underset{s<u_1<u_2<\cdots<u_i<t}{\cdots} \int dx_{u_1}\, dx_{u_2} \cdots dx_{u_i}$$

$$(1.15) \qquad + \int \underset{s<u_1<u_2<\cdots<u_{n+1}<t}{\cdots} \int A(A(\cdots A(y_{u_1})))\, dx_{u_1}\, dx_{u_2} \cdots dx_{u_{n+1}}$$

and using the boundedness of y on $[s, t]$, the factorial decay of the iterated integrals, and the geometric growth of the norm of the product of operators, one quickly shows that the remainder goes to zero with n and so we have the convergent series

$$(1.16) \qquad \pi_{s,t} = I + A \int_{s<u<t} dx_u + AA \iint_{s<u_1<u_2<t} dx_{u_1}\, dx_{u_2} + \cdots$$

and observe that the solution can be expressed as a inner product of a sequence of iterated integrals and "powers" of A.

This expansion (which occurs regularly in the literature over the last 50 years or so) underlines the importance of iterated integrals. We will see later that we will be able to associate infinite and rapidly decaying sequences of iterated integrals in settings where the paths are not smooth. In this case the series above can be used as a definition of the solution. However, it does not directly extend from the finite dimensional linear setting (1.11) to the fully nonlinear one (1.1) (for in this case the operators in the range of A are unbounded and do not have a common core). Additional ideas will be required at that point.

1.2.2. Preliminaries: Rough paths and smooth functions.

In this section we remind the reader of some basic analytic concepts. For our purposes a very convenient way of measuring the smoothness of rough paths is via the p-variation norm first introduced by Wiener. If we are to solve differential equations driven by rough paths, then it transpires that we must balance this by taking progressively smoother vector fields. For unique solutions in the classical case it suffices that the fields be Lipschitz. For our uniqueness results we will require that the fields are Lipschitz of order $\gamma > p$. Using the obvious definition, one might conclude that there were no non-constant functions satisfying the hypothesis. The definition we use follows Stein and seems particularly well adapted to the problem in hand. Any bounded function with n bounded derivatives is Lip[γ] for $\gamma \le n$.

Paths of finite p-variation. Suppose X_t is a path taking its values in a metric space. Following Wiener, one says that the p-variation of X_t on J is

$$\|X\|_{p,J}^p = \sup\Big\{\sum_j d(X_{t_j}, X_{t_{j+1}})^p,\ t_{j_1} < \cdots < t_{j_r} \in J\Big\}. \tag{1.17}$$

Definition 1.2.2. *We say that X_t has p-variation controlled by $\omega(s,t)$ if*

$$\|X\|_{p,[s,t]}^p \le \omega(s,t)\,, \qquad \text{for all } s < t\,. \tag{1.18}$$

A path is said to be of regular finite p-variation if ω can be chosen to be continuous near the diagonal, and zero on the diagonal.

Note that

$$\|X\|_{p,[s,t]}^p + \|X\|_{p,[t,u]}^p \le \|X\|_{p,[s,u]}^p \tag{1.19}$$

and so in this paper we only consider controlling ω that satisfy the inequality

$$\omega(s,t) + \omega(t,u) \le \omega(s,u)\,. \tag{1.20}$$

It makes sense to introduce a distance between two paths. Let Y_t denote a second path.

Definition 1.2.3. *We define the distance*[2] *between two paths to be finite if*

$$\|X,Y\|_{p,J} = \max\Big\{\sup_{t_{j_1}<\cdots<t_{j_r}\in J}\Big(\sum_j |d(X_{t_j}, X_{t_{j+1}}) - d(Y_{t_j}, Y_{t_{j+1}})|^p\Big)^{1/p},\ \sup_{t\in J} d(X_t, Y_t)\Big\} < \infty\,.$$

[2] In the more restricted situation where X takes its values in a Banach space there is a smaller norm where $|d(Y,Y) - d(X,X)|$ is replaced by $\|(X_{t_{j+1}} - X_{t_j}) - (Y_{t_{j+1}} - Y_{t_j})\|$. In fact it is this distance that we will use later.

As before we may talk about the distance being controlled by ω.

It is obvious from standard facts about sequence spaces that this distance is indeed a metric (and a norm if the original space were a Banach space) and that the space of paths of finite p-variation is complete in this metric providing the original metric space was complete. The space of regular paths is a closed subspace. The p-variation of a path and distance between two paths are monotone decreasing with increasing p. If X is continuous and of finite p-variation then X is regular for all $p' \geq p$. If X is not continuous the local p-variation never goes to zero and the path is never regular.

Example 1.2.1. *A path of bounded variation on a closed interval has finite* 1*-variation. Almost all Brownian paths* $X_t(\omega)$ *are of regular* p*-variation for all* $p > 2$ *but do not have finite* 2*-variation although the map* $t \longrightarrow X_t(\cdot)$, $\mathbb{R}^+ \longrightarrow L^2(\Omega, \mathbb{P})$ *does have finite* 2*-variation.*

Lipschitz functions. In [24, Chapter VI] Stein looked at the general problem of extending smooth functions from subsets of Euclidean space to the whole space. In particular, he considers the Whitney theorem which extends in a norm bounded way the space $\mathrm{Lip}(\gamma, F)$ of Lipschitz functions on a closed set F to the whole Euclidean space. In doing so he introduces a definition of $\mathrm{Lip}(\gamma, F)$ which is valid for any $\gamma > 0$ and not just for $\gamma \leq 1$. We recall a modification of this definition here; although we modify it slightly to be compatible with our notations; the resulting norms are equivalent.

Definition 1.2.4. *Suppose that* V, W *are normed vector spaces,* k *is a non-negative integer, and that* $k < \gamma \leq k+1$. *A function* $f = f_0$ *defined on a closed subset* $F \subset V$ *and taking values in* W *belongs to* $\mathrm{Lip}(\gamma, F)$ *if there exist symmetric multi-linear functions (formal derivatives)* $f^{(j)}(\boldsymbol{x})$, $0 \leq j \leq k$ *taking* $\overset{j}{\underset{1}{\otimes}} V$ *to* W *and satisfying the natural Taylor expansion type condition*

$$
\begin{aligned}
f^{(j)}(\boldsymbol{x}_t)(v) = \sum_{j+l \leq k} f^{(j+l)}(\boldsymbol{x}_s)\Big(v \otimes \int \underset{s<u_1<\cdots<u_l<t}{\int} d\boldsymbol{x}_{u_1} \cdots d\boldsymbol{x}_{u_l}\Big) \\
+ R_j(\boldsymbol{x}_s, \boldsymbol{x}_t)(v)\,. \qquad (1.21)
\end{aligned}
$$

for $v \in \overset{j}{\underset{1}{\otimes}} V$ *and where, as operators on the tensor product, the deriva-*

tives and remainder satisfy

$$\|f^{(j)}(\boldsymbol{x})\| \leq M\,, \tag{1.22}$$

$$\|R_j(\boldsymbol{x},\boldsymbol{y})\| \leq M\,|\boldsymbol{x}-\boldsymbol{y}|^{\gamma-j}\,, \qquad \boldsymbol{x},\boldsymbol{y}\in F\,. \tag{1.23}$$

We define the smallest M to be the $\operatorname{Lip}(\gamma, F)$ *norm of the sequence* $f^{(j)}(\boldsymbol{x})$, $0 \leq j \leq k$.

Some remarks are in order.

The terms

$$f^{(j+l)}(\boldsymbol{x}_s)\Big(v \otimes \int \underset{s<u_1<\cdots<u_l<t}{\cdots} \int d\boldsymbol{x}_{u_1}\cdots d\boldsymbol{x}_{u_l}\Big)\,, \tag{1.24}$$

are, for smooth paths/conventional integrals, independent of the choice of path and only depend on the values $(\boldsymbol{x}_s, \boldsymbol{x}_t)$. To prove this, observe first, that the dimension of W is irrelevant. Now consider the polynomial $p(\boldsymbol{x})$ of degree k whose Taylor expansion at $\boldsymbol{x}_s$ agrees with $\{f^{(j)}(\boldsymbol{x}_s)\}_{j=0,\dots,k}$. Expanding $p(\boldsymbol{x}_t)$ in terms of iterated integrals, as in the last section, we see that the expansion formulae is exact at level k and

$$p(\boldsymbol{x}_t) = \sum_{0<l\leq k} f^{(l)}(\boldsymbol{x}_s)\Big(\int \underset{s<u_1<\cdots<u_l<t}{\cdots} \int d\boldsymbol{x}_{u_1}\cdots d\boldsymbol{x}_{u_l}\Big)\,. \tag{1.25}$$

and as the left hand expression does not depend on the path nor can the right hand side. Similar arguments can be applied to the derivatives of $p(\boldsymbol{x})$ to obtain the invariance of the other expressions. An alternative, more algebraic proof of the result is to observe that the symmetric nature of the $\{f^{(j)}(\boldsymbol{x}_s)\}_{j=0,\dots,k}$ annihilates the antisymmetric components of a tensor and only these change when one perturbs the path. Either way, the observation is clear, and will be crucial to us.

The functions $\{f^{(j)}\}_{j=1,\dots,k}$ will not in general be unique given $f = f_0$. One only expects this if the set F is thick enough. In other words a function in $\operatorname{Lip}(\gamma, F)$ *is not a function on F but a sequence of functions representing* formal derivatives and satisfying these complex Taylor type bounds relating one term with the next. We will see that an essentially dual idea occurs when one considers paths of finite p-variation where $p > 2$. The definition we give above for p-variation, is in some sense wrong, as it fails to specify enough information.

Definition 1.2.5. *We have defined a* $\mathrm{Lip}(\gamma, F)$ *function; this definition easily extends to i-forms. A sequence* $f^{(j)}(\boldsymbol{x})$, $i \leq j \leq k$ *is a* $\mathrm{Lip}(\gamma, F)$ *i-form if all the higher Taylor expressions* (1.21) *satisfy the estimates set out in the definition above whenever* $i \leq j \leq k$.

Both definitions make sense if the functions or forms are vector valued.

Example 1.2.2. *If* θ *is a* 1*-form on* F *and* $1 < \gamma \leq 2$, *then we say it is in* $\mathrm{Lip}(\gamma, F)$ *if one has defined a* 2*-form* $d\theta$

$$\left\| \theta(X_t) - \theta(X_s) - \frac{1}{2} (d\theta)(X_s)(X_t - X_s) \right\| < M \|X_t - X_s\|^\gamma \tag{1.26}$$

and

$$\|d\theta(X_t) - d\theta(X_s)\| < M \|X_t - X_s\|^{\gamma - 1} . \tag{1.27}$$

However, some caution is now required as the resulting multi-linear maps are only required to have full symmetry in the $x_1, \ldots, x_l$ co-ordinates. One may compare this approach to defining $\mathrm{Lip}(\gamma, F)$ j-forms with the alternative approach which simply says a form valued function is a matrix valued function, and so we have already defined what we mean by $\mathrm{Lip}(\gamma, F)$. The two approaches give the same result.

2. The Finite-Dimensional Case - Linear Differential Equations.

2.1. Multiplicative Functionals - Introduction.

2.1.1. Multiplicative functionals - Introductory material and definitions.

Let V be a vector space, and suppose that X_t is a smooth path in V. The k-th iterated integral $\boldsymbol{X}^k_{s,t}$ of X_t over a fixed time interval $[s, t]$ is an element of the tensor product $V^{\otimes k}$. The sequence of iterated integrals

$$(\boldsymbol{X}^k_{s,t})_{k=0}^{\infty}$$

is far from being a generic collection of tensors; there are complicated algebraic dependencies between the terms in the sequence. To fully understand the collection of iterated integrals one must treat them as a single object.

The tensor algebra. We start by recalling some rather elementary facts about tensor algebras. Consider the space T of sequences $\boldsymbol{a} = (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \dots)$ with $\boldsymbol{a}_k \in V^{\otimes k}$. That is

$$T = \bigoplus_{k=0}^{\infty} V^{\otimes k} . \tag{2.1}$$

(We take the zero order tensor product to be the field of scalars.) Then T is an associative algebra with unit, which we shall refer to as the tensor algebra over V. If $\boldsymbol{a} = (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \dots)$ and $\boldsymbol{b} = (b_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \dots)$ are two elements of T then we may define their sum, (tensor) product, and the action of scalars in the obvious way

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= (a_0 + b_0, \boldsymbol{a}_1 + \boldsymbol{b}_1, \boldsymbol{a}_2 + \boldsymbol{b}_2, \dots) , \\ (\boldsymbol{a} \otimes \boldsymbol{b})_i &= \sum_{0 \le j \le i} \boldsymbol{a}_j \otimes \boldsymbol{b}_{i-j} , \\ \alpha \, \boldsymbol{a} &= (\alpha \, a_0, \alpha \, \boldsymbol{a}_1, \alpha \, \boldsymbol{a}_2, \dots) . \end{aligned} \tag{2.2}$$

The space T with these operations is an associative algebra. Suppose that $\boldsymbol{a} = (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \dots)$ is any element of the algebra with $a_0 > 0$ then $\boldsymbol{a}$ is invertible using the usual geometric power series approach

$$\begin{aligned} \boldsymbol{a} &= a_0 \, (1, \boldsymbol{b}_1, \boldsymbol{b}_2, \dots) = a_0 \, (1 + \boldsymbol{c}) , \\ \boldsymbol{a}^{-1} &= \frac{1 - \boldsymbol{c} + \boldsymbol{c}^2 - \boldsymbol{c}^3 + \cdots}{a_0} , \qquad a_0 \in \mathbb{R} , \end{aligned} \tag{2.3}$$

where

$$\begin{aligned} \boldsymbol{b}_i &= \frac{\boldsymbol{a}_i}{a_0} , \\ \mathbf{1} &= (1, \mathbf{0}, \mathbf{0}, \dots) , \\ \boldsymbol{c} &= (0, \boldsymbol{b}_1, \boldsymbol{b}_2, \dots) . \end{aligned} \tag{2.4}$$

Now $c_0 = 0$, hence $\{\boldsymbol{c}^j\}_k = 0$ providing $k < j$; therefore the k-tensor component of any power series in $\boldsymbol{c}$ and in particular $1 + \boldsymbol{c} + \boldsymbol{c}^2 + \boldsymbol{c}^3 + \cdots$ is a sum including only finitely many non zero terms and so has meaning. Similarly, providing $a_0 > 0$, we may define the logarithm of $\boldsymbol{a}$ by

$$\log \boldsymbol{a} = \log a_0 + \mathbf{1} + \boldsymbol{c} - \frac{\boldsymbol{c}^2}{2} + \frac{\boldsymbol{c}^3}{3} + \cdots \tag{2.5}$$

Both of these definitions are pure algebra, and no analysis is required. The exponential function is defined for all elements of T, but the series defining the k-tensor component involves a genuinely infinite sum (which always converges).

$$\exp \boldsymbol{a} = 1 + \boldsymbol{a} + \frac{\boldsymbol{a}^2}{2!} + \frac{\boldsymbol{a}^3}{3!} + \cdots \tag{2.6}$$

One can check that $\exp(-\boldsymbol{a}) = (\exp \boldsymbol{a})^{-1}$ and that $\exp \log \boldsymbol{a} = \boldsymbol{a}$, $\log \exp \boldsymbol{a} = \boldsymbol{a}$, etc. Because the space

$$D_n = \bigotimes_{k=n+1}^{\infty} V^{\otimes k} \tag{2.7}$$

of tensors of degree greater than n form an ideal we may also study the truncated tensor algebra $T^{(n)}$ obtained by quotienting out by D_n. We make the identification

$$T^{(n)} = \bigoplus_{k=0}^{n} V^{\otimes k}. \tag{2.8}$$

The full tensor algebra is an adequate algebraic object, but because it ignores any notion of convergence of the infinite sequences it is a rather poor analytic object. We will mainly work with the truncated tensor algebras $T^{(n)}$ where the analytic and algebraic structures are completely compatible, the fine analytic information will come from understanding the way objects in these finite dimensional quotients piece together.

At this point we record only the basic facts. If $m > n$, then there is a natural projection π of $T^{(m)}$ onto $T^{(n)}$ given by

$$\pi : (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m) \longmapsto (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n). \tag{2.9}$$

The map π *is an algebra homomorphism*. Moreover, the definitions of log, exp, $\boldsymbol{a}^{-1}$ extend to $T^{(n)}$ and their actions commute with that of π

so that for example $\pi(\exp \boldsymbol{a}) = \exp(\pi(\boldsymbol{a}))$. The inclusion ι of $T^{(n)}$ into $T^{(m)}$ given by

$$\iota : (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n) \longmapsto (a_0, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n, \mathbf{0}, \mathbf{0}, \ldots) . \tag{2.10}$$

is linear *but is not an algebra homomorphism.*

The free Lie algebra and free nilpotent groups. One can build certain Lie algebras inside $T^{(n)}$ and T. The product

$$[\boldsymbol{a}, \boldsymbol{b}] = \boldsymbol{a} \otimes \boldsymbol{b} - \boldsymbol{b} \otimes \boldsymbol{a} . \tag{2.11}$$

defines a Lie Bracket on T and $T^{(n)}$. Of particular interest is the Lie algebra generated by V. This is comprised of linear combinations of finite sequences of Lie brackets of elements of V

$$\mathfrak{A} = 0 \oplus V \oplus [V, V] \oplus [V, [V, V]] \oplus \cdots$$

where for example $[V, [V, V]]$ is the linear subspace of $V^{\otimes 3}$ spanned by

$$[\boldsymbol{v}_1, [\boldsymbol{v}_2, \boldsymbol{v}_3]] , \qquad \boldsymbol{v}_i \in V .$$

One may trivially prove that it has the special property that if S is a *linear* map from V into a Lie algebra $\mathfrak{B}$ then there is a unique extension of the map to a Lie algebra map from $\mathfrak{A}$ to $\mathfrak{B}$. In other words it is the free Lie algebra we identified earlier. The corresponding Lie algebra $\mathfrak{A}^{(n)} \subset T^{(n)}$ has the same extension property providing one restricts attention to maps into Lie algebras of nilpotency rank at most n (*i.e.* all Lie products involving n or more elements of the algebra are identically zero).

Theorem 2.1.1. *Let $G^{(n)} = \exp \mathfrak{A}^{(n)} \subset T^{(n)}$ then $G^{(n)}$ is a group called the free nilpotent group of step n. $G^{(0)} = \mathbb{R}$ and $G^{(1)} = V$. The exponential map from the Lie algebra $\mathfrak{A}^{(n)}$ to the Lie group $G^{(n)}$ is one to one and onto. The restriction of the map π to a map from $G^{(m)} \longrightarrow T^{(n)}$, $m > n$ defines a group homomorphism from $G^{(m)} \longrightarrow G^{(n)}$, $m > n$. On the other hand the map ι takes $G^{(n)} \longrightarrow T^{(m)}$, $m > n$ but intersects $G^{(m)}$ only at the identity.*

Remark 2.1.1. The above theorem and indeed everything in 2.1.1 is standard, proofs can be found in, for example, [22]. The properties of $\mathfrak{A}^{(n)}$ are by no means all easy to derive, and for example it is an interesting, nontrivial, and a numerically worthwhile exercise to compute the dimension of $\mathfrak{A}^{(n)}$, to find explicit bases for the space (such calculations go back to Hall and Linden), and even to decompose the space into GL(V)-invariant subspaces [22] according to the different irreducible representations.

Every element of the free Lie algebra is of finite degree and an element of $T^{(n)}$ for some n. We may exponentiate the free Lie algebra into the full tensor algebra and the map is injective, but the range is not a group or even multiplicatively closed. On the other hand, we can introduce the (highly non-separable) Lie algebra of infinite sequences of Lie elements. In this case, we see that the exponential map has a range comprising solely of elements of the tensor algebra which are carried by each of the projections $\pi : T \longrightarrow T^{(n)}$ into the corresponding group $G^{(n)}$. This subset of the full tensor algebra is the inverse limit of our nilpotent groups and is clearly itself a group which we denote $G^{(*)}$.

Definition 2.1.1. *We say an element of the full tensor algebra is group like if it is an element of* $G^{(*)}$.

Unfortunately this group is very big and its Lie algebra is no longer the free algebra.

Any attempt to use a linear map from V into the Lie algebra of a Lie group H to define a homomorphism of this enormous Lie group $G^{(*)}$ (or some part thereof) into the group H in a unique way must involve analysis. This paper can be viewed as an attempt to provide this analytic content.

Paths and multiplicative functionals - the definition. Let $\boldsymbol{X}_t$ be a fixed smooth path in V, and consider the sequence of iterated integrals

$$\boldsymbol{X}_{s,t}^{(n)} = 1 + \int\limits_{s<u<t} dx_u + \int \int\limits_{s<u_1<u_2<t} dx_{u_1} \otimes dx_{u_2} + \cdots$$

$$+ \int \int\limits_{s<u_1<u_2<\cdots<u_n<t} dx_{u_1} \otimes dx_{u_2} \otimes \cdots \otimes dx_{u_n} \in T^{(n)} . \tag{2.13}$$

Let $\boldsymbol{X}_{s,t}$ denote the infinite sequence. Suppose now that one wants to describe in detail the relationship between the iterated integrals over

$[r,t]$ and those over $[r,s]$ and $[s,t]$ where $s \in [r,t]$. If one starts to calculate in coordinates one will quickly become engulfed in terms and conclude that this is a horribly complicated thing to do, however this is really because the main features are best derived without taking coordinates. Now K. T. Chen [5] observed two essential features of the process $\boldsymbol{X}_{s,t}$ which we now state as a theorem.

Theorem 2.1.2. *For smooth paths and conventional integrals, the process $\boldsymbol{X}_{s,t}$ is multiplicative. That is to say*

$$\boldsymbol{X}_{r,s} \otimes \boldsymbol{X}_{s,t} = \boldsymbol{X}_{r,t} \,. \tag{2.14}$$

Moreover, it is group like, so that for each n,

$$\boldsymbol{X}_{s,t}^{(n)} \in G^{(n)}\,, \qquad \log\left(\boldsymbol{X}_{s,t}^{(n)}\right) \in A^{(n)}\,. \tag{2.15}$$

PROOF. The proof that $\boldsymbol{X}_{s,t}$ is multiplicative is instructive. Let the i-th component of $\boldsymbol{X}_{s,t}$ be denoted by $\boldsymbol{X}_{s,t}^i$, etc. Then

$$\begin{aligned}
\boldsymbol{X}_{r,t}^i &= \int\limits_{r<u_1<u_2<\cdots<u_i<t}\!\!\int dx_{u_1}\,dx_{u_2}\cdots dx_{u_i} \\
&= \sum_{0\le j\le i} \int\limits_{s<u_{j+1}<\cdots<u_i<t}\!\!\int \Big(\int\limits_{r<u_1<\cdots<u_j<s}\!\!\int dx_{u_1}\cdots dx_{u_j}\Big) \\
&\qquad\qquad \cdot dx_{u_{j+1}}\cdots dx_{u_i} \\
&= \sum_{0\le j\le i} \Big(\int\limits_{r<u_1<\cdots<u_j<s}\!\!\int dx_{u_1}\cdots dx_{u_j}\Big) \\
&\qquad \cdot \int\limits_{s<u_{j+1}<\cdots<u_i<t}\!\!\int dx_{u_{j+1}}\cdots dx_{u_i} \\
&= \sum_{0\le j\le i} \boldsymbol{X}_{rs}^j \otimes \boldsymbol{X}_{st}^{i-j}\,,
\end{aligned} \tag{2.16}$$

which establishes the multiplicative identity.

To prove that the iterated integral sequence is group like one needs a different approach. Because iterated integrals are integrals and our paths are smooth, it is an easy consequence of the fundamental theorem of calculus that they satisfy the system of differential equations

$$(2.17)\qquad \begin{cases} d\boldsymbol{X}_{0,t}^{(n)} = \boldsymbol{X}_{0,t}^{(n)} \otimes dX_t \in T^{(n)}\,, \\ \boldsymbol{X}_{0,0}^{(n)} = \mathbf{1} = (1, \mathbf{0}, \mathbf{0}, \dots)\,. \end{cases}$$

Suppose g is an element of the group $G^{(n)}$ thought of as a sub-manifold of $T^{(n)}$. Then left tensor multiplication by g is a linear map of $T^{(n)}$ which takes the group $G^{(n)}$ to itself, and 1 to g. It follows that the derivative of this map takes the tangent space to the group $G^{(n)}$ at 1 to the tangent space to g. However the derivative of a linear map is the map itself, and V is in the tangent space to $G^{(n)}$ at $\mathbf{1}$. Hence any solution to the differential equation $d\boldsymbol{g}_t = \boldsymbol{g}_t \otimes dX_t \in T^{(n)}$ will remain in the group $G^{(n)}$ if it starts there. It follows that $\boldsymbol{X}_{s,t}$ is a group like element.

REMARKS 2.1.1. The proof of the above result yields a certain amount of extra information.

1) From the differential equation (2.17) (which of course is of a very fundamental kind) we observe that the iterated integrals over a fixed time interval are insensitive to re-parameterisation of the underlying path, and by solving the differential equation backwards in time we see that the inverse group element is produced. The map from piecewise smooth path segment to iterated integral sequence is a homomorphism of the semi-group of path segments (multiplication is concatenation) to the group like elements. Identify re-parameterisations of paths, and the inverse of path segment with the path run in the reverse direction and one makes the path segments into a group. Chen proved that in this case the map into the group like elements is injective. Therefore, the infinite algebraic sequence $\boldsymbol{X}_{s,t}$ contains (in code!) all the information from x_u, $u \in [s,t]$ required to determine the solution y_t from y_s.

2) The proof of the first part of the theorem holds in wide generality. The first integral identity relies only on additivity of the integral over disjoint domains of a nice kind. The second term depends on a multiplicative linearity of the integral. In fact these properties (of linearity and additivity) are so basic that (2.14) is true for any sensible choice of integral (Itô, etc.) and in some sense captures what one means when one talks about an integral. Because the multiplicative property

seems so widely characteristic of integrals we make it our basic object of study.

Definition 2.1.2. *A multiplicative functional is a map from pairs* (s,t) *of real numbers to* $\boldsymbol{X}_{st} = (X^0_{st}, \boldsymbol{X}^1_{st}, \boldsymbol{X}^2_{st}, \dots)$ *in* $T^{(n)}$ *satisfying* $\boldsymbol{X}_{rs} \otimes \boldsymbol{X}_{st} = \boldsymbol{X}_{rt}$ *and* $X^0_{st} \equiv 1$. *We say a multiplicative functional is geometric if it takes its values in the group like elements.*

Suppose that $\boldsymbol{X}_t = (1, \boldsymbol{X}^1_t, \boldsymbol{X}^2_t, \dots, \boldsymbol{X}^n_t) \in T^{(n)}$ is a path in the space of n-tensors with unit scalar component. Then we say that $\boldsymbol{X}_{s,t} = (\boldsymbol{X}_s)^{-1} \otimes \boldsymbol{X}_t$ is the multiplicative functional determined by $\boldsymbol{X}_t$. Conversely, given a multiplicative functional $\boldsymbol{X}_{s,t}$ and a point $\boldsymbol{x}$ in $T^{(n)}$, we say that $\boldsymbol{X}_t = \boldsymbol{x} \otimes \boldsymbol{X}_{0,t}$ is the path in $T^{(n)}$ starting at $\boldsymbol{x}$ determined by $\boldsymbol{X}_{s,t}$. Given this almost one to one correspondence between paths and multiplicative functionals in $T^{(n)}$ it is reasonable to question the sense of introducing the concept of multiplicative functional at all. However, we will see later that it will be fundamental to the process of constructing an integral or of solving a differential equation that one can go from an almost multiplicative functional to a multiplicative functional and hence to a path. Almost multiplicative functionals will have no direct path-wise interpretation.

The logarithmic flow. As a simple application of the algebraic ideas set out so far, we go back to a question we raised earlier, suppose that one would like to know how to construct the logarithm of a flow. We can easily derive an asymptotic formulae for the logarithm of the flow (proving that it converges to a Lie element is of course a different question). Recall our basic equation

$$dy_t = f(y_t)\, dx_t \ , \tag{2.18}$$

where f is the linear map from V to a space of vector fields and suppose the fields form a Lie algebra (*e.g.* they are smooth). Can we construct a fixed vector field which, if we flow along it for unit time, gives the same homeomorphism as solving the inhomogeneneous differential equation over the interval $[s,t]$? Now f is a linear map from V into the smooth vector fields on some general target space. Because of the universal property of $\mathfrak{A}$ the map f extends to a unique Lie map f_* from $\mathfrak{A}$ into the vector fields with $f_*([v_1,[v_2,v_3]]) = [f(v_1),[f(v_2),f(v_3)]]$. The logarithm of the flow should be given by $f_*(\log(\boldsymbol{X}_{0,t}))$. However, this calculation is formal because one quietly slips from finite to infinite sequences. On the other hand one can always compute

$$f_*(\log(\boldsymbol{X}_{0,t}^{(n)}))\,, \qquad \text{where } \log(\boldsymbol{X}_{0,t}^{(n)})$$

is regarded as an element of $\mathfrak{A}^n \subset T^{(n)}$. These form a sequence of explicit and readily calculable vector fields providing an asymptotic expansion for the logarithmic vector field. A number of the optimal algorithms for solving sde's numerically are based on this idea [4].

Rough and smooth multiplicative functionals. Although our prime examples were obtained by computing the iterated integrals of a smooth path, the underlying definition of a multiplicative functional is at present a purely algebraic one. We now wish to consider rough and smooth multiplicative functionals. Equivalently we wish to consider rough or smooth monic paths in the truncated tensor algebras. For this we need a notion of distance between tensors in $T^{(n)}$. For all further discussion, suppose that V, and more generally $V^{\otimes n}$ are Banach spaces and that they have compatible norms $\|\cdot\|$ so that $\|\boldsymbol{u}\otimes\boldsymbol{v}\| \le \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$, and that the norms are invariant under permutations of the indices of the tensors. (Given a norm on V there are many norms one could take on the tensor products so that this property holds). Let $\boldsymbol{c} = (0, \boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_n)$ be an element of the radical

$$D_0^{(n)} = \bigoplus_{k=1}^{n} V^{\otimes k}$$

of $T^{(n)}$, then for any sequence $\lambda = (\lambda_1, \lambda_2, \dots)$ of strictly positive weights we may define a homogeneous distance function

$$|||\boldsymbol{c}|||_\lambda = \max\{(\lambda_i\,\|\boldsymbol{c}_i\|)^{1/i} :\ 1 \le i \le n\}\,. \tag{2.19}$$

It is clear that $|||\boldsymbol{c}+\boldsymbol{d}|||_\lambda \le |||\boldsymbol{c}|||_\lambda + |||\boldsymbol{d}|||_\lambda$ and so we define a metric on the radical by $d(\boldsymbol{c},\boldsymbol{d}) = |||\boldsymbol{c}-\boldsymbol{d}|||_\lambda$. The metrics on $D_0^{(n)}$ are uniformly equivalent for alternative choices of the constants λ, however this is only true for fixed and finite n. Although the metric is not a norm if $n > 1$ it has the very important property that it has the same homogeneity properties as our sequence of iterated integrals when we scale the underlying path.

Consider the element of the radical $\boldsymbol{X}_{s,t}^{(n)}(\eta) - \mathbf{1}$ generated by the sequence of iterated integrals of a smooth path η_s. Now scale the path, then the individual iterated integrals transform according to their degree and

$$|||\boldsymbol{X}_{s,t}^{(n)}(\varepsilon\,\eta) - \mathbf{1}||| = \varepsilon\,|||\boldsymbol{X}_{s,t}^{(n)}(\eta) - \mathbf{1}|||\,.$$

If we are only interested in fixed n we will frequently take $\lambda \equiv 1$ to avoid complicated expressions. If we wish to prove that the Einstein expansion for the solution of a linear equation converges one will need to control the behaviour as n goes to infinity. For this one requires a choice of λ very well adapted to the problem. In this more critical work we find $\lambda_i = \beta\,(i/p)!$ to be an excellent choice, where $\beta > 0$, $p > 1$ are to be chosen later. For notational convenience, we will use the notation $|||\cdot|||$ to denote either metric. It will not cause significant confusion.

Suppose that we have a monic path $\boldsymbol{X}_t$ in the truncated tensor algebra and its associated multiplicative functional $\boldsymbol{X}_{s,t}$. Then we could introduce a distance $\rho(\boldsymbol{X}_s, \boldsymbol{X}_t) = |||\boldsymbol{X}_{s,t} - \mathbf{1}|||$. In general this will not be a metric (although it is good enough) because it fails the symmetry condition and the triangle inequality. However, it is clear from the neo-classical inequality (see later) that if $\beta \geq 2^p\,p^2$ then it will satisfy the triangle inequality. For group like elements, it is obvious from Remark 2.1.1., the inverse being obtained from the path run backwards and the invariance of the norm under re-ordering of the tensors, that the inverse of a group like element has the same modulus as the original element. In this case it is clearly a metric.

We ignore the fact that this distance is a metric or not (because it follows that it is always equivalent to one). In any case we may follow section 2.1.1., and use it to define monic paths and multiplicative functionals of finite p-variation (controlled by a regular super additive function $\omega(s,t)$ etc.) and to provide a distance between two paths.

Lemma 2.1.1. *A multiplicative functional $\boldsymbol{X}_{s,t}$ in $T^{(n)}$ is of finite p-variation controlled by ω if and only if it satisfies the inequality*

$$\|\boldsymbol{X}_{s,t}^{i}\| < \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!}\,, \qquad i \leq n\,.$$

The proof is immediate from the definition. We include it as a convenient formulation.

2.2. Multiplicative functionals - the first main theorems.

Overview. We have introduced the idea of a multiplicative functional in $T^{(n)}$ of finite p-variation without making any direct connection between the degree n of the multiplicative functional and the roughness of the path as described by $p \geq 1$.

The theorems in this section, which are fundamental to our approach, demonstrate the central role played by the class of multiplicative functionals for which the degree n is the integer part $[p]$ of the variation p.

We have already observed that if we take a smooth path in a vector space and take its first k iterated integrals then we have constructed a multiplicative functional of degree k; computing the next iterated integral gives a method of extending the multiplicative functional to (a geometric) one of the next degree. This extension map is continuous as a function of the underlying path in p-variation metric if and only if $p < 2$.

By way of an extension of this result, the theorems in this section show by restriction that, for any $p \geq 1$, if we regard as our basic object the smooth path *and its iterated integrals of degree up to* $[p]$ then the higher iterated integrals are uniformly continuous functions in the metric of finite p-variation. The uniform continuity allows one to extend the definition of iterated integral to this class.

These results are the first step towards our main theorem that the Itô map[1] is uniformly continuous as a function of the sequence comprising a smooth path and its iterated integrals of degree up to $[p]$ where one takes the metric of finite p-variation. So providing a natural analytic extension of the Itô map to the class of geometric paths of finite p-variation and degree $[p]$.

The application to stochastic Stratonovich differential equations is realized by taking $3 > p > 2$; where these results reduce to the statement that the Itô map is continuous in the pair comprising the path and its Lévy area.

2.2.1. The First Theorem.

Theorem 2.2.1. *Let* $\boldsymbol{X}_{s,t}^{(n)}$ *be a multiplicative functional in* $T^{(n)}$ *of finite* p*-variation controlled by a regular* $\omega(s,t)$ *on an interval* J *where*

[1] defined by a differential equation with smooth enough coefficients

$n = [p]$. There exists a multiplicative extension $\boldsymbol{X}_{s,t}^{(m)}$ to $T^{(m)}$, $m > n$ which is of finite p-variation, the extension is unique in this class.

Moreover, this unique extension satisfies a rather precise estimate. Suppose that the p-variation norm of $\boldsymbol{X}_{s,t}^{(n)}$ is controlled by $\omega(s,t)$ so that for all pairs of times in an interval we have

$$\|\boldsymbol{X}_{s,t}^{i}\| < \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!}\,, \qquad i \le p\,, \tag{2.20}$$

then, providing β is large enough the same inequality

$$\|\boldsymbol{X}_{s,t}^{i}\| < \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!}\,, \qquad i > p\,, \tag{2.21}$$

holds in all degrees and p-variation norm of $\boldsymbol{X}_{s,t}^{(m)}$ is controlled by $\omega(s,t)$ without any sort of factor for all m.[2]

REMARKS 2.2.1. 1) It suffices for the above theorem that

$$\beta > p^2\Big(1 + 2^{([p]+1)/p}\Big(\zeta\Big(\frac{[p]+1}{p}\Big) - 1\Big)\Big)\,, \tag{2.22}$$

where

$$\zeta(z) = \sum_{1}^{\infty} \frac{1}{n^z} \tag{2.23}$$

is the traditional Riemann zeta function.

2) It is a more or less trivial remark that, in the case where $n < [p]$, if a multiplicative functional of degree n and finite p-variation has an extension to a multiplicative functional of degree m of finite p-variation, then the extension will never be unique. On the other hand in the case where $n > [p]$ the above theorem shows by restriction the existence and uniqueness of an extension of finite p-variation.

In the remainder of this section we outline the proof.

Two key results under-pin our argument. The first is completely elementary.

---

[2] where $x! = \Gamma(x+1)$

Lemma 2.2.1. *Suppose*

$$D = \{s = t_0 < t_1 < \cdots < t_r = t\}$$

is a dissection of $[s,t]$. *Then there is a* j *such that*

$$\omega(t_{j-1}, t_{j+1}) \le \begin{cases} \dfrac{2\,\omega(s,t)}{r-1}\,, & r > 2\,, \\ \omega(s,t)\,, & r = 2\,. \end{cases} \tag{2.24}$$

PROOF. ω is super-additive and so when $r > 2$

$$\sum_{1}^{r-1} \omega(t_{j-1}, t_{j+1}) \le 2\,\omega(s,t) \tag{2.25}$$

and at least one term in a sum is dominated by the mean so the result is clear. On the other hand when $r = 2$

$$\omega(t_{j-1}, t_{j+1}) = \omega(s,t)\,, \qquad \text{if } j = 1\,.$$

A neo-classical inequality.

Lemma 2.2.2. *The following extension of the binomial theorem holds*

$$\left(\frac{1}{p}\right)^2 \sum_{j=0}^{n} \frac{x^{j/p}}{(j/p)!} \, \frac{y^{(n-j)/p}}{((n-j)/p)!} \le \frac{(x+y)^{n/p}}{(n/p)!}\,, \tag{2.27}$$

where $n \in \mathbb{N}$, $x, y > 0$, $p \ge 1$.

We postpone the proof of this inequality which is quite non-trivial. Notice that since $(x/p)!$ is roughly $(x!)^{1/p}$, the lemma loosely asserts that we have a sequence of numbers satisfying $\sum a_j = b$ from the binomial theorem and $\sum a_j^{1/p} \le b^{1/p}$. In general the inequality would be reversed.

PROOF. **Existence.** Our intention is to proceed by induction. Fix $m \ge [p]$. As initial data consider a multiplicative functional

$$\boldsymbol{X}^{(m)}_{s,t} = (1, \boldsymbol{X}^1_{st}, \ldots, \boldsymbol{X}^{[p]}_{s,t}, \boldsymbol{X}^{[p]+1}_{s,t}, \ldots, \boldsymbol{X}^m_{s,t})$$

satisfying (2.32), we wish to construct a multiplicative functional $\boldsymbol{X}_{s,t}^{(m+1)}$ satisfying the same constraints.

Consider

$$\widehat{\boldsymbol{X}}_{s,t} = i(\boldsymbol{X}_{s,t}^{(m)}) = (1, \boldsymbol{X}_{s,t}^1, \dots, \boldsymbol{X}_{s,t}^m, \mathbf{0})\,. \tag{2.28}$$

Of course $\widehat{\boldsymbol{X}}_{st}$ is not multiplicative, but at least it is in $T^{(m+1)}$. Fix a dissection $D = \{s \le t_1 \le \cdots \le t_{i-1} \le t\}$ of $[s,t]$ and define

$$\widehat{\boldsymbol{X}}_{s,t}^{D} = \widehat{\boldsymbol{X}}_{s,t_1} \otimes \widehat{\boldsymbol{X}}_{t_1,t_2} \otimes \cdots \otimes \widehat{\boldsymbol{X}}_{t_{i-1},t} \tag{2.29}$$

using the multiplication in $T^{(m+1)}$. It suffices to show the existence of $\lim_{\mathrm{mesh}(D)\to 0} \widehat{\boldsymbol{X}}_{s,t}^{D}$, for this limit, if it exists, will surely be multiplicative.

To check this last point observe that if the limit exists over $[s,u]$, then it can be attained via dissections D all of which include a fixed $t \in (s,u)$, and so we have

$$\widehat{\boldsymbol{X}}_{s,u}^{D} = \widehat{\boldsymbol{X}}_{s,t}^{D\cap[s,t]} \otimes \widehat{\boldsymbol{X}}_{t,u}^{D\cap[t,u]}\,. \tag{2.30}$$

Taking this limit as the mesh size of D converges to zero we see that we have

$$\big(\lim_{D\to 0} \widehat{\boldsymbol{X}}_{s,t}^{D}\big) = \big(\lim_{D\to 0} \widehat{\boldsymbol{X}}_{s,t}^{D\cap[s,t]}\big) \otimes \big(\lim_{D\to 0} \widehat{\boldsymbol{X}}^{D\cap[t,u]}\big)\,. \tag{2.31}$$

To prove the convergence of $\widehat{\boldsymbol{X}}^{D}$ we see that the difficulty rests in understanding the terms $\big(\widehat{\boldsymbol{X}}_{s,t}^{D}\big)^{m+1}$ for $\big(\widehat{\boldsymbol{X}}_{s,t}^{D}\big)^{j} = \boldsymbol{X}_{s,t}^{j}$ for all $j \le m$ since $\boldsymbol{X}_{s,t}^{(m)}$ is multiplicative.

The heart of our argument is a maximal inequality, the existence of the limit follows by a secondary argument. Our aim is to prove, under the induction hypothesis

$$\begin{aligned}
&\boldsymbol{X}_{u,v}^{(m)} = \{1, \boldsymbol{X}_{u,v}^1, \dots, \boldsymbol{X}_{u,v}^m\} \in T^{(m)}\,,\\
&\boldsymbol{X}_{u,w}^{(m)} = \boldsymbol{X}_{u,v}^{(m)} \otimes \boldsymbol{X}_{v,w}^{(m)}\,,\\
&\|\boldsymbol{X}_{u,v}^{i}\| \le \Big(\frac{(\omega(u,v))^{i/p}}{\beta(i/p)!}\Big)\,, \qquad \text{for all } u<v,\ i \le m\,,
\end{aligned} \tag{2.32}$$

that for *any* dissection D of $[s,t]$

$$\left\| \left(\overset{D}{\widehat{\boldsymbol{X}}}_{s,t} \right)^j \right\| \le \frac{(\omega(s,t))^{j/p}}{\beta(j/p)!}, \qquad \text{for all } j \le m+1 . \tag{2.33}$$

The case where $j < m+1$ is a trivial consequence of our induction hypothesis. The $(m+1)$-tensor $\left(\overset{D}{\widehat{\boldsymbol{X}}}_{s,t}\right)^{m+1}$ is the focus of our attention. Now from the triangle inequality

$$\|(\boldsymbol{X}^D_{st})^{m+1}\| \le \|(\boldsymbol{X}^D_{s,t} - \boldsymbol{X}^{D'}_{s,t})^{m+1}\| + \|(\boldsymbol{X}^{D'}_{st})^{m+1}\| , \tag{2.34}$$

where D' is any other dissection. Suppose that it is obtained from D by dropping a single point from the dissection (this trick seems to be due to L. C. Young). By choosing the point to omit from the dissection carefully, and repeating this deletion procedure until we have the trivial dissection we will obtain our result.

Fix

$$D = \{s = t_0 < t_1 < \cdots < t_r = t\}$$

and use Lemma 2.2.1 to choose j so that

$$\omega(t_{j-1}, t_{j+1}) \le \begin{cases} \left(\dfrac{2}{r-1}\right) \omega(s,t) , & r \ge 3 , \\ \omega(s,t) , & r = 2 . \end{cases} \tag{2.35}$$

Let D' be $D \setminus \{t_j\}$ and consider $\overset{D}{\widehat{\boldsymbol{X}}}_{s,t} - \overset{D'}{\widehat{\boldsymbol{X}}}_{s,t}$. Now[3]

$$\begin{aligned} \overset{D}{\widehat{\boldsymbol{X}}}_{s,t} &= \left(\widehat{\boldsymbol{X}}_{s,t_1} \cdots \widehat{\boldsymbol{X}}_{t_{j-2},t_{j-1}}\right) \widehat{\boldsymbol{X}}_{t_{j-1},t_j} \widehat{\boldsymbol{X}}_{t_j,t_{j+1}} \\ &\qquad \cdot \left(\widehat{\boldsymbol{X}}_{t_{j+1},t_{j+2}} \cdots \widehat{\boldsymbol{X}}_{t_{j-1},t_r}\right) \\ &= \overset{D^-}{\widehat{\boldsymbol{X}}}_{s,t_{j-1}} \widehat{\boldsymbol{X}}_{t_{j-1},t_j} \widehat{\boldsymbol{X}}_{t_j,t_{j+1}} \overset{D^+}{\widehat{\boldsymbol{X}}}_{t_{j+1},t} , \end{aligned} \tag{2.36}$$

while

$$\overset{D'}{\widehat{\boldsymbol{X}}}_{st} = \overset{D^-}{\widehat{\boldsymbol{X}}}_{s,t_{j-1}} \widehat{\boldsymbol{X}}_{t_{j-1},t_{j+1}} \overset{D^+}{\widehat{\boldsymbol{X}}}_{t_{j+1},t} \tag{2.37}$$

[3] To shorten expressions we henceforth drop the use of the $\otimes$ to denote multiplication in the tensor algebra.

and so

$$\widehat{\boldsymbol{X}}^{D}_{st} - \widehat{\boldsymbol{X}}^{D'}_{st} = \widehat{\boldsymbol{X}}^{D^-}_{s,t_{j-1}} \boldsymbol{Z}_{t_{j-1}\,t_j\,t_{j+1}} \widehat{\boldsymbol{X}}^{D^+}_{t_{j+1},t}\,, \tag{2.38}$$

where

$$\boldsymbol{Z}_{t_{j-1}t_jt_{j+1}} = \widehat{\boldsymbol{X}}_{t_{j-1},t_j}\,\widehat{\boldsymbol{X}}_{t_{j-1},t_{j+1}} - \widehat{\boldsymbol{X}}_{t_{j-1},t_{j+1}} \tag{2.39}$$

and using the definition of $\widehat{\boldsymbol{X}}$ and the multiplicative nature of $\boldsymbol{X}$ one has

$$\boldsymbol{Z}_{t_{j-1}t_jt_{j+1}} = \Big(0,\ldots,\mathbf{0},\sum_{1}^{m}\boldsymbol{X}^{i}_{t_{j-1},\,t_j}\,\boldsymbol{X}^{m+1-i}_{t_j,\,t_{j+1}}\Big)\,. \tag{2.40}$$

The only products which yield nonzero results in this tensor multiplication are those where the sum of the degrees of the individual factors is at most m; it follows that we have the reasonably simple expression for the difference

$$\begin{aligned}\widehat{\boldsymbol{X}}^{D}_{s,t} - \widehat{\boldsymbol{X}}^{D'}_{s,t} &= (1,\ldots)\Big(0,\ldots,\mathbf{0},\sum_{1}^{m}\boldsymbol{X}^{i}_{t_{j-1},\,t_j}\,\boldsymbol{X}^{(m+1)-i}_{t_j,\,t_{j+1}}\Big)(1,\ldots)\\ &= \Big(0,\ldots,\mathbf{0},\sum_{1}^{m}\boldsymbol{X}^{i}_{t_{j-1},\,t_j}\,\boldsymbol{X}^{(m+1)-i}_{t_j,\,t_{j+1}}\Big)\,.\end{aligned} \tag{2.41}$$

We can estimate this difference

$$\Big\|\sum_{1}^{m}\boldsymbol{X}^{i}_{t_{j-1}t_j}\boldsymbol{X}^{m+1-i}_{t_jt_{j+1}}\Big\| \le \sum_{1}^{m}\|\boldsymbol{X}^{i}_{t_{j-1}t_j}\|\,\|\boldsymbol{X}^{m+1-i}_{t_jt_{j+1}}\| \tag{2.42}$$

and so using our a priori bound (2.32) for the magnitudes of these tensors

$$\begin{aligned}&\Big\|\sum_{i=1}^{m}\boldsymbol{X}^{i}_{t_{j-1},t_j}\boldsymbol{X}^{m+1-i}_{t_j,t_{j+1}}\Big\|\\ &\qquad\le \sum_{i=0}^{m+1}\Big(\frac{\omega(t_{j-1},t_j)^{i/p}}{\beta(i/p)!}\Big)\Big(\frac{\omega(t_j,t_{j+1})^{(m+1-i)/p}}{\beta((m+1-i)/p)!}\Big)\end{aligned} \tag{2.43}$$

and by the Neo-Classical inequality, Lemma 2.2.2, and superadditivity this is

$$\leq \frac{p^2}{\beta^2}\,\frac{(\omega(t_{j-1},t_j)+\omega(t_jt_{j+1}))^{(m+1)/p}}{((m+1)/p)!} \tag{2.44}$$

$$\leq \frac{p^2}{\beta^2}\,\frac{(\omega(t_{j-1},t_{j+1}))^{(m+1)/p}}{((m+1)/p)!}\,. \tag{2.45}$$

We now recall that we chose our j carefully so that (2.35) held so that if $r > 2$ one has

$$\Big\|\sum_{i=1}^{m}\boldsymbol{X}^{i}_{t_{j-1}t_j}\boldsymbol{X}^{m+1-i}_{t_jt_{j+1}}\Big\| \leq \Big(\frac{2}{r-1}\Big)^{(m+1)/p}\,\frac{p^2}{\beta}\,\frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!} \tag{2.46}$$

and if $r = 2$ one has the similar

$$\Big\|\sum_{i=1}^{m}\boldsymbol{X}^{i}_{t_{j-1}t_j}\boldsymbol{X}^{m+1-i}_{t_jt_{j+1}}\Big\| \leq \frac{p^2}{\beta}\,\frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!}\,. \tag{2.47}$$

Successively dropping points we see that

$$\begin{aligned}\Big\|\big(\overset{D}{\widehat{\boldsymbol{X}}}_{st}\big)^{m+1} - \big(\widehat{\boldsymbol{X}}_{s,t}\big)^{m+1}\Big\| &\\ &\leq \frac{p^2}{\beta}\Big(1+\sum_{r=3}^{\infty}\Big(\frac{2}{r-1}\Big)^{(m+1)/p}\Big)\Big(\frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!}\Big)\\ &= \frac{p^2}{\beta}\,\Big(1+2^{(m+1)/p}\Big(\xi\Big(\frac{m+1}{p}\Big)-1\Big)\Big)\Big(\frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!}\Big)\,.\end{aligned} \tag{2.48}$$

Observing that as $\big(\widehat{\boldsymbol{X}}_{st}\big)^{m+1} = \mathbf{0}$ and

$$\Big(1+\sum_{r=3}^{\infty}\Big(\frac{2}{r-1}\Big)^{(m+1)/p}\Big)$$

is monotone in m and finite because $m+1>p$ we have

$$\begin{aligned}\Big\|\big(\overset{D}{\widehat{\boldsymbol{X}}}_{s,t}\big)^{m+1}\Big\| &\\ &\leq \frac{p^2}{\beta}\Big(1+2^{([p]+1)/p}\Big(\zeta\Big(\frac{[p]+1}{p}\Big)-1\Big)\Big)\,\frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!}\end{aligned} \tag{2.49}$$

(where $\zeta(z) = \sum_1^\infty 1/n^z$ is the traditional Riemann zeta function). Thus if we choose

$$\beta \ge p^2\Big(1 + 2^{([p]+1)/p}\Big(\zeta\Big(\frac{[p]+1}{p}\Big) - 1\Big)\Big)\,, \tag{2.50}$$

we get the estimate

$$\big\|\big(\widehat{\boldsymbol{X}}^D_{s,t}\big)^{m+1}\big\| \le \frac{\omega(s,t)^{(m+1)/p}}{\beta((m+1)/p)!}\,, \tag{2.51}$$

for *all choices of dissection* D. This completes the proof of the maximal inequality.

Now we must show convergence of the products. It is at this point that we require our control ω on the p-variation to be regular. We will show that our sequence $\widehat{\boldsymbol{X}}^D$ satisfies a Cauchy convergence principle. Consider two dissections D, $\tilde{D}$ both having mesh size less than δ. We can always find a common refinement $\hat{D}$ of D and $\tilde{D}$. We fix some interval in $[t_j, t_{j+1}] \in D$; then the refinement $\hat{D}$ breaks the interval up into a number of pieces $t_j \le s_{j_1} \le \cdots \le s_{j_r} = t_{j+1}$; call the dissection $\hat{D}_j$. Then, we know from the maximal inequality, how to estimate

$$\big(\widehat{\boldsymbol{X}}^{\hat{D}_j}_{t_j t_{j+1}} - \widehat{\boldsymbol{X}}_{t_j t_{j+1}}\big)^{m+1}$$

and all terms of degree less than $m+1$ in the difference are zero because $\boldsymbol{X}$ is multiplicative. Therefore

$$\big\|\widehat{\boldsymbol{X}}^{\hat{D}_j}_{t_j t_{j+1}} - \widehat{\boldsymbol{X}}_{t_j t_{j+1}}\big\| \le \frac{\omega(t_j, t_{j+1})^{m+1/p}}{\beta((m+1)/p)!}\,. \tag{2.52}$$

So the total difference[4]

$$\big(\widehat{\boldsymbol{X}}^{\hat{D}_j} - \widehat{\boldsymbol{X}}^{D}\big)^{m+1} \tag{2.53}$$

is controlled in norm by

$$\begin{aligned}
&\Big(\sum_D \frac{\omega(t_j, t_{j+1})^{(m+1)/p}}{\beta((m+1)/p)!}\Big)\\
&\quad\le \frac{1}{\beta((m+1)/p)!}\,\max_D(\omega(t_j,t_{j+1}))^{(m+1)/p-1}\sum_D \omega(t_j,t_{j+1})\\
&\quad\le \frac{1}{\beta((m+1)/p)!}\,\max_D(\omega(t_j,t_{j+1}))^{(m+1)/p-1}\,\omega(s,t)\,,
\end{aligned} \tag{2.54}$$

[4] using a simple extension of the argument used in (2.38)-(2.40) which drastically limited the range of terms which contribute to the difference of the products.

which is independent of $\hat{D}$ and by the regularity of ω this converges uniformly to zero as the mesh size of D converges to zero. Applying the triangle inequality, we have a uniform bound on $\boldsymbol{X}^{\tilde{D}} - \boldsymbol{X}^{D}$ as required. It follows that we have established the existence of a multiplicative functional satisfying all the requirements of the induction.

Uniqueness. We must show that if $\boldsymbol{X}_{s,t}$ and $\boldsymbol{Y}_{s,t}$ are two multiplicative functionals which agree up to the m-th degree, so that $\boldsymbol{X}^i_{st} = \boldsymbol{Y}^i_{st}$, $i \leq m$, and which are both of regular finite p-variation where $(m+1)/p > 1$ then they agree. The following algebraic lemma makes the situation clear.

Lemma 2.2.3. *Suppose that $\boldsymbol{X}_{s,t}$ and $\boldsymbol{Y}_{s,t}$ are multiplicative functionals in $T^{(m+1)}$ which agree up to the m-th degree so that $\boldsymbol{X}^i_{st} = \boldsymbol{Y}^i_{st}$, $i \leq m$. The difference function $\boldsymbol{\Psi}_{s,t}$*

$$\boldsymbol{\Psi}_{s,t} = \boldsymbol{X}^{m+1}_{s,t} - \boldsymbol{Y}^{m+1}_{s,t} \in \bigotimes_{i=1}^{m+1} V \tag{2.55}$$

is additive

$$\boldsymbol{\Psi}_{s,t} + \boldsymbol{\Psi}_{t,u} = \boldsymbol{\Psi}_{s,u} \ . \tag{2.56}$$

Conversely, if $\boldsymbol{X}_{s,t}$ is a multiplicative functional in $T^{(m+1)}$ and $\boldsymbol{\Psi}_{s,t}$ is additive in $V^{\otimes m+1}$ then $\boldsymbol{X}_{s,t} + \boldsymbol{\Psi}_{s,t}$ is also a multiplicative functional.

REMARK 2.2.1. This easy result reflects the nilpotent nature of the algebraic structures we are interested in, the function $\boldsymbol{\Psi}_{s,t}$ lies in the centre.

PROOF. Use the multiplicative property for $\boldsymbol{X}_{s,t}$ and $\boldsymbol{Y}_{s,t}$ to observe that

$$\begin{aligned}
(\boldsymbol{Y}_{s,u})^{m+1} &= (\boldsymbol{Y}_{s,t} \otimes \boldsymbol{Y}_{t,u})^{m+1} \\
&= \boldsymbol{Y}^{m+1}_{s,t} + \boldsymbol{Y}^{m+1}_{t,u} + (\boldsymbol{X}_{s,t} \otimes \boldsymbol{X}_{t,u})^{m+1} - \boldsymbol{X}^{m+1}_{s,t} - \boldsymbol{X}^{m+1}_{t,u} \\
&= (\boldsymbol{Y}_{s,u})^{m+1} \\
&= (\boldsymbol{X}_{s,u})^{m+1} + (\boldsymbol{Y}^{m+1}_{s,t} - \boldsymbol{X}^{m+1}_{s,t}) + (\boldsymbol{Y}^{m+1}_{t,u} - \boldsymbol{X}^{m+1}_{t,u})
\end{aligned} \tag{2.57}$$

and so our claim is verified

$$\boldsymbol{\Psi}_{s,u} = \boldsymbol{\Psi}_{s,t} + \boldsymbol{\Psi}_{t,u} \ . \tag{2.58}$$

The same identity also makes it clear that if $\boldsymbol{X}_{s,t}$ is multiplicative on $T^{(n+1)}$ and $\boldsymbol{\Psi}$ satisfies (2.56) and is in $\otimes_i^{n+1} V$, then $\boldsymbol{X}_{s,t} + \boldsymbol{\Psi}_{s,t}$ is also multiplicative.

Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ have finite p-variation controlled by a regular ω where $(m+1)/p > 1$. By assumption, there is a constant so that

$$\|\boldsymbol{\Psi}_{st}\| \leq c\,\omega(s,t)^{(m+1)/p}\,. \tag{2.59}$$

and so $\boldsymbol{\Psi}_{0,t}$ is a conventional path of finite $(m+1)/p$-variation. If $(m+1)/p > 1$ and ω is regular, it follows that $\boldsymbol{\Psi}$ is identically zero and uniqueness follows.

These calculations also establish the remarks we made on the non-uniqueness of extensions of multiplicative functionals if $(m+1)/p \leq 1$ as in this case perturbing an extension by a continuous additive $\boldsymbol{\Psi}_{s,t}$ of bounded variation will produce a different extension of finite p-variation.

2.2.2. Continuity.

We have shown that the high order multiplicative functionals are uniquely determined by the low order ones if we impose a p-variation condition. We also defined a natural distance between paths of finite p-variation. The map we have defined is continuous, and there is a very explicit estimate for the modulus of continuity.

Theorem 2.2.2. *Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are multiplicative functionals in $T^{(n)}$ of finite p-variation controlled by ω where $(n+1)/p > 1$. Suppose further that for some $\varepsilon < 1$ one has*

$$\|\boldsymbol{X}^i_{s,t} - \boldsymbol{Y}^i_{s,t}\| \leq \varepsilon\,\frac{\omega(s,t)^{i/p}}{\gamma(i/p)!}\,, \tag{2.60}$$

for all $i \leq n$. Then for a suitable choice of γ,

$$\gamma \geq 3\,p^2\Big(1 + 2^{([p]+1)/p}\Big(\zeta\Big(\frac{[p]+1}{p}\Big) - 1\Big)\Big) \tag{2.61}$$

will do, one has

$$\|\boldsymbol{X}^i_{s,t} - \boldsymbol{Y}^i_{s,t}\| \leq \varepsilon\,\frac{\omega(s,t)^{i/p}}{\gamma(i/p)!}\,, \tag{2.62}$$

for all $i < \infty$ where $\boldsymbol{X}^i$ and $\boldsymbol{Y}^i$ are, for $i > n$, the components in $V^{\otimes i}$ of the multiplicative extension of finite p-variation.

PROOF. Proceed by induction. Suppose $n+1 > p$. Recall how we constructed $\boldsymbol{X}^{m+1}$ and $\boldsymbol{Y}^{m+1}$ from $\boldsymbol{X}^{(m)}$ and $\boldsymbol{Y}^{(m)}$ by taking the limit of the products $\boldsymbol{X}^D$, $\boldsymbol{Y}^D$. Recall in particular, that our choice of dissection in the proof of the maximal inequality depended on ω alone and not on $\boldsymbol{X}^{(n)}$ or $\boldsymbol{Y}^{(n)}$. So we may select the same coarsening sequence of dissections in the analysis bounding $\boldsymbol{X}^D$ and $\boldsymbol{Y}^D$. We may also use this sequence of dissections to estimate $\|\boldsymbol{X}^D - \boldsymbol{Y}^D\|$. As we coarsen the dissection we have

$$\begin{aligned} \|(\boldsymbol{X}^D - \boldsymbol{Y}^D)^{m+1}\| \leq \|(\boldsymbol{X}^D - \boldsymbol{X}^{D'})^{m+1} - (\boldsymbol{Y}^D - \boldsymbol{Y}^{D'})^{m+1}\| \\ + \|(\boldsymbol{X}^{D'} - \boldsymbol{Y}^{D'})^{m+1}\| . \end{aligned} \tag{2.63}$$

Estimate the first term on the right side of the expression.

$$(\boldsymbol{X}^{D'}_{s,t} - \boldsymbol{X}^D_{s,t})^{n+1} = \sum_{1\leq j\leq n} \boldsymbol{X}^j_{t_{i-1}t_i} \boldsymbol{X}^{n+1-j}_{t_i t_{i+1}} \tag{2.64}$$

and

$$\boldsymbol{Y}^j_{s,t} = \boldsymbol{X}^j_{s,t} + \boldsymbol{R}^j_{s,t} , \tag{2.65}$$

so

$$\begin{aligned} &((\boldsymbol{X}^{D'}_{s,t} - \boldsymbol{X}^D_{s,t}) - (\boldsymbol{Y}^{D'}_{s,t} - \boldsymbol{Y}^D_{s,t}))^{n+1} \\ &\quad = \sum_{1\leq j\leq n} (\boldsymbol{X}^j_{t_{i-1},t_i} \boldsymbol{X}^{n+1-j}_{t_i,t_{i+1}} \\ &\qquad - (\boldsymbol{X}^j_{t_{i-1},t_i} + \boldsymbol{R}^j_{t_{i-1},t_i}) (\boldsymbol{X}^{n+1-j}_{t_i,t_{i+1}} + \boldsymbol{R}^{n+1-j}_{t_i,t_{i+1}})) \end{aligned} \tag{2.66}$$

and by exploiting induction and the neo-classical inequality one has

$$\begin{aligned} &\|((\boldsymbol{X}^{D'}_{s,t} - \boldsymbol{X}^D_{s,t}) (\boldsymbol{Y}^{D'}_{st} - \boldsymbol{Y}^D_{s,t}))^{n+1}\| \\ &\qquad \leq (2\,\varepsilon + \varepsilon^2)\, p^2 \frac{\omega(t_{i-1},t_{i+1})^{(n+1)/p}}{\gamma^2((n+1)/p)!} \end{aligned} \tag{2.67}$$

and as before, summing over our carefully chosen and successively coarsening partitions one has

$$\begin{aligned} \|\boldsymbol{X}^D_{s,t} - \boldsymbol{Y}^D_{s,t}\| \leq \frac{1 + 2^{(n+1)/p}(\zeta((n+1)/p) - 1)}{\gamma^2} \\ \cdot p^2 \frac{\omega(s,t)^{(n+1)/p}}{((n+1)/p)!} (2\,\varepsilon + \varepsilon^2) . \end{aligned} \tag{2.68}$$

So for

$$\gamma \geq \Bigl(1 + 2^{([p]+1)/p}\Bigl(\xi\Bigl(\frac{[p]+1}{p}\Bigr) - 1\Bigr)\Bigr)p^2\,(2\,\varepsilon + \varepsilon^2)$$

the result follows. In particular if $\varepsilon < 1$ the required estimate holds. This completes the induction step and this rather explicit continuity result follows.

Remark 2.2.2. It might be thought that we have introduced a variety of topologies on the space of paths of finite p-variation in the above theorem; however, they can all be pasted together in the most natural way.

Definition 2.2.1. *We say a pair of paths $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $T^{(n)}$ which have regular finite p-variation are at most a distance ε apart if*

$$\begin{aligned}&\omega^{X,Y}(s,t)\\&\quad = \sup_{s,t\in J}\Bigl\{\Bigl(\sum_j |||\boldsymbol{X}_{t_j,t_{j+1}} - \boldsymbol{Y}_{t_j,t_{j+1}}|||^p\Bigr),\ s \leq t_{j_1} < \cdots < t_{j_r} \leq t\Bigr\} \leq \varepsilon\,,\end{aligned}$$

$$\sup\{|||\boldsymbol{X}_t - \boldsymbol{Y}_t|||,\ t \in J\} \leq \varepsilon\,.$$

It is elementary that such a distance is complete, and that if a sequence converges in the sense that we introduced and exploited in the preceding lemma then it also converges in this new sense.

Consider a sequence $U_t^{(n)}$ of paths converging to a path U_t^0. Then the p-variation of $U_t^{(n)}$, denoted by $\omega^{U^{(n)}}(s,t)$, and $\omega^{U^{(n)},U^{(0)}}(s,t)$ are continuous and zero on the diagonal because of the regularity of the paths. We may choose and re-label a subsequence so that

$$\sup_{s,t}\,\omega^{U^{(n)},U^{(0)}}(s,t) < 4^{-n}$$

on J. Consider the new superadditive functional

$$\Psi(s,t) = \sup_n\,\omega^{U^{(n)}}(s,t) + \sum_n 2^n\,\omega^{U^{(n)},U^{(0)}}(s,t) \tag{2.69}$$

and observe that it is continuous (note that the supremum of a sequence of continuous and uniformly converging functions is itself continuous), it is obviously superadditive and zero on the diagonal; it therefore provides a regular control on the p-variation of all the paths we are considering, and most importantly, satisfies $\omega^{U^{(n)},U^{(0)}}(s,t) \leq 2^{-n}\,\psi(s,t)$. This essentially concludes the remark. Every convergent sequence in the weaker sense has a subsequence converging in this stronger dominated sense, and so we see that the notions of convergent sequence must correspond.

In a metric space, the topology is determined by the convergent sequences.

2.2.3. The neo-classical inequality: a proof.

Theorem 2.2.3. *The following inequality holds uniformly in $p \geq 1$, n*

$$\frac{1}{p^2}\sum_{j=0}^{n} \frac{a^{j/p}\, b^{(n-j)/p}}{(j/p)!\,((n-j)/p)!} \leq \frac{(a+b)^{n/p}}{(n/p)!}\,, \qquad a, b > 0\,. \tag{2.70}$$

REMARK 2.2.3. For our application we only require this inequality with some constant in place of $1/p^2$ which is independent of a and n. However, it is interesting to ask what is the best uniform estimate in all the variables. All numerical evidence and proofs of special cases suggest the inequality is true with $1/p$ in place of $1/p^2$ and that in this form the inequality is very strongly saturated (with equality to the n-th degree as p approaches one if $a = b$. When $p = 1$, we have equality of the left and right expressions by the binomial theorem in either form. When $p = n$ we can prove the result in its strengthened form with $1/p$.

PROOF. To prove the inequality in the form stated, it suffices to establish that

$$\frac{1}{p}\sum_{j=0}^{n} x^{j/p}\,(1-x)^{(n-j)/p} \frac{(n/p)!}{(j/p)!\,((n-j)/p)!} \leq p\,, \tag{2.71}$$

because the expression (2.70) is homogeneous under scaling of a and b. Moreover, we have an integral expression for the special functions

$$\begin{aligned}\left(\frac{x!\,y!}{(x+y)!}\right)^{-1} &= \frac{1}{(x+y+1)\,\beta(x+1,y+1)} \\ &= \frac{1}{(x+y+1)\displaystyle\int_0^1 u^x(1-u)^y\,du}\,.\end{aligned} \tag{2.72}$$

We may rewrite the left hand of the expression (2.71)

$$
\begin{aligned}
&\frac{1}{p}\sum_{0}^{n} x^{j/p}\,(1-x)^{(n-j)/p}\,\frac{(n/p)!}{(j/p)!\,((n-j)/p)!}\\
&\qquad = \frac{1}{n+p}\sum_{0}^{n}\frac{x^{j/p}\,(1-x)^{(n-j)/p}}{\int_0^1 (u^j\,(1-u)^{n-j})^{1/p}\,du}\\
&\qquad = \frac{1}{n+p}\sum_{0}^{n}\frac{1}{\int_0^1 \left(\frac{u}{x}\right)^{j/p}\left(\frac{1-u}{1-x}\right)^{(n-j)/p} du}\,.
\end{aligned}
$$

We now make a substitution: $v = p/n$, $\theta_j = j/n$. Then the individual terms in the above sum are derived from

$$
F_\theta(x,v) = \frac{1}{n\,(v+1)}\,\frac{x^{\theta/\nu}\,(1-x)^{(1-\theta)/\nu}}{\int_0^1 (u^\theta\,(1-u)^{1-\theta})^{1/v}\,du}\,. \tag{2.73}
$$

By the binomial theorem

$$
\sum_{0}^{n} F_{\theta_j}\Big(x,\frac{1}{n}\Big) \equiv 1\,, \tag{2.74}
$$

for all n and all $x \in [0,1]$. If we could also prove that

$$
\sum_{0}^{n} F_{\theta_j}(x,v) \le 1\,, \tag{2.75}
$$

for all $v > 1/n$, and for all x then we would have established the stronger result which we believe is true. To do this it would suffice to show that

$$
\Big(\frac{\partial}{\partial x}\,(x\,(1-x))\,\frac{\partial}{\partial x} - \frac{\partial}{\partial v}\Big)F_\theta \ge 0\,, \tag{2.76}
$$

for in this case we could use the maximum principle for sub-parabolic functions to deduce that any positive linear combination of F_θ, taken over varying θ, attains its maximum over the region $v > 1/n$, $x \in (0,1)$ on its parabolic boundary. In particular we could conclude that $\sum_0^n F_{\theta_j}(x,v) \le 1$, for all x and for all $v > 1/n$.

But F_θ is not a subsolution. On the positive side, we can prove that

$$\frac{1}{v}\, x^{\theta/v}\,(1-x)^{(1-\theta)/v}\Big(\int (u^{\theta}\,(1-u)^{1-\theta})^{1/v}\,du\Big)^{-1} \tag{2.77}$$

is a subsolution for any choice of θ. We can therefore apply a maximum principle argument to prove that if $\theta_j = j/n$ then

$$\begin{aligned}\sum_{j=0}^{n}\frac{v+1}{v}\,F_{\theta_j}(v,u) &\le \sup_{u\in[0,1]}\sum_{j=0}^{n}\frac{\frac{1}{n}+1}{\frac{1}{n}}\,F_{\theta_j}\Big(\frac{1}{n},u\Big)\\ &= \frac{\frac{1}{n}+1}{\frac{1}{n}}\\ &= n+1\,,\end{aligned} \tag{2.78}$$

for $v > 1/n$. We may cross-multiply and substitute to obtain

$$\sum_{j=0}^{n} F_{\theta_j}(v,u) \le \frac{v\,(n+1)}{v+1} = \frac{p\,(n+1)}{p+n}\,. \tag{2.79}$$

As the inequality $v > 1/n$ is equivalent to $p \ge 1$, we may deduce that for $v > 1/n$ and $u \in [0,1]$ the inequality

$$\sum_{j=0}^{n} F_{\theta_j}(v,u) \le p \tag{2.80}$$

holds, concluding our main argument.

However, it remains to prove that our expression (2.77) is indeed a subsolution to the parabolic equation (2.76). This is elementary, but relatively delicate.

Because our expression is positive, we may work with its logarithm. Observe that as a general fact a parabolic operator applied to an exponential has a simple form

$$\begin{aligned}Le^{U} &:= \frac{\partial}{\partial u}\,\varphi\,\frac{\partial}{\partial u}\,e^{U} - \frac{\partial e^{U}}{\partial v}\\ &= \Big(\frac{\partial}{\partial u}\,\varphi\,\frac{\partial}{\partial u}\,U + |\nabla_{\varphi}U|^{2} - \frac{\partial U}{\partial v}\Big)e^{U}\,,\end{aligned} \tag{2.81}$$

where we define

$$|\nabla_\varphi u|^2 = \varphi\, |\nabla u|^2\,. \tag{2.82}$$

To show that the exponential e^U is a subsolution it suffices to show that

$$\Big(\frac{\partial}{\partial u}\,\varphi\,\frac{\partial}{\partial u}\,U + |\nabla_\varphi U|^2 - \frac{\partial U}{\partial v}\Big) \geq 0\,. \tag{2.83}$$

The log of the expression (2.77) is

$$\begin{aligned} &-\log v + \frac{\theta}{v}\log x + \frac{1-\theta}{v}\log(1-x)\\ &-\log\int_0^1 (u^\theta(1-u)^{1-\theta})^{1/v} du\,. \end{aligned} \tag{2.84}$$

Let us apply our identity for Le^U, one term at a time, with U given by the expression (2.84) above.

$$\frac{\partial}{\partial x}\,x\,(1-x)\,\frac{\partial}{\partial x}\,U = \frac{\partial}{\partial x}\Big(\frac{\theta}{v}\,(1-x)\Big) - \frac{\partial}{\partial x}\Big(\frac{1-\theta}{v}\,x\Big) = -\frac{1}{v} \tag{2.85}$$

and

$$\begin{aligned} x\,(1-x)\,\Big|\frac{\partial}{\partial x}\,U\Big|^2 &= x\,(1-x)\Big(\frac{\theta}{v}\,\frac{1}{x} - \frac{1-\theta}{v}\,\frac{1}{1-x}\Big)^2\\ &= \Big(\frac{\theta-x}{v}\Big)^2\,\frac{1}{x\,(1-x)}\,. \end{aligned} \tag{2.86}$$

On the other hand the expression (2.84) can also be rewritten as

$$-\log v - \log\Big(\int\Big(\Big(\frac{u}{x}\Big)^\theta\Big(\frac{1-u}{1-x}\Big)^{1-\theta}\Big)^{1/v}du\Big)\,. \tag{2.87}$$

So

$$\begin{aligned} &-\frac{\partial}{\partial v}\,U\\ &= \frac{1}{v} - \frac{1}{v^2}\,\frac{\int \Big(\theta\log\Big(\frac{u}{x}\Big) + (1-\theta)\log\Big(\frac{1-u}{1-x}\Big)\Big)\Big(\Big(\frac{u}{x}\Big)^\theta\Big(\frac{1-u}{1-x}\Big)^{1-\theta}\Big)^{1/v} du}{\int\Big(\Big(\frac{u}{x}\Big)^\theta\Big(\frac{1-u}{1-x}\Big)^{1-\theta}\Big)^{1/v}du}\\ &= \frac{1}{v} - \frac{1}{v^2}\Big(\theta\log\Big(\frac{\theta}{x}\Big) + (1-\theta)\log\Big(\frac{1-\theta}{1-x}\Big)\Big)\\ &\quad + \frac{1}{v^2}\,\frac{\int\Big(\theta\log\Big(\frac{\theta}{u}\Big) + (1-\theta)\log\Big(\frac{1-\theta}{1-u}\Big)\Big)\Big(\Big(\frac{u}{x}\Big)^\theta\Big(\frac{1-u}{1-x}\Big)^{1-\theta}\Big)^{1/v}du}{\int\Big(\Big(\frac{u}{x}\Big)^\theta\Big(\frac{1-u}{1-x}\Big)^{1-\theta}\Big)^{1/v}du} \end{aligned} \tag{2.88}$$

and applying Jensen's inequality to the convex function $x\log x$,

$$\left(\theta\log\left(\frac{\theta}{u}\right)+(1-\theta)\log\left(\frac{1-\theta}{1-u}\right)\right) \tag{2.89}$$

in the last integral, and hence the integral itself is always positive. Collecting the terms together we see that (2.83) will hold providing we can show

$$f(\theta,x)=\frac{(\theta-x)^2}{x\,(1-x)}-\left(\theta\log\left(\frac{\theta}{x}\right)+(1-\theta)\log\left(\frac{1-\theta}{1-x}\right)\right)\geq 0\,, \tag{2.90}$$

for all pairs θ, $x\in[0,1]$. This will follow through a study of $\partial f(\theta,x)/\partial x$. This derivative is 0 at $x=\theta$. If we prove it to be positive for $x\geq\theta$ and negative for $x<\theta$ then the result follows since $f(\theta,\theta)=0$. But

$$\begin{aligned}\frac{\partial f}{\partial x}&=(x-\theta)\,\frac{x\,(1-x)+(1-2\,x)\,(\theta-x)}{(x\,(1-x))^2}\\&=(x-\theta)\,\frac{(x-\theta)^2+(\theta-\theta^2)}{(x\,(1-x))^2}\end{aligned} \tag{2.91}$$

and the second factor in the last expression is positive because $\theta\in[0,1]$.

This completes the proof of the neo-classical inequality.

2.3. Multiplicative functionals - The basic spaces of paths.

We can now identify the basic classes of objects which drive differential equations.

Definition 2.3.1. *A p-multiplicative functional is a multiplicative functional of degree $[p]$ and finite p-variation, taking its values in $T(V)^{([p])}$. We denote the set of such paths by $\Omega(V)^p$. The elements of $\Omega(V)^p$ with $\boldsymbol{X}_{s,t}\in G^{([p])}$ for all pairs of times s, t are the geometric p-multiplicative functionals denoted by $\Omega G(V)^p$.*

Within these spaces, we will often refine our interest and consider only multiplicative functionals which are controlled by a given regular ω.

The constraint defining $\Omega G(V)^p$ as a subspace of $\Omega(V)^p$ is a purely algebraic one; and for this reason it is obvious that it defines a closed subset. On the other hand $\Omega G(V)^p$ has a very important analytic interpretation. The class $S(V)$ of piecewise smooth paths can be lifted to *a subset $S(V)_p$ of $\Omega(V)^p$ in a canonical way using the first $[p]$ iterated integrals and, as we have shown, Chen observed that the embedding is actually into $\Omega G(V)^p$.*

Lemma 2.3.1. *The closure of $S(V)_p$ in $\Omega(V)^p$ is $\Omega G(V)^p$.*

The proof of this lemma is quite routine and so we only sketch it. Fix a group-like multiplicative functional $\boldsymbol{X}$. Suppose that it has finite p-variation controlled by a regular ω. We must construct piecewise smooth paths whose iterated integrals approximate it. However, given an element g of the group $G^{(n)}$ there is always a smooth path whose first n iterated integrals at time one agree with g. Among these paths the one with shortest projected distance in V has been closely studied [30]. In any case, its p-variation in a compact neighborhood of the identity in $G^{(n)}$ will be uniformly comparable[5] with $|||g-1|||$. As a consequence, we see that the paths obtained by taking the original multiplicative functional, fixing a dissection, and then replacing the intermediate segments of the multiplicative functional by these "chords" re-parameterised so that they are transversed according to the times in our dissection provide an approximating family of piecewise smooth multiplicative functionals. The regularity of ω ensures convergence.

The class of geometric multiplicative functionals will be of great importance later. A number of questions that remain open relate to the possible extension of theorems from $\Omega G(V)^p$ to $\Omega(V)^p$. Such an extension corresponds to the extension from Stratonovich to Itô in the classical probabilistic setting. In this paper, we will frequently use the above lemma to obtain results for the geometric p-functionals that we do not know how to prove more generally. We hope to understand matters better, and return to this issue in a later paper.

2.3.1. Inhomogeneous degrees of smoothness.

Consider the equation

$$dy_t = \sum_i f^i(y_t)\,dx_t^i + f^0(y_t)\,dt\,, \tag{2.92}$$

by taking our driving signal to be (x_t, t) everything we said previously applies. However, this is an analytically wasteful approach as we fail

[5] The bound will depend on the values of n and p.

to take advantage of the smoother character of one of the co-ordinates in contrast with the others. So we remark now that at the price of increased notational complexity, one may introduce a notion of multiplicative functional $\boldsymbol{X}_{s,t}$ of finite $p = (p_1, \ldots, p_d)$ variation controlled by ω.

Definition 2.3.2. *A path $\boldsymbol{X}_{s,t}$ in $T(V^1 \oplus \cdots \oplus V^d)$ is of finite $p = (p_1, \ldots, p_d)$ variation controlled by ω providing the component*

$$\boldsymbol{X}_{s,t}^{(r_1,\ldots,r)} \in V^{r_1} \otimes \cdots \otimes V^{r_l}\,, \tag{2.93}$$

where $r_i \in \{1, \ldots, d\}$ satisfies

$$\|\boldsymbol{X}_{s,t}^{(r_1,\ldots,r)}\| \leq \frac{\omega(s,t)^{l_1/p_1+\cdots+l_d/p_d}}{\beta^d\,(l_1/p_1)!\cdots(l_d/p_d)!}\,, \tag{2.94}$$

where

$$l_j = \frac{|\{i:\ r_i = j\}|}{l}\,. \tag{2.95}$$

In this case it is easy to see that essentially the same arguments and definitions can be applied to get existence, uniqueness and continuity theorems. The crucial point is that to get existence, and a uniqueness theorem, one must know all the components of the multiplicative functional for which $l_1/p_1 + \cdots + l_d/p_d < 1$. The arguments vary scarcely at all.

2.4. Differential equations driven by rough signals - The linear case.

2.4.1. The flow induced by a rough multiplicative functional.

We now draw out some applications of our first theorems on multiplicative functionals.

Recall that a linear differential equation is one where the target manifold (where y_t takes its values) is a Banach space, and the linear map from the space V carrying the driving signal x_t has as its range vector fields that are *bounded linear* maps

$$x \longrightarrow A(\,)\,x : V \longmapsto \hom(W, W)\,. \tag{2.96}$$

In our general form an equation can be reparameterised in this way if the vector fields define a finite dimensional Lie algebra.

If x_t is a smooth path then, as we saw previously, the linear flow associated to the linear equation

$$\begin{cases} dy_t = A(y)\,dx_t\,, \\ d\pi_t = A(\cdot)\,dx_t\,\pi_t\,, \end{cases} \tag{2.97}$$

can be recovered as the sum of the convergent Einstein series

$$\pi_{s,t} = I + A \int\limits_{s<u<t} dx_u + AA \iint\limits_{s<u_1<u_2<t} dx_{u_1}\,dx_{u_2} + \cdots \tag{2.98}$$

The theorems in the last section associate to any element X in $\Omega(V)^p$ a unique multiplicative functional $\boldsymbol{X}_{s,t} = (1, \boldsymbol{X}^1_{s,t}, \ldots, \boldsymbol{X}^{[p]}_{s,t}, \boldsymbol{X}^{[p]+1}_{s,t}, \ldots)$ of arbitrarily high (and hence of infinite) degree and finite p-variation. Because the terms $\boldsymbol{X}^i_{s,t}$ decay like

$$\frac{1}{(i/p)!} \tag{2.99}$$

and this is faster than any geometric series grows, the series

$$\pi_{s,t} = I + A\boldsymbol{X}^1_{s,t} + AA\boldsymbol{X}^2_{s,t} + \cdots \tag{2.100}$$

converges absolutely to an operator in $\hom(W,W)$. Moreover the mapping is obviously continuous from $\Omega(V)^p$.

Lemma 2.4.1. *The map*

$$\pi_{s,t} = I + A\boldsymbol{X}^1_{s,t} + AA\boldsymbol{X}^2_{s,t} + \cdots \tag{2.101}$$

from $\Omega(V)^p$ *to* $\hom(W,W)$ *respects multiplication. That is to say* $\pi_{s,t}\,\pi_{t,u} = \pi_{s,u}$.

Remark 2.4.1. The fact that we can find a multiplicative extension of our map from geometric paths to all p-multiplicative paths indicates that the role of $\Omega(V)^p$ relative to $\Omega G(V)^p$ is very similar to that of the enveloping algebra to the Lie group.

PROOF. If s, t, u are in V then

$$(A\,v)(AA\,t\otimes u) = AAA\,v\otimes t\otimes u\,, \qquad \text{etc.} \tag{2.102}$$

From this observation, the multiplicative property of X, and the absolute convergence of all the series the result is immediate.

We could state a more abstract form of the above result.

Corollary 2.4.1. *Suppose A is a bounded map from a Banach space V into any Banach algebra Q then the map*

$$d\pi_{t_0,t} = \pi_{t_0,t}\,A\,dx_t\ , \qquad \pi_{t_0,t_0} = 1\,, \tag{2.103}$$

defined on smooth paths in V extends in a unique continuous way to the geometric multiplicative functionals of finite p-variation in $\Omega G(V)^p$ and more generally to any regular multiplicative functional of p-variation. The map is multiplicative on $\Omega(V)^p$.

Although this allows us to give a meaning to (2.97) for elements of $\Omega(V)^p$, we only feel 100% confident about calling it a *solution* in the case where $\boldsymbol{X}$ is an element of $\Omega G(V)^p$. The reason for our nervousness is that if we apply the functional that we have just identified to an element of $\Omega(V)^p$ that is not geometric, then the resulting operator is no longer a path in the underlying Lie group, but an element of the enveloping algebra. In other words, the natural solution to an Itô equation is not a randomly evolving flow on the manifold, but rather an evolving differential operator. Only the use of a connection can bring it back to a flow.

Iterated integrals for solutions to linear equations.[6] We have established that the Itô functional associated to a linear differential equation can be extended to a continuous multiplicative function from $\Omega G(V)^p$ in a unique way. But our solution was a flow, or a path π_t in the algebra of linear homomorphisms of W to itself. By evaluating it against a single vector w we get the solution $y_t = \pi_t w$ which starts at w. At least in the linear case it would seem that all is complete. But this is not really the case. The point is that we would like the

[6] The remarks in this section are far more significant than the reader might appreciate on first inspection.

solutions to our equation to be of the same class as the driving signal. It is obvious from our estimates that the solution y_t is a path in W of finite p-variation. But we have seen that such paths are not the correct objects with which to drive differential equations, we also require the iterated integrals of low degree.

For smooth driving paths x_t we can obviously construct all the iterated integrals of y_t and the joint iterated integrals of x_t with y_t. This defines a map from $S(V)$ into $\Omega G(V \oplus W)^p$. The question we aim to answer in this section is the following: can we extend that definition to one valid for any path in $\Omega G(V)^p$, or even to any path in $\Omega(V)^p$? We only have a general answer in the former case which we now explain. (Understanding how to make the extension to $\Omega(V)^p$ is the key to generalising Itô's type of differential equation to rougher paths).

Consider the equation (2.97) driven by a piecewise smooth path. The solution is again piecewise smooth, moreover the series solution converges locally uniformly at the level of derivatives. Therefore we have the expression for the iterated integrals of y

$$\begin{aligned} \boldsymbol{Y}^i_{s,t} &= \iint\limits_{s<u_1<\cdots<u_i<t} dy_{u_1}\cdots dy_{u_i} \\ &= \iint\limits_{s<u_1<\cdots<u_i<t} \sum_{l_1=1}^{\infty} A^{l_1}(d\boldsymbol{X}^{l_1}_{s,u_1})\cdots \sum_{l_1=1}^{\infty} A^{l_i}(d\boldsymbol{X}^{l_i}_{s,u_i})\; y_s^{\otimes i}\,. \end{aligned} \tag{2.104}$$

Providing we can justify changing the order of summation of the series we have the alternative expression

$$\sum_{S=r}^{\infty} \sum_{\substack{l_1+\cdots+l_i=S \\ l_j\geq 1}} A^{l_1}\otimes\cdots\otimes A^{l_i} \iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1}\cdots d\boldsymbol{X}^{l_i}_{s,u_i}\, y_s^{\otimes i} \tag{2.105}$$

where

$$A^{l_1}\otimes\cdots\otimes A^{l_i} : V^{\otimes(l_1+\cdots+l_i)} \longrightarrow \hom\left(W^{\otimes i}, W^{\otimes i}\right) \tag{2.106}$$

is the obvious induced map.

To obtain the absolute convergence of the series, and the continuous extension of the map to $\Omega G(V)^p$, we must look a bit more closely at the expression for

$$\iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i} \; . \tag{2.107}$$

At this point we exploit in a critical way the fact that we are dealing with iterated integrals of the classical kind and are not working with abstract multiplicative functionals. Now

$$\begin{aligned} &\iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i} \\ &\qquad = \iint\limits_{\substack{s<u_1<\cdots<u_i<t \\ s<u_{1,1}<\cdots<u_{l_1}=u_{1,l_1} \\ \cdots}} dx_{u_{1,1}} \cdots dx_{u_{i,l_i}} \; , \end{aligned} \tag{2.108}$$

and the domain of integration in this second expression can be partitioned into disjoint simplexes. Given a sequence of distinct real numbers $u_{1,1} = v_1, \ldots, u_{i,l_i} = v_S$ let π be the unique rearrangement of $1, \ldots, S$ so that $v_{\pi j}$ are monotone decreasing. More generally, consider the set of all rearrangements Π_l of $1, \ldots, S$ that arise as one reorders sequences $u_{1,1}, \ldots, u_{i,l_1}$ satisfying $s < u_1 < \cdots < u_i < t$, $s < u_{1,1} < \cdots < u_1 = u_{1,l_1}$, etc. until $s < u_{i,1} < \cdots < u_i = u_{i,l_i}$. These are in one to one correspondence with the number of ways to partition $1, \ldots, S$ into exactly i components. The correspondence with $(l_1, \ldots, l_i)$ is achieved by ordering the components according to their last surviving element, (the component that becomes extinct first is the first component etc.) and putting l_j equal to the number of elements in the j-th component. Each element $\pi \in \Pi_l$ induces a linear map of $V^{\otimes(l_1+\cdots+l_i)}$ to itself, and this map π^* is an isometry. Because the domain of integration is the sum of the disjoint simplexes associated with the rearrangements, and the integral is the sum of the integrals over these disjoint domains, we have

$$\iint_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_1} = \sum_{\pi \in \Pi_l} \pi^* \boldsymbol{X}^{S}_{s,t} \; . \tag{2.109}$$

As we will see, this expression is easy to estimate, and we can readily conclude that the expression (2.105) converges absolutely. So for smooth paths we have the identity

$$\boldsymbol{Y}^{i}_{s,t} = \sum_{S=i}^{\infty} \sum_{\substack{l_1+\cdots+l_i=S \\ l_j \geq 1}} A^{l_1} \otimes \cdots \otimes A^{l_i} \sum_{\pi \in \Pi_l} \pi^* \boldsymbol{X}^{S}_{s,t}\, y_s^{\otimes i}\ , \tag{2.110}$$

which has the considerable attraction that the right hand side involves x_t only through it's associated multiplicative functional and is essentially a function on the infinite tensor algebra.

However, *this expression should carry a government health warning.* Certainly, the right hand side is (as we shall see) defined for any multiplicative functional in $\Omega(V)^p$ and is a continuous function on that space. For piecewise smooth paths, it defines a multiplicative functional because it coincides with the iterated integrals of the piecewise smooth path y_t, using the continuity of the map it also defines a multiplicative functional for any element of $\Omega G(V)^p$; indeed that path is geometric.

It is therefore tempting to assume the expression has a natural interpretation for any multiplicative functional in $\Omega(V)^p$, but this is a mistake. The result will not be multiplicative, and so fails the most basic property we expect of iterated integrals, and their substitutes in the rougher case. The point is that the expression on the right in (2.110) is the unique *linear* function yielding the desired value on group-like elements in the tensor algebra. However, although the functions on smooth paths obtained by taking iterated integrals are linearly independent (when regarded as elements of the space of functions on the space of smooth paths), they are certainly not algebraically independent. There are many different algebraic expressions that agree on the sequences of iterated integrals corresponding to geometric multiplicative paths.

Observe that (2.110) defines a multiplicative map from the group-like elements in the tensor algebra of infinite degree into an associative algebra. Arguing formally, we may differentiate to induce a Lie map from the Lie elements of the tensor algebra into the associative algebra. Again arguing formally, the tensor algebra is the enveloping algebra of this embedded Lie algebra, and so exploiting the universal property of enveloping algebras, there should exist a unique multiplicative extension of the Lie map to the full tensor algebra.

If there is a unique extension of (2.110) to a continuous and multiplicative map from $T(W)$ it will not be linear. It's construction would allow us to give a unified treatment of differential equations of Itô and Stratonovich type. We would then be confident that there was good sense in extending the Itô functional beyond geometric paths, and allowing any multiplicative functional in $\Omega(V)^p$ to be the driving signal.

At the time of writing, we believe we understand the correct approach to the identification of such an extension in an analytically useful form. (In the piecewise smooth case each iterated integral of y_t solves a differential equation over x_t, and we may compute the Lie algebra associated to it. In fact this Lie algebra is always finite dimensional. Therefore, after a non-linear change of co-ordinates, we may express the iterated integral as a Taylor series as we have mapped out earlier. By computing these changes of co-ordinates the new expression would be multiplicative for all $\boldsymbol{X}_{s,t}$ in $\Omega(V)^p$); confirmation and explicit determination of the formulae one obtains requires calculations we have not carried through and must wait for a later paper.

Theorem 2.4.1. *The series* (2.110) *and* (2.105) *converge absolutely for any multiplicative functional* $\boldsymbol{X}_{s,t}$ *in* $\Omega(V)^p$ *and define continuous functions. The resulting sequence* $\mathbf{Y}_{s,t} = \{\mathbf{Y}^i_{s,t}\}_{i=0}^n$ *is of finite p-variation and*

$$\|\mathbf{Y}^i_{s,t}\| \le K^i \frac{i^i}{i!} \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!} \sum_{S=0}^{\infty} K^S\, i^S \frac{\omega(s,t)^{S/p}}{(S/p)!} \|y_s\|^i \,. \tag{2.111}$$

If $\boldsymbol{X}_{s,t}$ *in* $\Omega G(V)^p$ *is multiplicative, then* $\mathbf{Y}_{s,t} = \{\mathbf{Y}^i_{s,t}\}_{i=0}^n$ *is multiplicative, and we have the asymptotically improved bound*

$$\|\mathbf{Y}^i_{s,t}\| \le \frac{(U_p\,\omega(s,t))^{i/p}}{\beta\,(i/p)!} \,. \tag{2.112}$$

PROOF. Let $K = \|A\|$ be the operator norm of A regarded as a linear map $A : V \longrightarrow \hom(W,W)$. The number of partitionings of an ordered set of S elements into exactly i non-empty subsets is bounded above by

$$\frac{i^S}{i!} \tag{2.113}$$

and so

$$\|\mathbf{Y}^i_{s,t}\| \le \sum_{S=i}^{\infty} K^S \frac{i^S}{i!} \|\boldsymbol{X}^S_{s,t}\| \, \|y_s\|^i \,. \tag{2.114}$$

If $\boldsymbol{X}_{s,t}$ is in $\Omega(V)^p$ and has variation controlled by ω then one has the estimate

$$\begin{aligned} \|\boldsymbol{Y}^i_{s,t}\| &\le \sum_{S=i}^{\infty} K^S \frac{i^S}{i!} \frac{\omega(s,t)^{S/p}}{\beta\,(S/p)!} \|y_s\|^i \\ &= K^i \frac{i^i}{i!} \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!} \sum_{S=0}^{\infty} K^S\, i^S \frac{\omega(s,t)^{S/p}\,(i/p)!}{((S+i)/p)!} \|y_s\|^i \\ &\le K^i \frac{i^i}{i!} \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!} \sum_{S=0}^{\infty} K^S\, i^S \frac{\omega(s,t)^{S/p}}{(S/p)!} \|y_s\|^i , \end{aligned} \tag{2.115}$$

showing the series converges absolutely and bounding the individual terms in a way that makes it clear that $\boldsymbol{Y}^{(n)}_{s,t}$ has finite p-variation controlled by a multiple of ω on any interval where ω is bounded. A virtually identical argument shows the uniform continuity of the sequence under variation of $\boldsymbol{X}_{s,t}$. However, the constants in these estimates explode with the degree.

On the positive side, the continuity ensures that if $\boldsymbol{X}_{s,t}$ is in $\Omega G(V)^p$ then $\boldsymbol{Y}^{(n)}_{s,t}$ is multiplicative; our results in Section 2.2 and particularly Theorem 2.2.1 then give the much stronger and more useful estimate that for $n > p$, ω bounded by L, and with

$$U_p = \max_{j \le [p]} \|y_s\|^j\, K^p \frac{j^p}{(j!)^{p/j}} \Big(\sum_{S=0}^{\infty} \frac{K^S\, j^S\, L^{S/p}}{(S/p)!} \Big)^{p/j} , \tag{2.116}$$

choosing β large enough, we have

$$\|\boldsymbol{Y}^i_{s,t}\| \le \frac{(U_p\,\omega(s,t))^{i/p}}{\beta\,(i/p)!} \tag{2.117}$$

completing the proof of the theorem.

Cross terms. We have therefore seen that for linear equations the Itô functional can be extended in a unique continuous way as a map from $\Omega G(V)^p$ to $\Omega G(W)^p$. However, for technical reasons that will become apparent later, we would like also to know that the iterated integrals between solution and driving noise also exist. This is readily done by extending the original differential equation, in other words we solve the equation

$$\begin{cases} dc_t = 0\, c_t\, dx_t\ , \\ d\hat{x}_t = c_t\, dx_t\ , \\ dy_t = A(y)\, dx_t\ , \end{cases} \tag{2.118}$$

with $c_0 = 0$, $\hat{x}_0 = x_0$. The equation is still linear and so we can use the approach above to construct the iterated integrals of $\hat{x}$ and y and see that they have unique continuous extension to $\Omega G(V)^p$.

In this way we see that if we wish to record the full structure associated to our differential equation *we should regard the Itô map as an extension map lifting paths in* $\Omega G(V)^p$ *to paths in* $\Omega G(V \oplus W)^p$*.*

2.4.2. The stochastic example.

What do the results we have proved so far say in the context of Brownian motion and stochastic differential equations?

Suppose that $X_t \in V$ is a continuous path in Euclidean space, chosen randomly according to Wiener measure (in which case we say it is a Brownian path) or more generally according to some measure which makes the underlying stochastic process a martingale or semimartingale (when we say X_t is a martingale or semimartingale path). Then it is standard [11] that, with probability one, the forward and symmetric Riemann sums

$$
\begin{aligned}
\boldsymbol{X}_{s,t}^{2,\text{ito}} &= \lim_{n\to\infty} \sum_{s<k/2^n}^{k/2^n<t} X_{k/2^n} \otimes \left(X_{k/2^n} - X_{(k+1)/2^n}\right), \\
\boldsymbol{X}_{s,t}^{2,\text{strat}} &= \lim_{n\to\infty} \sum_{s<k/2^n}^{k/2^n<t} \frac{X_{k/2^n} + X_{(k+1)/2^n}}{2} \otimes \left(X_{k/2^n} - X_{(k+1)/2^n}\right),
\end{aligned}
\tag{2.119}
$$

converge uniformly in the time co-ordinates and define two distinct multiplicative functionals

$$
\begin{aligned}
\boldsymbol{X}_{s,t}^{\text{ito}} &= (1, \boldsymbol{X}_s - \boldsymbol{X}_t, \boldsymbol{X}_{s,t}^{2,\text{ito}}), \\
\boldsymbol{X}_{s,t}^{\text{strat}} &= (1, \boldsymbol{X}_s - \boldsymbol{X}_t, \boldsymbol{X}_{s,t}^{2,\text{strat}}),
\end{aligned}
\tag{2.120}
$$

corresponding to the Itô and Stratonovich integrals. A simple Borel-Cantelli lemma shows that with probability one they are both in $\Omega(V)^p$ for every $p > 2$. The two multiplicative functionals agree in degree one, so their difference is an additive function with values in two tensors. It is referred to by probabilists as the *quadratic variation process*

$$\langle X, X\rangle_{s,t} = \frac{1}{2}\lim_{n\to\infty}\sum_{s<k/2^n}^{k/2^n<t}\left(X_{k/2^n} - X_{(k+1)/2^n}\right) \otimes \left(X_{k/2^n} - X_{(k+1)/2^n}\right), \tag{2.121}$$

it has finite variation with probability one. Exploiting the Itô and Stratonovich integrals further, one may construct higher order iterated integrals. These sequences $\boldsymbol{X}_{s,t}^{\text{ito}}$ and $\boldsymbol{X}_{s,t}^{\text{strat}}$ define multiplicative functionals of finite p-variation and arbitrarily high degree.

By our theorems these higher iterated integrals etc. are continuous functions of the path and its second iterated integral. *The difference between the Itô and Stratonovich equations driven by Brownian motion depends entirely on the choice of multiplicative functional of degree two that we use to extend Brownian motion.*

To understand clearly the possibilities and choices made in extending our Brownian path to a multiplicative functional of degree two and finite p-variation where $2 < p < 3$, we must look more carefully at the symmetric and anti-symmetric components of $\boldsymbol{X}_{st}^2$.

Decomposing the second integral - the area or anti-symmetric part. In our discussion of the iterated integrals of a smooth path, we saw that the symmetric part of the classical second iterated integral of a smooth path is

$$\frac{1}{2}\left(X_t - X_s\right)\otimes\left(X_t - X_s\right) \tag{2.122}$$

and as this is a continuous function in the uniform topology this relation will hold true for any geometric path. (One readily checks that for the Stratonovich integral the symmetric component of the second integral is precisely this continuous extension.)

To create a geometric multiplicative functional of degree two it is therefore sufficient to construct the anti-symmetric two tensor process, and to be multiplicative this must satisfy the algebraic relationship

$$\boldsymbol{A}_{s,u} = \boldsymbol{A}_{s,t} + \boldsymbol{A}_{t,u} + \text{Area}\left(\overline{X_s\, X_t\, X_u}\right), \tag{2.123}$$

where

$$\text{Area}\left(\overline{PQR}\right), \tag{2.124}$$

is the area of the triangle interpolating the three points P, Q, R. (Observe that the *area* associated to a loop formed by taking a chord and the trajectory of the path along the time interval $[s,t]$ would obviously satisfy this relationship).

Definition 2.4.1. *We call an anti-symmetric two tensor process satisfying* (2.123) *an area process relative to the path* X_s.

Suppose $2 < p < 3$, then for any path of finite p-variation in V, and associated area process $\boldsymbol{A}_{s,t}$ (having the correct modulus of continuity) the multiplicative functional

$$\left(1, \boldsymbol{X}_t - \boldsymbol{X}_t, \frac{1}{2}\left(\boldsymbol{X}_t - \boldsymbol{X}_s\right) \otimes \left(\boldsymbol{X}_t - \boldsymbol{X}_s\right) + \boldsymbol{A}_{s,t}\right) \tag{2.125}$$

defines a *geometric* multiplicative functional in $\Omega G(V)^2$. The geometric condition does not imply any sort of uniqueness or canonical choice for the the area process given the underlying path, this is in contrast to the unique continuous choice for symmetric component. Even if X_t is smooth, there are many elements of $\Omega G(V)^2$ lying over the path. Consider the multiplicative functional $\boldsymbol{Y}_{s,t} = (1, 0, \psi(t) - \psi(s))$ constructed by taking the limit of the increments and second integrals of the smooth paths $\exp\left(n^2 \pi\, i\, \psi(t)\right)/(n\pi)$. The result is geometric, non-trivial, and for smooth enough ψ will be in $\Omega G(\mathbb{R}^2)^2$, however it projects to the constant path.

The key, then, to defining stochastic differential equations is the choice of this area integral. It really is a choice even in the Brownian case, the work [25] demonstrates just how tenuous the connection between Lévy area and geometric area of smooth paths really is.

The Itô and Stratonovich second iterated integrals only differ in the symmetric bracket process, they share a common area process - the *Lévy Area.* The Stratonovich multiplicative functional is geometric.

Theorem 2.4.2. *Let* X_t *be a semi-martingale and* $\boldsymbol{A}_{s,t}$ *be its Lévy area. The linear stochastic differential equation*

$$dy_t = A(y_t, dX_t) + B(y_t)\, dt\,, \tag{2.126}$$

where $x \longrightarrow A(\ ,x)$ *in* $\operatorname{hom}(V, \operatorname{hom}(W,W))$, *and* $B(\,)$ *in* $\operatorname{hom}(W,W)$, *are bounded operators which can be regarded as the composition of a continuous function on* $\Omega(V,\mathbb{R})^{2+\varepsilon,1}$ *and the random multiplicative functional*

$$\left(1, \boldsymbol{X}_t - \boldsymbol{X}_s, \frac{1}{2}\left(\boldsymbol{X}_t - \boldsymbol{X}_s\right) \otimes \left(\boldsymbol{X}_t - \boldsymbol{X}_s\right) + \boldsymbol{A}_{s,t}\right). \tag{2.127}$$

In particular, all equations can be solved simultaneously with only a single null set. The equations can be chosen to depend on the path, end point of the solution etc.

PROOF. There is little to say. The driving signal is (X_t, t), so that if we consider the inhomogeneous p-variation introduced in (2.3.1), (The cross-iterated integrals against t are all canonically defined) we deduce that the differential equation can be extended from the class of smooth paths in an unique way to $\Omega G(V, \mathbb{R})^{2+\varepsilon,1}$. The multiplicative functional (2.127), with probability one, takes its values in $\Omega G(V, \mathbb{R})^{2+\varepsilon,1}$ [23]. We claim that this construction obtained by taking the composition of the two maps coincides with the Stratonovich solution which probabilists construct.

Fortunately, the very continuity of the map from $\Omega G(V, \mathbb{R})^{2+\varepsilon,1}$ ensures this. It is well known that one may solve a Stratonovich differential equation in probability, by replacing the semimartingale path by its dyadic piecewise linear approximations, and then taking the solutions to the equation driven by these piecewise linear equations [11].

On the other hand, our definition of the Lévy area makes it clear that it is the limit of the areas associated to these piecewise linear paths, a Borel-Cantelli argument ([8], Sipiläinen) shows that the rate of convergence is fast enough for the piecewise linear paths, and their iterated integrals to converge in $\Omega G(V, \mathbb{R})^{2+\varepsilon,1}$. By our continuity results, we see that our solution and the conventional probabilistic one agree with probability one.

Finally observe that our solution is obtained by composing a deterministic function depending on the coefficients of the equation with a random multiplicative functional constructed almost surely, but with a null set that is independent of the coefficients of the equation. In particular we may solve all such equations simultaneously and can choose the equation so as to depend on the path without difficulty of interpretation. No predictability condition is involved.

REMARKS 2.4.1. GENERALIZING THE EQUATION. We will in due course prove that we can develop continuity results in the fully nonlinear situation where the vector fields in the differential equation are $\mathrm{Lip}\,(2+\varepsilon, V)$ so the remarks above apply in much greater generality than the linear case proven so far.

REMARKS 2.4.2. GENERALIZING THE NOISE. There are a number of directions in which one could generalize the noise. One should certainly consider jumps; in general these are a little easier, because pure jump random processes tend to have finite variation for $p < 2$. The area integral does not come into the picture [31]. In another direction, one could look at other Markov processes as driving processes for dynamical systems. Here, matters still seem relatively open, except that one can say there are wide classes of Markov processes which extend, like Brownian motion, to admit Lévy area processes, and hence Stratonovich differential equations; but which are definitely not semi-martingales and cannot be attacked via the standard Itô theory.

In these situations where the usual theory simply does not apply [10] an alternative approach is required to construct the Lévy area. Now, Lévy proved, if one takes the piecewise linear approximation to the path X_s that agrees at 2^n equally spaced points and look at the sequence of areas as one refines the dyadic partitions. Then if X_s is Brownian motion, *this sequence forms a martingale* over the filtration obtained by revealing X_s at the 2^n equally spaced points. In many other situations, one can still show that it is a convergent semi-martingale. The classical martingale techniques are still important - but not the time ordered filtration.

The symmetric part of $\mathbf{X}^2_{s,t}$ and Itô Equations. By now a persistent reader might understand enough to guess that constructing different second order multiplicative extensions to Brownian motion is essentially equivalent to varying our notion of solution to our stochastic differential equation. Even so we are at least superficially surprised that the distinction between Itô and Stratonovich second integrals is not in the discontinuous Lévy area, but in the symmetric part; the part which has a natural continuous choice for all continuous paths!

The difference between the Itô and Stratonovich approaches lie in the symmetric additive functional known as the quadratic variation or bracket process.

The earlier results about iterated integrals apply to the Itô equation and it is easy to write down series solutions etc. but now those series involve the bracket process. To apply our approach we express our equations in co-ordinate invariant form. An Itô equation

$$dy_t = f^i(y_t)\,dx^i_t + f^0(y_t)\,dt\,, \tag{2.128}$$

always requires a connection before it makes good sense, but can then be rewritten in the Stratonovich form

$$dy_t = f^i(y_t)\,dx^i_t + f^0(y_t)\,dt + \nabla_{f^i} f^j\, d\langle X, X\rangle^{i,j}_{0,t} \tag{2.129}$$

and one can deduce all the theorems one had before, but now the dependence includes the bracket process separately.

Theorem 2.4.3. *Consider the linear Itô stochastic differential equation* (2.126)

$$dy_t = A(y_t, dx_t) + B(y_t)\,dt\,, \tag{2.130}$$

where $x \longrightarrow A(\ ,x)$ *in* $\hom(V, \hom(W,W))$*, and* $B(\ )$ *in* $\hom(W,W)$*, are bounded operators. This map can be regarded as the composition of a continuous function on* $\Omega(V, V\hat{\otimes}V, \mathbb{R})^{2+\varepsilon,1,1}$ *and a random multiplicative functional depending only on the path, its Lévy area, and its bracket process. In particular, all equations can be solved simultaneously with only a single null set. The equations can be chosen to depend on the path, end point of the solution etc.*

In particular, this perspective suggests that for robust numerical solution of stochastic differential equations, one should not try to implicitly simulate the bracket process locally as the quadratic variation of the path, (as one does when one solves an Itô equation directly using Euler type methods) but treat it separately as a known quantity and go via Stratonovich methods. We think this is definitely true in some cases, although it is not the whole story, and a complete understanding of differential equations driven by (non-geometric) multiplicative functionals will be required to give a better answer.

3. Integration against a rough path.

In this section we move from the linear/real analytic setting to the truly non-linear/rough setting. Our objective is to define the integral of a rough path against a one form.

3.1. Almost multiplicative functionals - The construction of an integral.

We have shown in Section 2.2 that if $\boldsymbol{X}_{st} \in T^{(n)}$ is p-multiplicative (where we will use the convention $n = [p]$) then it extends in a unique way to a multiplicative functional $\boldsymbol{X}_{st}$ of finite p -variation in $T^{(m)}$ for all $m \geq n$ and if $\boldsymbol{X}_{st}$ is controlled by ω in $T^{(n)}$ so that

$$\|\boldsymbol{X}^j_{st}\| \leq \frac{\omega(s,t)^{j/p}}{\beta\,(j/p)!}\,, \qquad \text{for all } j \leq n\,, \tag{3.1}$$

then one has the same estimate for all $j < \infty$. (Here β is an appropriately chosen constant depending only on p). Similar estimates reflect the continuity of this extension map.

We will now explain how, with some loss of quantitative control, this result can be seen as a special case of a more general one concerning almost multiplicative functionals.

Definition 3.1.1. *Suppose $\boldsymbol{X}_{st}$ is any functional taking values in $T^{(n)}$, we say it is of finite p-variation controlled by ω if, for all s,t,*

$$\|\boldsymbol{X}^j_{st}\| \leq \frac{\omega(s,t)^{j/p}}{\beta\,(j/p)!}\,, \qquad \textit{for all } j \leq n\,, \tag{3.2}$$

In addition we say that such an $\boldsymbol{X}_{st}$ is an almost multiplicative functional if for any compact interval J there is a θ and a K such that for all s,t and u in J we have

$$\|(\boldsymbol{X}_{st}\,\boldsymbol{X}_{tu} - \boldsymbol{X}_{su})^j\| \leq K\,\omega(s,u)^{\theta}\,, \qquad \textit{for all } j \leq n,\ \theta > 1\,. \tag{3.3}$$

OBSERVATIONS 3.1.1. We have already seen an almost multiplicative functional. The lift $\hat{\boldsymbol{X}}_{st} = (1, \boldsymbol{X}^1_{st}, \ldots, \boldsymbol{X}^n_{st}, \mathbf{0})$ defined in the proof of Theorem 2.2.1 is an almost multiplicative functional controlled by ω providing $\boldsymbol{X}_{s,t} = (1, \ldots, \boldsymbol{X}^n_{st})$ is a multiplicative functional of finite p-variation where $n \geq [p]$. We see therefore that (ignoring the quality of the estimates) Theorem 2.2.1 is a special case of the following.

Theorem 3.3.1. *Suppose $\boldsymbol{X}_{st}$ is a bounded almost multiplicative functional controlled by ω on the compact interval J of degree n. Then there exists a unique multiplicative functional $\hat{\boldsymbol{X}}_{st}$ on J and a constant*

$$C\,(L, K, \theta, \max_{s,t\in J} \omega(s,t), n)\,,$$

such that

$$\|(\hat{\boldsymbol{X}}_{st} - \boldsymbol{X}_{st})^i\| \leq C\,(L, K, \theta, \max_{s,t\in J} \omega(s,t), n)\,\omega(s,t)^{\theta}\,, \tag{3.4}$$

for all $i \leq n$. *There is at most one multiplicative functional* $\hat{\boldsymbol{X}}_{st}$ *that can satisfy* (3.4) *regardless of the choice of* C. *Here* θ, K *and* ω *are the terms in the definition of almost multiplicative and* L *is the uniform bound on the components of* $\boldsymbol{X}_{st}$.

Corollary 3.1.1. *In addition, if* $\boldsymbol{X}_{st}$ *has finite p-variation controlled by* ω *then* $\hat{X}$ *has p-variation controlled by* $C_1\,\omega$ *where* C_1 *only depends on* K, θ, $\max\{\omega(s,t),\ s,t \in J\}$ *and* n.

Proof of the theorem. We proceed by induction and suppose the projection of $\boldsymbol{X}_{st}$ into $T^{(j)}$ has the multiplicative property. Presuming for a moment existence of the limit, define $\tilde{X}$ as follows

$$(\tilde{\boldsymbol{X}}_{st})^{j+1} = \lim_{\mathrm{mesh}(D)\to 0} (\boldsymbol{X}_{st_1}\boldsymbol{X}_{t_1t_2}\cdots\boldsymbol{X}_{t_{r-1}t})^{j+1} \tag{3.5}$$

and for all $i \neq j+1$ take $(\tilde{\boldsymbol{X}}_{st})^i = (\boldsymbol{X}_{st})^i$. In this case it is clear that $\tilde{X}$ will be multiplicative on $T^{(j+1)}$. If we show the existence of $(\tilde{\boldsymbol{X}}_{st})^{j+1}$, establish that $\tilde{X}$ is almost multiplicative, and compare it with $\boldsymbol{X}_{st}$ we will have established the induction step. Iterating it completes the proof.

We proceed in a similar way to before. Let

$$\boldsymbol{X}^D_{s,t} = \boldsymbol{X}_{s,t_1}\boldsymbol{X}_{t_1,t_2}\cdots\boldsymbol{X}_{t_{r-1},t}\ ,$$

where $D = \{s, t_1, \ldots, t_{r-1}, t\}$ is a dissection of $[s,t]$. First we bound

$$(\boldsymbol{X}^D_{st} - \boldsymbol{X}_{st})^{j+1}\,, \tag{3.6}$$

independently of the choice of dissection D, and then we will show the convergence of the products as the mesh size of the dissections tends to zero, always providing ω is regular. Observe first that in the case where the dissection is trivial, $r = 2$, the difference in (3.6) is zero. Assume the dissection is nontrivial, and choose an interior point t_i of the dissection D so that

$$\omega(t_{i-1}, t_{i+1}) \leq \frac{2}{(r-2)}\,\omega(s,t)$$

or equals $\omega(s,t)$ in the case where $r = 3$. Let $D' = D - \{t_i\}$. If we estimate $((\boldsymbol{X}^D_{st} - \boldsymbol{X}^{D'}_{st}))^{j+1}$ and the similar terms as we successively remove all the interior points of the dissection, we may use the triangle inequality to estimate (3.6); we will obtain a bound, which in analogy with our previous arguments, is easily seen to be dissection independent. Now

$$((\boldsymbol{X}_{st}^{D} - \boldsymbol{X}_{st}^{D'}))^{j+1} \tag{3.7}$$

$$= (\boldsymbol{X}_{s,t_{i-1}}^{D\cap[s,t_{i-1}]}(\boldsymbol{X}_{t_{i-1},t_i}\boldsymbol{X}_{t_i,t_{i+1}} - \boldsymbol{X}_{t_{i-1},t_{i+1}})\boldsymbol{X}_{t_{i+1},t}^{D\cap[t_{i+1},t]})^{j+1} \tag{3.8}$$

and the multiplicative nature of $\boldsymbol{X}$ ensures that

$$\boldsymbol{X}_{t_{i-1}t_i}\boldsymbol{X}_{t_it_{i+1}} - \boldsymbol{X}_{t_it_{i+1}} = (\underbrace{0,\dots,0}_{j+1 \text{ terms}}, \boldsymbol{R}_{t_{i-1},t_i,t_{i+1}}^{j+1},\dots) \tag{3.9}$$

and so

$$((\boldsymbol{X}_{st}^{D} - \boldsymbol{X}_{st}^{D'}))^{j+1} = \boldsymbol{R}_{t_{i-1},t_i,t_{i+1}}^{j+1}\,. \tag{3.10}$$

But the almost multiplicative property then gives the estimate

$$\|\boldsymbol{R}_{t_{i-1},t_i,t_{i+1}}^{j+1}\| \le K\,\omega(t_{i-1},t_{i+1})^{\theta} \le K\Big(\frac{2}{r-2}\Big)^{\theta}\,\omega(s,t)^{\theta}\,, \tag{3.11}$$

for $r > 3$ and the similar estimate for $r = 3$. Summing these error estimates as one drops points from the dissection leads to the, by now, familiar estimate

$$\|(\boldsymbol{X}_{s,t}^{D} - \boldsymbol{X}_{s,t})^{j+1}\| \le K\,(2^{\theta}(\zeta(\theta)-1)+1)\,\omega(s,t)^{\theta}\,, \tag{3.12}$$

and the consequential argument that if ω is regular, then the X^D converge as the mesh size of the dissection goes to zero. In particular we may define

$$(\tilde{\boldsymbol{X}}_{s,t})^{j+1} = \lim_{\mathrm{mesh}(D)\to 0}(\boldsymbol{X}_{st}^{D})^{j+1}\,.$$

It follows that if

$$\boldsymbol{R}_{st}^{j+1} = (\tilde{\boldsymbol{X}}_{st} - \boldsymbol{X}_{st})^{j+1}\,, \tag{3.13}$$

then

$$\|\boldsymbol{R}_{st}^{j+1}\| \le K\,(2^{\theta}(\zeta(\theta)-1)+1)\,\omega(s,t)^{\theta}\,. \tag{3.14}$$

To see that $\tilde{\boldsymbol{X}}$ is almost multiplicative, observe that

$$(3.15) \qquad (\tilde{\boldsymbol{X}}_{st}\,\tilde{\boldsymbol{X}}_{tu} - \tilde{\boldsymbol{X}}_{su})^i = \begin{cases} 0\,, & i \le j+1\,, \\ (\boldsymbol{X}_{st}\boldsymbol{X}_{tu} - \boldsymbol{X}_{su})^i & \\ +\boldsymbol{R}_{st}^{j+1}\,\boldsymbol{X}_{tu}^{i-(j+1)} & \\ +\boldsymbol{X}_{st}^{i-(j+1)}\,\boldsymbol{R}_{tu}^{j+1}\,, & i > j+1\,. \end{cases}$$

and providing $\|\boldsymbol{X}_{st}^i\| \le L$ for all $i \le n$, $s,t \in J$ we have the estimate

$$(3.16) \qquad \begin{aligned} \|(\tilde{\boldsymbol{X}}_{st}\tilde{\boldsymbol{X}}_{tu} - \tilde{\boldsymbol{X}}_{su})^i\| \le K\,\omega(s,u)^\theta & \\ + 2\,L\,K(2^\theta(\varsigma(\theta)-1)+1)\,\omega(s,u)^\theta\,, & \end{aligned}$$

completing the proof that $\tilde{\boldsymbol{X}}$ is almost multiplicative, but with the new constant

$$(3.17) \qquad \tilde{K} \le K\,(1 + 2\,L\,(2^\theta(\xi(\theta)-1)+1))\,,$$

and a new uniform bound

$$(3.18) \qquad \tilde{L} \le L + K\,(2^\theta(\zeta(\theta)-1)+1)\,\max_{s,t\in J}\omega(s,t)^\theta\,.$$

As $\tilde{\boldsymbol{X}}$ is also almost multiplicative controlled by a multiple of ω and bounded on J this completes the basic induction step. Observing that the theorem is trivial if $j = 0$ and repeating the step n times completes the construction of the multiplicative functional.

To see uniqueness of the functional, it is enough to show that if one has two multiplicative functionals $\widehat{\boldsymbol{X}}_{st}$, $\tilde{\tilde{\boldsymbol{X}}}_{st}$ and they satisfy

$$(3.19) \qquad \|(\widehat{\boldsymbol{X}}_{st} - \tilde{\tilde{\boldsymbol{X}}}_{st})^i\| \le C\,\omega(s,t)^\theta\,, \qquad \text{for all } s,t \in J,\ i \le n\,,$$

then they agree for all $i \le n$. The proof is also an induction argument, fix the smallest i for which the two multiplicative functionals differ. Then putting

$$\psi(s,t) = (\widehat{\boldsymbol{X}}_{st} - \tilde{\tilde{\boldsymbol{X}}}_{st})^i\,,$$

one obtains from the multiplicative property that

$$\psi(s,u) = \psi(s,t) + \psi(t,u)$$

and hence for any dissection one has the estimate that

$$\psi(s,t) \leq (\max_D \omega(t_i, t_{i+1}))^{\theta-1}\, \omega(s,t)\,.$$

For regular ω on an interval J this forces $\psi(s,t) \equiv 0$ contradicting the induction hypothesis.

The theorem and its proof only require a boundedness assumption on $\boldsymbol{X}$ and regularity assumption on ω.

Proof of the Corollary. Suppose now that $\boldsymbol{X}$ is of finite p-variation controlled by ω on $T^{(n)}$, where $n/p \leq \theta$. Then it is a simple application of the triang le inequality to see that $\hat{\boldsymbol{X}}$ is also of finite p-variation on $T^{(n)}$. If $n > p$ then one may repeat the uniqueness induction argument we have just given to deduce that the new multiplicative functional we have constructed in this theorem agrees with the unique multiplicative extension of finite p-variation we constructed in Theorem 2.2.1.

3.3.1. Applications and extensions.

A) The map from p-almost multiplicative functional to p-multiplicative functional is a uniformly continuous one. However, this is not a consequence of the result so much as of the proof. Suppose that $\boldsymbol{X}$, $\boldsymbol{Y}$ are two almost multiplicative functionals controlled by the same K, ω, θ. And suppose that they are close to each other in the sense that

$$\|(\boldsymbol{X}_{s,t} - \boldsymbol{Y}_{s,t})^i\| < \varepsilon\, \omega(s,t)^{i/p}\,, \qquad i \leq [p]\,, \tag{3.20}$$

then of course by the triangle inequality

$$\|(\hat{\boldsymbol{X}}_{s,t} - \hat{\boldsymbol{Y}}_{s,t})^i\| < \varepsilon\, \omega(s,t)^{i/p} + C\,\omega(s,t)^{\theta} \tag{3.21}$$

and for $i \leq [p]$ this looks adequate. But for $\varepsilon < C\,\omega(s,t)^{\theta-[p]/p}$ or less seriously $\omega(s,t) \gg 1$ the estimate deteriorates. The key to the proof of a continuity result is to observe that at each stage in the construction of the multiplicative functional out of the almost multiplicative functional, we can control the difference between the two approximations. We then obtain the following theorem.

Theorem 3.1.2. *Suppose that $\boldsymbol{X}$, $\boldsymbol{Y}$ are two almost multiplicative functionals controlled by the same K, ω, θ, and that $\omega(s,t) < L$ for $s,t \in J$. Suppose further that $\boldsymbol{X}, \boldsymbol{Y}$ are close in the p-variation sense so that*

$$\|(\boldsymbol{X}_{s,t} - \boldsymbol{Y}_{s,t})^i\| < \varepsilon\, \omega(s,t)^{i/p}\,, \qquad i \leq [p]\,, \tag{3.22}$$

then there is a continuous, increasing function $\delta(\varepsilon)$ depending only on K, L, θ, p and satisfying $\delta(0) = 0$ so that the associated multiplicative functionals satisfy

$$\|(\hat{\boldsymbol{X}}_{s,t} - \hat{\boldsymbol{Y}}_{s,t})^i\| < \delta(\varepsilon)\, \frac{\omega(s,t)^{i/p}}{(i/p)!}\,, \tag{3.23}$$

for all i.

PROOF. Because of Theorem 3.1.1, it is sufficient that we deal with the case $i \leq [p]$.

Suppose $\boldsymbol{X}_{st}$, $\boldsymbol{Y}_{st}$ are almost multiplicative and multiplicative up to degree $j < [p]$; and that they satisfy the hypotheses of the theorem. Define $\tilde{\boldsymbol{X}}$, $\tilde{\boldsymbol{Y}}$ by

$$(\tilde{\boldsymbol{X}}_{st})^{j+1} = \lim_{\mathrm{mesh}(D)\to 0} (\boldsymbol{X}_{st_1}\boldsymbol{X}_{t_1t_2}\cdots\boldsymbol{X}_{t_{r-1}t})^{j+1} \tag{3.24}$$

and for all $i \neq j+1$ take $(\tilde{\boldsymbol{X}}_{st})^i = (\boldsymbol{X}_{st})^i$, and similarly for $\tilde{\boldsymbol{Y}}$. We will show that these are close in the sense of the conclusion. Repeating the argument the required finite number of times, the result will follow.

Define $\boldsymbol{X}^D_{s,t} = \boldsymbol{X}_{s,t_1}\boldsymbol{X}_{t_1,t_2}\cdots\boldsymbol{X}_{t_{r-1},t}$, etc. where

$$D = \{s, t_1, \ldots, t_{r-1}, t\}\,.$$

We will estimate $\|(\tilde{\boldsymbol{X}}_{s,t} - \tilde{\boldsymbol{Y}}_{s,t})^{j+1}\|$ by controlling $(\boldsymbol{X}^D_{st} - \boldsymbol{Y}^D_{st})^{j+1}$ in a uniform way and passing to the limit. Now as before we may successively drop points from the dissection. By making a careful choice of the point to drop (but note that the choice depends on ω alone, and can be common for both functionals) we have the following two estimates; because of the almost multiplicativeness we have

$$\|(\boldsymbol{X}^D_{st} - \boldsymbol{X}^{D'}_{st})^{j+1}\| \leq K\left(\frac{2}{r-2}\right)^{\theta} \omega(s,t)^{\theta}\,. \tag{3.25}$$

for $r > 3$ and the similar estimate for $r = 3$. Combining the estimates for $\boldsymbol{Y}$ we have

$$\|((\boldsymbol{X}_{st}^{D} - \boldsymbol{Y}_{st}^{D}) - (\boldsymbol{X}_{st}^{D'} - \boldsymbol{Y}_{st}^{D'}))^{j+1}\| \leq 2\,K \left(\frac{2}{r-2}\right)^{\theta} \omega(s,t)^{\theta}\,, \tag{3.26}$$

while using the closeness hypothesis, (and some crude version of the neo-classical inequality) one obtains

$$\begin{aligned}&\|((\boldsymbol{X}_{st}^{D} - \boldsymbol{Y}_{st}^{D}) - (\boldsymbol{X}_{st}^{D'} - \boldsymbol{Y}_{st}^{D'}))^{j+1}\| \\ &\qquad \leq A(p)\,(\varepsilon + \varepsilon^2)\left(\frac{1}{r-2}\right)^{(j+1)/p} \omega(s,t)^{(j+1)/p}\,,\end{aligned} \tag{3.27}$$

so combining the two and using the uniform bound that $\omega(s,t) < L$ one has

$$\begin{aligned}&\|((\boldsymbol{X}_{st}^{D} - \boldsymbol{Y}_{st}^{D}) - (\boldsymbol{X}_{st}^{D'} - \boldsymbol{Y}_{st}^{D'}))^{j+1}\| \\ &\qquad \leq \Bigg(A(p)\,(\varepsilon + \varepsilon^2)\left(\frac{1}{r-2}\right)^{(j+1)/p} \\ &\qquad\qquad \wedge\, 2\,K\,L^{\theta-(j+1)/p}\left(\frac{2}{r-2}\right)^{\theta}\Bigg)\,\omega(s,t)^{(j+1)/p}\end{aligned} \tag{3.28}$$

and summing this over r yields the required uniform estimate.

B) As a second simple, but rather important corollary of Theorem 3.1.1, we see that it is possible to vary one multiplicative functional in the direction of a second. In particular, suppose that $\boldsymbol{X}_{st}$ is a multiplicative functional of finite p-variation controlled by ω and that $\boldsymbol{H}_{st}$ is a second; and suppose further that $\|(\boldsymbol{H}_{st})^j\| < K(\omega(s,t))^{\phi}$ for all $j \leq [p]$ and for some $\phi > 1 - 1/p$. In this case, the neo-classical inequality shows $\boldsymbol{H}_{st}\boldsymbol{X}_{st}$ to be of finite p-variation an d more elementary considerations show it to be almost multiplicative. Moreover $\|(\boldsymbol{H}_{st}\boldsymbol{X}_{st} - \boldsymbol{X}_{st}\boldsymbol{H}_{st})^j\| < K(\omega(s,t))^{\phi+1/p}$, and so Theorem 3.1.1 shows that the multiplicative functional associated to the left or right hand perturbations of $\boldsymbol{X}_{st}$ coincide. We denote this modification by $\boldsymbol{X}_{s,t}^{H}$. Although we do not have time in this paper to pursue the matter, it will be useful if we want to differentiate functionals on path space.

3.2. Integrating a one-form - A most important almost multiplicative functional.

Our intention is to solve equations of the type

$$dY = f(Y)\, dX\,, \qquad Y_0 = a\,, \tag{3.29}$$

where X and Y are multiplicative functionals and where $Y_s = Y_{0s} + Y_0$. We wish to adopt an approach based on Picard iteration, in other words we treat our equation as an integral equation and construct a solution by iterating the function F

$$F(Y)_t = a + \int_0^t f(Y_s)\, dX_s\,. \tag{3.30}$$

Although such an approach is almost universal, it is apparently unnatural from a geometric perspective. Every term in our differential equation is meaningful without a choice of co-ordinates for the space where Y takes its values and one would hope that the solution had the same properties. However, the functional in (3.30) certainly involves a choice of co-ordinate chart, and different choices produce different maps F.

To succeed in our Picard iteration we now follow up these two separate but closely related points. We must make sense of the concept of an integral, and we must understand its behaviour under changes of variable.

3.2.1. Integrating a one form.

We will now prove that a one form can be integrated against a multiplicative functional in a natural way. We do this via the construction of an almost multiplicative functional. The reader should be warned that our methods are currently limited, and in general we can only treat *geometric* multiplicative functionals of finite p-variation. However, for the case where the paths have p-variation satisfying $p < 3$ (and so degree is $n \leq 2$) then the next section will extend these results to all multiplicative functionals. This improvement in the case $n = 2$ is important because it allows one to treat the Itô approach to differential equations in common with the Stratonovich approach. We believe our failure to extend the result to all multiplicative functionals in the general case reflects a lack of understanding on our part, inspection and guess work allow one to treat $n = 2$ but do not point to the general picture.

But before explaining the analysis, for the sake of precision, we need some simple notation.

Definition 3.2.1. *We say that a multiplicative functional* $\boldsymbol{X}_{s,t} \in T^{(n)}$ *lies above a path* $X_t \in V$ *if* $\boldsymbol{X}^1_{s,t} = X_t - X_s$.

It is clear that there always is such a path under any multiplicative functional and that it is unique, once we have determined its value at a single time. In what follows we will use the notation (vector font, normal font) to express this relationship without further mention.

Main Lemma. Notation. A W-valued 1-form ϑ on V is a function on V whose value at any point is a linear homomorphism from V to W, that is $\vartheta(v) \in \hom(V, W)$. Suppose that ϑ is smooth enough that one can differentiate it. Denote by

$$\text{(3.31)} \qquad \begin{cases} \vartheta^1 = \vartheta\,, & \vartheta^1(v) \in \hom(V, W)\,, \\ \vartheta^2 = d\vartheta\,, & \vartheta^2(v) \in \hom(V, \hom(V, W)) \\ & \qquad \cong \hom(V \otimes V, W)\,, \\ \vartheta^k = d\vartheta^{k-1}\,, & \vartheta^k(v) \in \hom(\bigotimes_1^k V, W)\,. \end{cases}$$

Now, the multilinear map $\vartheta^k(v)$ is not symmetric in all its coefficients - and so one must have some convention on the order in which they appear. We adopt the convention that $\vartheta^k(v)\,(v_1, v_2, \ldots, v_k)$ is defined so that for smooth paths and conventional integrals

$$\text{(3.32)} \qquad \int_{s<u<t} \vartheta^k(x_u)\,(dx_u, v_2, \ldots, v_k) = \vartheta^{k-1}(x_t)\,(v_2, \ldots, v_k) - \vartheta^{k-1}(x_s)\,(v_2, \ldots, v_k)\,.$$

Recall that ϑ is a Lip$(\gamma - 1)$ one form with norm at most M providing that for $1 \leq j < \gamma$ one has the Taylor series style expression

$$\text{(3.33)} \qquad \vartheta^j(x_t)\,(v_1, v_2, \ldots, v_j) = \sum_{0 \leq i < \gamma - j} \vartheta^{j+i}(x_0)\,(x^i_{0,t}, v_1, v_2, \ldots, v_j) + R^j(x_0, x_t)\,(v_1, v_2, \ldots, v_j)\,,$$

where $\vartheta^i(x)$ and $R^i(x,y)$ are bounded in operator norm on

$$\hom\Big(\bigotimes_1^i V, W\Big)$$

with the controls

$$\begin{aligned}\|\vartheta^i(x)\| &\leq M\,,\\ \|R^i(x,y)\| &\leq M\|x-y\|^{\gamma-i}\,.\end{aligned} \tag{3.34}$$

As we noted (1.2.2), the remainder only depends on x_0, and x_t, and not on the intermediate smooth path segment. Exploiting this point, and taking a limit, we see that the identity (3.33) and estimate (3.34) hold for any sequence $(1, \boldsymbol{x}_t = \boldsymbol{x}^i_{0,t}, \ldots, \boldsymbol{x}^i_{0,t})$ arising from a geometric multiplicative functional.

We are now in a position to define the crucial almost multiplicative functional which will give us the integral. We start with a definition which is understandable for smooth paths, and then transform it in a combinatorial way so that it is clear that the functional is the restriction of a uniformly continuous function defined on all paths in $\Omega(V)^p$. This extension is easily seen to define an almost multiplicative functional when evaluated on $\Omega G(V)^p$ and this completes the definition/theorem. As a warning, this functional (which is linear) definitely does not give an almost multiplicative functional for a general element of $\Omega(V)^p$.

Predefinition 3.2.1. *For ϑ a* Lip$(\gamma-1)$ *one form with values in W, and $\boldsymbol{X}_{s,t}$ a geometric multiplicative functional of finite p-variation* (*obtained by taking a sequence of iterated integrals of a smooth path*), *define*

$$\begin{aligned}\boldsymbol{Y}^i_{s,t} = \iint\limits_{s<u_1<\cdots<u_i<t} &\sum_{l_1=1}^{[p]} \vartheta^{l_1}(X_s)\,(d\boldsymbol{X}^{l_1}_{s,u_1})\\ &\cdots \sum_{l_i=1}^{[p]} \vartheta^{l_i}(X_s)\,(d\boldsymbol{X}^{l_i}_{s,u_i})\,.\end{aligned} \tag{3.35}$$

Because the $\vartheta^{l_i}(X_s)$ are constants we may equate the expression with

$$\begin{aligned}\boldsymbol{Y}^i_{s,t} = &\sum_{l_1,\ldots,l_i=1}^{[p]} \vartheta^{l_1}(X_s)\otimes\cdots\otimes\vartheta^{l_i}(X_s)\\ &\cdot \iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1}\cdots d\boldsymbol{X}^{l_i}_{s,u_i}\,.\end{aligned} \tag{3.36}$$

Focus attention on $\iint_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i}$. *For our smooth path one has*

$$\begin{aligned}(3.37)\qquad &\iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i} \\ &\qquad = \iint\limits_V dX_{u_{1,1}} \cdots dX_{u_{1,l_1}} \cdots dX_{u_{i,1}} \cdots dX_{u_{i,l_i}} \,,\end{aligned}$$

where the domain of integration V *is given by*

$$(3.38)\qquad \begin{aligned} V = s < u_1 < \cdots < u_i < t\,, \\ s < u_{1,1} < \cdots < u_{1,l_1} = u_1 \,, \\ \vdots \\ s < u_{i,1} < \cdots < u_{i,l_i} = u_i \,. \end{aligned}$$

But this domain of integration is a product of simplexes and can be represented as a union of disjoint simplexes obtained by shuffling. Fix $\underline{l} = (l_1, \ldots, l_i)$ *and let* $\underline{u} = (u_{1,1}, \ldots, u_{1,l_1}, \ldots, u_{i,1}, \ldots, u_{i,l_i})$ *be any distinct sequence satisfying the constraints of* (3.38). *Let* $\pi_{\underline{u}}$ *denote the permutation that would reorder the numbers* $\underline{u}$ *to be increasing, and let* $\Pi_{\underline{l}}$ *denote the range of this function as a subset of the group* $\Sigma_{\|\underline{l}\|}$ *of permutations of* $\|\underline{l}\|$ *elements where* $\|\underline{l}\| = \sum_{j=1,\ldots,i} l_j$. *We can expand our integral as a sum*

$$\begin{aligned}(3.39)\qquad &\iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i} \\ &\qquad = \sum_{\pi\in\Pi_{\underline{l}}} \ \iint\limits_{s<v_1<\cdots<v_{\|\underline{l}\|}<t} dX_{v_{\pi(1)}} \cdots dX_{v_{\pi(\|\underline{l}\|)}} \,.\end{aligned}$$

Now the group Σ_n *acts on* $\bigotimes_n V$ *in the obvious way taking* $(v_1, \ldots, v_n)$ *to* $(v_{\pi(1)}, \ldots, v_{\pi(n)})$. *It follows that*

$$\begin{aligned}(3.40)\qquad &\iint\limits_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1} \cdots d\boldsymbol{X}^{l_i}_{s,u_i} \\ &\qquad = \sum_{\pi\in\Pi_{\underline{l}}} \pi\Big(\iint\limits_{s<v_1<\cdots<v_{\|l\|}<t} dX_{v_1} \cdots dX_{v_{\|\underline{l}\|}} \Big) \\ &\qquad = \sum_{\pi\in\Pi_{\underline{l}}} \pi(\boldsymbol{X}^{\|\underline{l}\|}_{s,t})\end{aligned}$$

and so finally we have reduced the integral to an expression involving only the multiplicative functional. Regarding this calculation as motivation, we give our formal definition for $\boldsymbol{Y}_{s,t}$.

Definition 3.2.2. *For any multiplicative functional $\boldsymbol{X}_{s,t}$ in $\Omega G(V)^p$ define,*

$$\boldsymbol{Y}^i_{s,t} = \sum_{l_1,\ldots,l_i=1}^{[p]} \vartheta^{l_1}(X_s)\otimes\cdots\otimes\vartheta^{l_i}(X_s)\sum_{\pi\in\Pi_{\underline{l}}}\pi(\boldsymbol{X}^{\|\underline{l}\|}_{s,t})\,. \tag{3.41}$$

Theorem 3.2.1. *For any multiplicative functional $\boldsymbol{X}_{s,t}$ in $\Omega G(V)^p$ and any one-form $\theta\in\mathrm{Lip}[\gamma-1,\{X_u,u\in[s,t]\}]$ with $\gamma>p$ the sequence $\boldsymbol{Y}_{s,t}=(1,\boldsymbol{Y}^1_{s,t},\ldots,\boldsymbol{Y}^{[p]}_{s,t})$ defined above is almost multiplicative and of finite p-variation; if $\boldsymbol{X}_{s,t}$ is controlled by ω on J where ω is bounded by L, and the $\mathrm{Lip}[\gamma-1]$ norm of θ is bounded by M, then the almost multiplicative and p-variation properties of $\boldsymbol{Y}$ are controlled by multiples of ω which depend only on γ,p,L,M.*

PROOF. Note that we also have the trivial estimate based on the size of the permutation group that

$$\begin{aligned}\left\|\iint_{s<u_1<\cdots<u_i<t} d\boldsymbol{X}^{l_1}_{s,u_1}\cdots d\boldsymbol{X}^{l_i}_{s,u_i}\right\| &< |\Pi_{\underline{l}}|\,\frac{\omega(s,t)^{\|\underline{l}\|/p}}{\beta\left(\|l\|/p\right)!}\\ &< \|\underline{l}\|!\,\frac{\omega(s,t)^{\|\underline{l}\|/p}}{\beta\left(\|\underline{l}\|/p\right)!}\,.\end{aligned} \tag{3.42}$$

We must now prove that $\boldsymbol{Y}_{s,t}$ is almost multiplicative when restricted to $\Omega G(V)^p$. For motivation of our calculations we again start by formally regarding our multiplicative functional as a sequence of iterated integrals.

$$\begin{aligned}\boldsymbol{Y}^i_{s,u} &= \iint_{s<u_1<\cdots<u_i<u}\sum_{l_1=1}^{[p]}\vartheta^{l_1}(X_s)\,(d\boldsymbol{X}^{l_1}_{s,u_1})\cdots\sum_{l_i=1}^{[p]}\vartheta^{l_i}(X_s)\,(d\boldsymbol{X}^{l_i}_{s,u_i})\\ &= \sum_{r=1,\ldots,i}\iint_{t<u_{r+1}<\cdots<u_i<u}\Bigg(\iint_{s<u_1<\cdots<u_r<t}\sum_{l_1=1}^{[p]}\vartheta^{l_1}(X_s)\,(d\boldsymbol{X}^{l_1}_{s,u_1})\\ &\qquad\cdots\sum_{l_r=1}^{[p]}\vartheta^{l_r}(X_s)\,(d\boldsymbol{X}^{l_r}_{s,u_r})\Bigg)\end{aligned}$$

(3.43)

$$\otimes \sum_{l_{r+1}=1}^{[p]} \vartheta^{l_{r+1}}(X_s)\,(d\boldsymbol{X}^{l_{r+1}}_{s,u_{r+1}})$$

$$\cdots \sum_{l_i=1}^{[p]} \vartheta^{l_i}(X_s)\,(d\boldsymbol{X}^{l_i}_{s,u_i})$$

$$= \Big(\boldsymbol{Y}_{s,t} \otimes \Big(\iint_{t<u_1<\cdots<u_j<u} \sum_{l_1=1}^{[p]} \vartheta^{l_1}(X_s)\,(d\boldsymbol{X}^{l_1}_{s,u_1})$$

$$\cdots \sum_{l_j=1}^{[p]} \vartheta^{l_j}(X_s)\,(d\boldsymbol{X}^{l_j}_{s,u_j})\Big)_{j=0}^{[p]}\Big)^i$$

This expression looks close to our target, but we must move the reference point in the second half of the expression from the time point s to the time point t. This follows from the Taylor type expression. Consider the terms $\vartheta^l(X_s)\,d\boldsymbol{X}^l_{s,u}$ where $u > t$. Then again by linearity of tensor multiplication one gets $d\boldsymbol{X}^l_{s,u} = (\boldsymbol{X}_{s,t} \otimes d\boldsymbol{X}_{t,u})^l$ and so

$$\sum_{l=1}^{[p]} \vartheta^l(X_s)\,(d\boldsymbol{X}^l_{s,u_1}) = \sum_{l=1}^{[p]} \sum_{i=0}^{l-1 \text{ or } l} \vartheta^l(X_s)\,(\boldsymbol{X}^i_{s,t} \otimes d\boldsymbol{X}^{l-i}_{t,u})$$

(3.44)
$$= \sum_{j=1}^{[p]} \sum_{i=0}^{[p]-j} \vartheta^{i+j}(X_s)\,(\boldsymbol{X}^i_{s,t} \otimes d\boldsymbol{X}^j_{t,u})$$

$$= \sum_{j=1}^{[p]} \vartheta^j(X_t)\,(d\boldsymbol{X}^j_{t,u}) + \sum_{j=1}^{[p]} R^j(X_0, X_t)\,(d\boldsymbol{X}^j_{t,u})$$

[1] and so we have

$$\boldsymbol{Y}_{s,u} = \boldsymbol{Y}_{s,t} \otimes \boldsymbol{Y}_{t,u}$$

$$+ \boldsymbol{Y}_{s,t} \otimes \Big(\sum_{l_1,\ldots,l_i=1}^{[p]} \Big(\sum_{\beta} \beta^{l_1}(X_s, X_t) \otimes \cdots \otimes \beta^{l_i}(X_s, X_t)$$

---

[1] It is exactly at this point that one is assuming the derivatives of the one form contract with the iterated integrals to produce a result that depends on the path chosen only through it's initial and terminal values. In other words, the iterated integrals differ from those of the chord $X_t - X_0$ by an element in the enveloping algebra.

$$(3.45) \qquad \cdot \iint_{t<u_1<\cdots<u_i<u} d\boldsymbol{X}^{l_1}_{t,u_1} \cdots d\boldsymbol{X}^{l_i}_{t,u_i} \Big)\Big)_{i=0}^{i=[p]}$$

where the sum is over all sequences β where

$$(3.46) \qquad \beta^l \in \{\vartheta^l(X_s), R^l(X_0, X_t)\}$$

and where for each l, one has, for at least one of l_j, that $\beta^l = R^l(X_0, X_t)$. It is then an easy matter to estimate the size of this term and see that the functional is almost multiplicative.

One has that

$$\begin{aligned}
&\Big\| \sum_{l_1,\dots,l_i=1}^{[p]} \Big(\sum_{\beta} \beta^{l_1}(X_s, X_t) \otimes \cdots \otimes \beta^{l_i}(X_s, X_t) \\
&\qquad \cdot \iint_{t<u_1<\cdots<u_i<u} d\boldsymbol{X}^{l_1}_{t,u_1} \cdots d\boldsymbol{X}^{l_i}_{t,u_i} \Big) \Big\| \\
&\le M^i \sum_{l_1,\dots,l_i=1}^{[p]} (2^i - 1)\, |\Pi_{\underline{l}}|\, (1 + |X_s - X_t|^{\gamma-1})^{i-1}\, |X_s - X_t|^{\gamma-1} \\
&\qquad \cdot \frac{\omega(t,u)^{(l_1+\cdots+l+\cdots l_i)/p}}{\beta\,(|\underline{l}|/p)!} \\
&\le M^i\, |X_s - X_t|^{\gamma-1}\, \omega(t,u)^{(l+i-1)/p} \\
&\quad \cdot \sum_{l_1,\dots,l_i=1}^{[p]} (2^i - 1)\, |\Pi_{\underline{l}}|\, (1 + |X_s - X_t|^{\gamma-1})^{i-1}\, (1 + \omega(t,u)^{(\gamma-1)/p})^{i-1}
\end{aligned}$$

where the passage from the first to second expression is based on the estimate given above for the iterated integral of iterated integrals, counting the number of β, and by exploiting the inequality

$$(3.47) \qquad \beta^{l_j}(\boldsymbol{X}_s, \boldsymbol{X}_t) < M\,(1 + \|\boldsymbol{X}_s - \boldsymbol{X}_t\|^{\gamma-1})$$

in all but one of the terms in the product, in the latter one uses the fact that the remainder type term appears at least, once to be more precise

$$\le M^i \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!} \Big) \omega(t,u)^{l+i-1/p} \Big(1 + \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!} \Big)^{\gamma-1} \Big)^{i-1}$$

$$\cdot \left(1 + \omega(t,u)^{(\gamma-1)/p}\right)^{i-1} \sum_{l_1,\dots,l_i=1}^{[p]} \frac{(2^i-1)\,|\Pi_{\underline{l}}|}{\beta\,(\|\underline{l}\|/p)!}$$

$$\le M^i \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!}\Big)^{\gamma-l} \omega(t,u)^{l+i-1/p} \Big(1 + \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!}\Big)^{\gamma-1}\Big)^{i-1}$$

$$\cdot \left(1 + \omega(t,u)^{(\gamma-1)/p}\right)^{i-1} \sum_{m=1}^{i[p]} \frac{(2^i-1)\,m!}{\beta\,(m/p)!}$$

$$\le \frac{(i\,[p]+1)!\,(2^i-1)}{\beta\,(1/p)!} M^i \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!}\Big)^{\gamma-l} \omega(t,u)^{l+i-1/p}$$

$$\cdot \Big(1 + \Big(\frac{\omega(t,u)^{1/p}}{(1/p)!}\Big)^{\gamma-1}\Big)^{i-1} \left(1 + \omega(t,u)^{(\gamma-1)/p}\right)^{i-1}$$

$$\le K(p,\beta)\, M^i\, \omega(s,u)^{\gamma+i-1/p} \Big(1 + \Big(\frac{\omega(t,u)}{(1/p)!}\Big)^{(\gamma-1)/p}\Big)^{2(i-1)}$$

and since $\gamma > p$ and $i \ge 1$ we have the estimate. The functional $\boldsymbol{Y}_{st}$ is almost multiplicative with power γ/p. It is interesting that the const ant grows so rapidly with the roughness of the path.

To finalize the argument, recall that we did some manipulations of $\boldsymbol{Y}_{st}$ where we used the representation of the terms in the iterated integral to motivate certain manipulations which were obvious for classical smooth integrals because of their general properties of linearity and additivity over disjoint simplexes. It is necessary to convince oneself that an integrated form of (3.44) holds when a geometric multiplicative functional is substituted for the iterated integrals of the smooth path. This is obvious for geometric multiplicative functionals because the algebraic identities clearly hold on a closed set containing the lifts of the smooth paths. By definition this includes $\Omega G(V)^p$.

As a consequence of this result we can define the integral of a 1-form.

Definition 3.2.3. *We say that $\hat{\boldsymbol{Y}}_{s,t}$ is the integral of the one form θ against X if $\hat{\boldsymbol{Y}}_{s,t}$ is the multiplicative functional associated to the almost multiplicative functional we defined above. In this case we write*

(3.48)
$$\hat{\boldsymbol{Y}}_{s,t} = \int_{s<u<t} \theta(X_u)\,\delta\boldsymbol{X}_u\,,$$
$$\delta\hat{\boldsymbol{Y}} = \theta(X)\,\delta\boldsymbol{X}\,.$$

We now have an integral. We also have a change of variable formula.

Corollary 3.2.1. *Suppose f is a* Lip$[\gamma]$ *map from $V \longrightarrow U$ then it induces a natural map of $\Omega G(V)^p \longrightarrow \Omega G(U)^p$ providing $\gamma > p$.*

PROOF. Apply the above theorem to the differential of f.

So just as semimartingales as a class are preserved by smooth maps, so is $\Omega G(V)^p$.

3.2.2. The two step case - p-variation less than 3.

The reader may be particularly interested in the special case which includes stochastic differential equations. For this reason we treat independently the case where one has a multiplicative functional of degree two, the more explicit approach developed here permits a stronger result. We show that it is possible to integrate any p-multiplicative functional against a Lip$[\gamma - 1]$ one form providing $\gamma > p$. Again our approach is to construct an almost multiplicative functional and although the result is almost contained in the previous section it seems worth the effort of doing the calculation explicitly in this important special case to identify the constants and (perhaps?) get a feel for how to generalise to the general case.

Even in this case there are many terms and the algebra is relatively complex. Mathematica was used by the author to keep track of some of the terms in the calculations.

Our basic idea can be summarized by saying we start with a multiplicative functional $\boldsymbol{X}$ of degree two which we think of as representing the integral and second iterated integral of a path X . We write down the obvious approximation to the integral and iterated integral of the integral of X against a 1-form. This is not multiplicative, but it is almost multiplicative. The unique multiplicative functional that is appropriately close is regarded as the integral of X against the 1-form.

Fix $2 \leq p < 3$. Then $\boldsymbol{X}_{st}$ is a multiplicative functional on J with p-variation controlled by ω if

$$\|\boldsymbol{X}^1_{ts}\| \leq \omega(t,s)^{1/p} \qquad \text{and} \qquad \|\boldsymbol{X}^2_{ts}\| \leq \omega(t,s)^{2/p}\,.$$

Suppose that θ is a 1-form that is Lip$[\gamma]$ where $p - 1 < \gamma < 2$.

By Taylor's theorem

$$\left\|\theta(X_t) - \theta(X_s) - \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^1_{st})\right\| < M\,\omega(t,s)^{\gamma/p}\,. \tag{3.49}$$

So if we wish to approximate the iterated integrals of $\boldsymbol{Y}$ the "integral" of $\boldsymbol{X}$ against θ, it makes sense to consider

$$\boldsymbol{Y}_{st} = \left\{1, \theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st}), \theta(X_s)\otimes\theta(X_s)(\boldsymbol{X}^2_{st})\right\}.$$

Clearly, $\boldsymbol{Y}_{s,t}$ has finite p-variation controlled by $2\,M\,\omega$. We will now establish the claim that it is also an almost multiplicative functional

$$\boldsymbol{Y}_{st} = \left\{1, \theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st}), \theta(X_s)\otimes\theta(X_s)\,(\boldsymbol{X}^2_{st})\right\},$$

$$\boldsymbol{Y}_{tu} = \left\{1, \theta(X_t)\,(\boldsymbol{X}^1_{tu}) + \frac{1}{2}\,(d\theta)\,(X_t)\,(\boldsymbol{X}^2_{tu}), \theta(X_t)\otimes\theta(X_t)\,(\boldsymbol{X}^2_{tu})\right\},$$

$$\begin{aligned}\boldsymbol{Y}_{st}\otimes\boldsymbol{Y}_{tu} = \Big\{1, &\,\theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st}) + \theta(X_t)\,(\boldsymbol{X}^1_{tu})\\ &+ \frac{1}{2}\,(d\theta)\,(X_t)\,(\boldsymbol{X}^2_{tu}),\\ &(\theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)(\boldsymbol{X}^2_{st}))\\ &\otimes(\theta(X_t)\,(\boldsymbol{X}^1_{tu}) + \frac{1}{2}\,(d\theta)\,(X_t)\,(\boldsymbol{X}^2_{tu}))\\ &+\theta(X_s)\otimes\theta(X_s)\,(\boldsymbol{X}^2_{st}) + \theta(X_t)\otimes\theta(X_t)\,(\boldsymbol{X}^2_{tu})\Big\},\end{aligned} \tag{3.51}$$

$$\begin{aligned}\boldsymbol{Y}_{su} - \boldsymbol{Y}_{st}\otimes\boldsymbol{Y}_{tu} = \Big\{0, &\,\theta(X_s)\,(\boldsymbol{X}^1_{su}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{su})\\ &-(\theta(X_s)\,(\boldsymbol{X}^1_{su}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st})\\ &+\theta(X_t)\,(\boldsymbol{X}^1_{tu}) + \frac{1}{2}\,(d\theta)\,(X_t)\,(\boldsymbol{X}^2_{tu}))\,,\\ &\theta(X_s)\otimes\theta(X_s)\,(\boldsymbol{X}^2_{su})\\ &-((\theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st}))\\ &\otimes(\theta(X_s)\,(\boldsymbol{X}^1_{su}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{su}))\\ &+\theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st})\\ &+\theta(X_t)\,(\boldsymbol{X}^1_{tu}) + \frac{1}{2}\,(d\theta)\,(X_t)\,(\boldsymbol{X}^2_{tu}))\Big\},\end{aligned} \tag{3.52}$$

now recalling Taylor's theorem

$$\theta(X_t) = \theta(X_s) + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^1_{st}) + r_1(t,s)\,,$$
$$\|r_1(t,s)\| < M\,\omega(t,s)^{\gamma/p}$$

and

$$d\theta(X_t) = d\theta(X_s) + r_2(t,s)\,, \qquad \|r_2(t,s)\| < M\,\omega(t,s)^{(\gamma-1)/p}\,. \tag{3.53}$$

We use these approximations to estimate $\boldsymbol{Y}_{su} - \boldsymbol{Y}_{st} \otimes \boldsymbol{Y}_{tu}$. Substituting both approximations into the 1-tensor component of $\boldsymbol{Y}_{st} \otimes \boldsymbol{Y}_{tu}$, substituting only the first into the 2-tensor component, and expanding out each term in $\boldsymbol{X}^i_{s,u}$ within $\boldsymbol{Y}_{su}$ in terms of $\boldsymbol{X}^j_{st}, \boldsymbol{X}^j_{tu}$ using the multiplicative proper ty for $\boldsymbol{X}$ one has after a tedious calculation with many terms (or using Mathematica after a relatively complex set of manipulations) the three terms of different tensor degree in $\boldsymbol{Y}_{su} - \boldsymbol{Y}_{st} \otimes \boldsymbol{Y}_{tu}$ in increasing order of complexity.

The zero'th order term is clearly zero.

The first order term is $r_1(s,t)\boldsymbol{X}^1_{t,u} + r_2(s,t)\,\boldsymbol{X}^2_{t,u}$ and

$$\begin{aligned}\|r_1(s,t)\,\boldsymbol{X}^1_{t,u} + r_2(s,t)\boldsymbol{X}^2_{t,u}\| \\ &\le M\,(\omega(s,t)^{\gamma/p}\,\omega(t,u)^{1/p} + \omega(s,t)^{(\gamma-1)/p}\,\omega(t,u)^{2/p}) \\ &\le 2\,M\,\omega(s,u)^{(\gamma+1)/p}\,,\end{aligned} \tag{3.54}$$

giving the required estimate.

The second order term breaks naturally (if somewhat painfully) into a sum of 15 terms, which under our assumptions are of 5 different magnitudes.

$$\begin{aligned}&(\theta(X_s) \otimes d\theta(X_s))\,(\boldsymbol{X}^1_{s,t} \otimes \boldsymbol{X}^2_{t,u}) \\ &+ (\theta(X_s) \otimes d\theta(X_t))\,(\boldsymbol{X}^1_{s,t} \otimes \boldsymbol{X}^2_{t,u}) \\ &+ (d\theta(X_s) \otimes \theta(X_s))\,(\boldsymbol{X}^1_{s,t} \otimes \boldsymbol{X}^2_{t,u} + \boldsymbol{X}^2_{s,t} \otimes \boldsymbol{X}^1_{t,u}) \\ &+ (\theta(X_s) \otimes d\theta(X_s))\,(\boldsymbol{X}^1_{s,t} \otimes \boldsymbol{X}^1_{s,t} \otimes \boldsymbol{X}^1_{t,u}) \\ &+ (\theta(X_s) \otimes r_1(s,t))\,(\boldsymbol{X}^2_{t,u} + \boldsymbol{X}^1_{t,u} \otimes \boldsymbol{X}^1_{s,t})\end{aligned} \tag{3.55}$$

$$+\left(r_1(s,t)\otimes\theta(X_s)\right)\left(\boldsymbol{X}^2_{t,u}\right) \tag{3.56}$$

$$\begin{aligned}&+\left(d\theta(X_s)\otimes d\theta(X_t)\right)\left(\boldsymbol{X}^2_{s,t}\otimes\boldsymbol{X}^2_{t,u}\right)\\&+\left(d\theta(X_s)\otimes d\theta(X_s)\right)\left(\boldsymbol{X}^1_{s,t}\otimes\boldsymbol{X}^1_{s,t}\otimes\boldsymbol{X}^2_{t,u}\right)\\&+\left(d\theta(X_s)\otimes d\theta(X_s)\right)\left(\boldsymbol{X}^1_{s,t}\otimes\boldsymbol{X}^2_{s,t}\otimes\boldsymbol{X}^1_{t,u}\right)\end{aligned} \tag{3.57}$$

$$\begin{aligned}&+\left(d\theta(X_s)\otimes r_1(s,t)\right)\left(\boldsymbol{X}^1_{t,u}\otimes\boldsymbol{X}^2_{s,t}\right)\\&+\left(d\theta(X_s)\otimes r_1(s,t)\right)\left(\boldsymbol{X}^1_{s,t}\otimes\boldsymbol{X}^2_{t,u}\right)\\&+\left(r_1(s,t)\otimes d\theta(X_s)\right)\left(\boldsymbol{X}^1_{s,t}\otimes\boldsymbol{X}^2_{t,u}\right)\end{aligned} \tag{3.58}$$

$$+\left(r_1(s,t)\otimes r_1(s,t)\right)\left(\boldsymbol{X}^2_{t,u}\right) \tag{3.59}$$

and so one has that the norm of the expression above is less than

$$\begin{aligned}M^2\,(5\,\omega(s,u)^{3/p}+3\,\omega(s,u)^{(\gamma+2)/p}+3\,\omega(s,u)^{4/p}\\+3\,\omega(s,u)^{(\gamma+3)/p}+\omega(s,u)^{(2\gamma+2)/p})\end{aligned} \tag{3.60}$$

and providing $\omega(s,u)<1$ we have the simpler bound

$$15\,M^2\,\omega(s,u)^{3/p}\,,\qquad \omega(s,u)<1\,. \tag{3.61}$$

Recalling our assumption that θ is a 1-form that is Lip$[\gamma]$ where $p-1<\gamma<2$ we see that both errors are controlled to a degree greater than one in ω. This leads us to conclude that Y is an almost multiplicative functional. Our approach used the multiplicative property, but never required the geometric property of $\boldsymbol{X}$. As above we define the integral

$$\int\limits_{s<u<t}\theta(X_u)\,\delta\boldsymbol{X}\,, \tag{3.62}$$

to be the associated multiplicative functional.

3.2.3. Continuity of the integral.

It is an immediate corollary of our results so far, that the integral

$$\int_{s<u<t} \theta(X_u)\, \delta \boldsymbol{X} \tag{3.63}$$

is a continuous map from (geometric) multiplicative functionals and Lip$[\gamma - 1]$ one forms to p-multiplicative functionals. Since it is clear that the integral of a smooth path produces a geometric functional, it follows from the continuity of the map that the integral against any element of $\Omega G(V)^p$ produces a multiplicative functional in $\Omega G(W)^p \subset \Omega(W)^p$.

In more detail, the almost multiplicative functional associated with a geometric functional

$$\boldsymbol{Y}^i_{s,t} = \sum_{l_1,\ldots,l_i=1}^{[p]} \vartheta^{l_1}(X_s) \otimes \cdots \otimes \vartheta^{l_i}(X_s) \sum_{\pi \in \Pi_{\underline{l}}} \pi(\boldsymbol{X}^{\|\underline{l}\|}_{s,t}) \tag{3.64}$$

is clearly continuous in the sense that if $\boldsymbol{X}, \boldsymbol{X}'$, are multiplicative functionals controlled by ω and satisfying

$$\|(\boldsymbol{X}_{s,t} - \boldsymbol{X}'_{s,t})^i\| < \varepsilon \, \frac{\omega(s,t)}{\beta\,(i/p)!}\ , \tag{3.65}$$

moreover the $\|X_u - X'_u\| \leq \varepsilon$, and the finitely many functions $x \longrightarrow \vartheta^{l_1}(x) \otimes \cdots \otimes \vartheta^{l_i}(x)$ have a uniform modulus of continuity $\sigma(\varepsilon, M, p)$, so one has the estimate on the almost multiplicative functionals

$$\begin{aligned} &\|\boldsymbol{Y}^i_{s,t} - \boldsymbol{Y}'^i_{s,t}\| \\ &\qquad \leq \Big(\sum_{l_1,\ldots,l_i=1}^{[p]} (\varepsilon\, M^{\|\underline{l}\|}\, |\Pi_{\underline{l}}| \\ &\qquad\qquad + \sigma(\varepsilon, M, p))\, \frac{\omega(s,t)^{|l|/p - i/p}}{\beta\,(|l|/p)!} \Big) \omega(s,t)^{i/p} \end{aligned} \tag{3.66}$$

and so we can apply the continuity theorem for the construction of a multiplicative functional from an almost multiplicative functional to deduce the following

Theorem 3.2.2. *If $\boldsymbol{X}, \boldsymbol{X}'$, are geometric multiplicative functionals of finite p-variation controlled by ω with $\omega(s,t) < L$ for $s,t \in J$, and θ is a one form with a* $\mathrm{Lip}[\gamma - 1]$ *norm at most M then there is a function $\delta(\varepsilon, L, M, p)$ continuous and zero if $\varepsilon = 0$ such that if*

$$\|(\boldsymbol{X}_{s,t} - \boldsymbol{X}'_{s,t})^i\| < \varepsilon \frac{\omega(s,t)^{i/p}}{\beta\,(i/p)!}\,, \qquad i \le [p] \tag{3.67}$$

and

$$\|X_u - X'_u\| \le \varepsilon\,, \tag{3.68}$$

then for $i \le [p]$

$$\begin{aligned}\Big\|\Big(\int\limits_{s<u<t} \theta(X_u)\,\delta\boldsymbol{X} - \int\limits_{s<u<t} \theta(X'_u)\,\delta\boldsymbol{X}'\Big)^i\Big\| \\ < \delta(\varepsilon, L, M, p)\,\frac{\omega(s,t)^i}{(i/p)!}\,.\end{aligned} \tag{3.69}$$

Similar estimates apply to the variation of the one form.

Corollary 3.2.2. *If $\boldsymbol{X} \in \Omega G(V)^p$ then*

$$\int\limits_{s<u<t} \theta(X_u)\,\delta\boldsymbol{X} \qquad \textit{is in } \Omega G(W)^p \subset \Omega(W)^p\,. \tag{3.70}$$

Continuity in the case $p < 3$ for non-geometric functionals. In the situation where $p < 3$ we have the alternative description of our almost multiplicative functional valid for any $\boldsymbol{X} \in \Omega(V)^p$, and we may check the continuity directly in this case as well. We explicitly compute the changes to the almost multiplicative functional

$$\begin{aligned}\boldsymbol{Y}_{st} = \Big\{1, \theta(X_s)\,(\boldsymbol{X}^1_{st}) + \frac{1}{2}\,(d\theta)\,(\boldsymbol{X}_s)\,(\boldsymbol{X}^2_{st}), \\ \theta(X_s) \otimes \theta(X_s)\,(\boldsymbol{X}^2_{st})\Big\}\,.\end{aligned} \tag{3.71}$$

Suppose that

$$\boldsymbol{X}_s = \hat{\boldsymbol{X}}_s + e\,0_s\,, \qquad \boldsymbol{X}^1_{st} = \hat{\boldsymbol{X}}^1_{st} + e\,1_{st}\,, \qquad \boldsymbol{X}^2_{st} = \hat{\boldsymbol{X}}^2_{st} + e\,2_{st}\,, \tag{3.72}$$

where the approximation errors satisfy

$$e\,0_s < \varepsilon\,, \qquad e\,1_{st} < \varepsilon\,\omega(s,t)^{1/p}\,, \qquad e\,2_{st} < \varepsilon\,\omega(s,t)^{2/p}\,. \tag{3.73}$$

Then

$$\begin{aligned} \boldsymbol{Y}_{st} - \hat{\boldsymbol{Y}}_{st} = \Big\{0,\; & \theta(X_s)\,(\boldsymbol{X}^1_{st}) - \theta(\hat{X}_s)\,(\hat{\boldsymbol{X}}^1_{st}) \\ & + \frac{1}{2}\,(d\theta)\,(X_s)\,(\boldsymbol{X}^2_{st}) - \frac{1}{2}\,(d\theta)\,(\hat{X}_s)\,(\hat{\boldsymbol{X}}^2_{st}), \\ & \theta(X_s) \otimes \theta(X_s)\,(\boldsymbol{X}^2_{st}) - \theta(\hat{X}_s) \otimes \theta(\hat{X}_s)\,(\hat{\boldsymbol{X}}^2_{st})\Big\} \end{aligned} \tag{3.74}$$

and so

$$\begin{aligned} \|(\boldsymbol{Y}_{st} - \hat{\boldsymbol{Y}}_{st})^i\| \le \Big\{0,\; & M\,\varepsilon\,\omega(s,t)^{1/p} + M\,\varepsilon\,\omega^{1/p}\,\omega(s,t)^{1/p} \\ & + \frac{1}{2}\,M\,(\varepsilon\,\omega(s,t)^{1/p})^{\gamma-1}\,\omega(s,t)^{2/p} \\ & + \frac{1}{2}\,M\,\varepsilon\,\omega(s,t)^{2/p}, \\ & M^2\,((\varepsilon^2+\varepsilon)\,\omega(s,t)^{2/p} \\ & \quad + 2\,(\varepsilon^2+\varepsilon)\,\omega(s,t)^{3/p} \\ & \quad + \varepsilon^3\,\omega(s,t)^{4/p})\Big\}\,, \end{aligned} \tag{3.75}$$

with $i = 0,1,2$, and providing $\omega(s,t) < 1$, $\varepsilon < 1$, one has the more intelligible inequality

$$\|(\boldsymbol{Y}_{st} - \hat{\boldsymbol{Y}}_{st})^i\| \le \{0, 3\,M\,\varepsilon^{\gamma-1}\,\omega(s,t)^{1/p}, 7\,M^2\,\varepsilon\,\omega(s,t)^{2/p}\}\,, \tag{3.76}$$

with $i = 0,1,2$, establishing the continuity of the map into almost multiplicative functionals.

4. Differential equations, putting it all together.

In this section we achieve our main objective of showing that the Itô functional extends uniquely to a continuous map defined on the rough paths in $\Omega G(V)^p$ providing the defining vector fields are $\mathrm{Lip}[\gamma]$

and $\gamma > p$. This permits, in a reasonably complete way, the solution of differential equations driven by rough (but geometric) multiplicative functionals. It completely removes the finite dimensional Lie algebra assumption.

The key estimate will be the one we established for the integration of one forms; this together with a reasonably delicate exploitation of inhomogeneity will show Picard's iteration scheme converges. The argument will be split into a number of distinct steps. But first we must be precise about our concept or definition of a solution!

4.1. Giving the differential equation meaning.

Take a smooth path X_t in V and a linear map f from V into the Lipschitz vector fields on a vector space W, then one may use schoolboy integration to define a solution to our basic equation. Classically, one could say the path Y_t solves the equation

$$dY_t = f(Y_t)\,dX_t\,, \qquad Y_0 = a\,, \tag{4.1}$$

providing Y_t satisfies the integral equation

$$Y_t = a + \int_{0<u<t} f(Y_u)\,dX_u\,. \tag{4.2}$$

Observe that we can reformulate this integral identity in a trivially different way

$$\begin{aligned} X_t &= X_0 + \int_{0<u<t} dX_u\,, \\ Y_t &= a + \int_{0<u<t} f(Y_u)\,dX_u\,. \end{aligned} \tag{4.3}$$

Consider the one form on $V \oplus W$ with values in $V \oplus W$ defined by

$$h((x,y))\,(dX,dY) = (dX, f(y)\,dX)\,. \tag{4.4}$$

Then for smooth paths the integral equation (4.3) can be rewritten as

$$(X_t, Y_t) = (X_0, a) + \int_0^t h(X_u, Y_u)\,(dX_u, dY_u)\,. \tag{4.5}$$

Putting $Z_t = (X_t, Y_t)$ we can say that a solution to (4.1) is a lift of the path X_t to a path in $V \oplus W$ satisfying

$$
\begin{aligned}
&Z_t - Z_0 = \int\limits_{0<u<t} h(Z_u)\, dZ_u \ , \\
&Z_0 = (X_0, a)\,.
\end{aligned}
\tag{4.6}
$$

Although this transformation may seem essentially trivial in the classical setting, for us it is not really so. We have no difficulty extending this characterisation to rough signals.

Definition 4.1.1. *Let $\boldsymbol{X} \in \Omega G(V)^p$ be a geometric multiplicative functional projecting onto the path X_t, and let f be a linear map from V into the* Lip$[\gamma - 1, W]$ *vector fields. A solution to the equation*

$$
d\boldsymbol{Y} = f(Y_t)\, d\boldsymbol{X}\,, \qquad Y_0 = a\,, \tag{4.7}
$$

is an extension of $\boldsymbol{X}$ to $\boldsymbol{Z} \in \Omega G(V \oplus W)^p$ such that $\boldsymbol{Z}$ projects onto $Z_t = (X_t, Y_t)$, $Y_0 = a$, and such that $\boldsymbol{Z}$ satisfies $\delta \boldsymbol{Z} = h(Z_t)\, \delta \boldsymbol{Z}$.

The main point to notice is that we do not treat the solution as an independent object, but rather as an extension of the original driving signal. In particular, we require the existence of cross iterated integrals between driving signal and solution to be constructed. On the one hand this seems a bonus, if we can construct integrals between solution and driving signal so much the better; on the other hand it is essential, we could not make sense of the integral at all for rough signals without some cross information between integrand and integrator. The author is remi nded of those induction arguments which only work if you prove a stronger result than you were aiming for. In any case, the definition is clearly consistent with the classical one. If $\boldsymbol{X} \in \Omega G(V)^p$ is a smooth path with its iterated integrals, the classical solution, its iterated integrals, together with the cross integrals with the driving signal, together satisfy the extended equation.

Our approach requires that the vector fields in the equation have a smoothness related to the roughness of the path. This was necessary for the integral to make sense. However, as in the classical situation, the smoothness required of the vector fields in the definition is less than that required for uniqueness.

The main purpose of this part of the paper, and indeed of the entire paper is to prove the following theorem.

Theorem 4.1.1. *Suppose that $f : V \longrightarrow \mathrm{Lip}[\gamma, W, W]$ is a linear map into Lipschitz vector fields. Then consider the Itô map $X \longrightarrow (X, Y)$ defined for smooth paths by*

$$dY_t = f(Y_t)\, dX_t\,, \qquad Y_0 = a\,. \tag{4.8}$$

Define the one form h by

$$h((x, y))\,(dX, dY) = h(y)\,(dX, dY) = (dX, f(y)\,dX)\,.$$

For any geometric multiplicative functional $\boldsymbol{X} \in \Omega G(V)^p$ with $1 \leq p < \gamma$ there is exactly one geometric multiplicative functional extension $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y}) \in \Omega G(V \oplus W)^p$ such that if $Y_t = \boldsymbol{Y}^1_{0,t} + a$ then $\boldsymbol{Z}$ satisfies the rough differential equation

$$\delta \boldsymbol{Z} = h(Y_t)\,\delta \boldsymbol{Z}\,. \tag{4.9}$$

Moreover this solution to the rough differential equation is constructed by Picard iteration, there is a small interval $[0, T]$ whose length can be controlled entirely in terms of the control on the roughness of X and of f and the rate so that the convergence of this iteration scheme is faster than the given exponential rate on the interval. The Itô map is uniformly continuous and the map $\boldsymbol{X} \longrightarrow \boldsymbol{Z}$ is the unique continuous extension of the Itô map from $\Omega G(V)^p$ to $\Omega G(V \oplus W)^p$.

Our convergence theorem for Picard iteration requires that $\gamma > p$, and constructively produces a unique solution; the extension of Peano's theorem to show existence under the weaker hypothesis $\gamma > p - 1$ is open (except in the case where $p < 2$; here a fixed point argument can be applied to show existence and A. M. Davie (Edinburgh - private communication) has given the author examples to show that the solution need not be unique $\gamma < p$ [14], [15]).

We may define Picard iteration as follows

$$\begin{aligned} \boldsymbol{Z}^{n+1}_{s,t} &= \int\limits_{s<u<t} h(Z^n_u)\,\delta \boldsymbol{Z}^n\,, \\ Z^n_0 &= (b, a)\,, \end{aligned} \tag{4.10}$$

where Z^n is uniquely determined by $Z^n_0 = (b, a)$; the choice of b is irrelevant to the definition as h does not depend in any way on the first coordinate of $\boldsymbol{Z}$. If we can prove that the multiplicative functionals $\boldsymbol{Z}^n$ converge in $\Omega G(V)^p$, then it is routine from our result about the continuity of the integration against one forms that the limit will be a fixed point of the functional and so our desired solution.

However, in contrast to the normal contraction mapping argument, it seems essential to consider a more complicated iteration so that we might keep track of the joint interactions of more terms.

Step 1. **Norms on tensor algebras over finite sums of vector spaces.** There are many different equivalent norms one could use on the tensor algebra over the space $V \oplus W$; we will use an induction argument where a choice adapted to the possibilities for independently scaling the different coordinates will simplify the proof.[1]

The tensors of fixed degree over a vector space admit a further direct sum decomposition if the underlying vector space is already a direct sum

$$\begin{aligned} T^{(n)}(V \oplus W) &= \bigoplus_{j=0}^{n} (V \oplus W)^{\otimes j}, \\ (V \oplus W)^{\otimes j} &= Z^{j,0} \oplus Z^{j-1,1} \oplus Z^{j-2,2} \oplus \cdots \oplus Z^{0,j}, \end{aligned} \tag{4.11}$$

where $Z^{j-k,k}$ comprises those tensors that are homogeneous of degree $j-k$ in V and k in W in whatsoever order.

REMARK 4.1.1.–REQUIREMENT. Let $\boldsymbol{z} = \boldsymbol{z}^{j,0} + \boldsymbol{z}^{j-1,1} + \boldsymbol{z}^{j-2,2} + \cdots + \boldsymbol{z}^{0,j}$ represent the decomposition of an element $\boldsymbol{z} \in (V \oplus W)^{\otimes j}$, then the norm on $(V \oplus W)^{\otimes j}$ should be chosen to have the property that $\|\boldsymbol{z}\| = \sup_{k \le j} \|\boldsymbol{z}^{j-k,k}\|$.

Definition 4.1.2. *A multiplicative functional* $\boldsymbol{Z}$ *in* $\Omega(V \oplus W)^p$ *is controlled by* ω *if*

$$\|\boldsymbol{Z}_{s,t}^{j-k,k}\| \le \frac{\omega(s,t)^{j/p}}{\beta\,(j-k/p)!\,(k/p)!}, \tag{4.12}$$

for all $j \le [p]$.

Of course this control is comparable with the one that ignores the inhomogeneity.

[1] See also the earlier section inhomogeneous degrees of smoothness.

Step 2. **Rescaling and Tensor Algebras.** If S is a linear automorphism on V then it induces a natural graded algebra homomorphism $\tilde{S}$ on the tensor algebra, taking $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ to $Sv_1 \otimes Sv_2 \otimes \cdots \otimes Sv_n$. Apply this to the scaling operators $S_\varepsilon(v) = \varepsilon\, v$. Their extensions act by multiplying the tensors of degree k by ε^k so that $\tilde{S}_\varepsilon(\boldsymbol{a}) = (1, \varepsilon\, \boldsymbol{a}_1, \varepsilon^2\, \boldsymbol{a}_2, \ldots, \varepsilon^n\, \boldsymbol{a}_n)$. These operators are very important to us, but the general notation is clumsy, so we shorten it.

Definition 4.1.2. *We will use the notation* $\varepsilon\, \boldsymbol{X}_{s,t}$ *for* $\tilde{S}_\varepsilon(\boldsymbol{X}_{s,t})$.

Because $\tilde{S}$ is always an algebra homomorphism $\varepsilon\, \boldsymbol{X}_{s,t}$ is also a multiplicative functional, leading to the slightly peculiar but correct notation $\varepsilon\, \boldsymbol{X}_{s,t} \otimes \varepsilon\, \boldsymbol{X}_{t,u} = \varepsilon\, \boldsymbol{X}_{s,u}$.

Consider the linear projections $P_V : V \oplus W \longrightarrow V$, and $P_W : V \oplus W \longrightarrow W$; then if $\boldsymbol{Z}$ is a multiplicative functional in the tensor algebra over $V \oplus W$, let $\boldsymbol{X} = P_V \boldsymbol{Z}$ and $\boldsymbol{Y} = P_W \boldsymbol{Z}$ be the associated multiplicative functionals. We will frequently use the notation $(\boldsymbol{X}, \boldsymbol{Y})$ for $\boldsymbol{Z}$ to remind the reader of the direct sum structure, however the multiplicative functional $(\boldsymbol{X}, \boldsymbol{Y})$ is *not determined* by $\boldsymbol{X}, \boldsymbol{Y}$ separately, as it involves cross terms.

It is possible to scale the complementary subspaces of a direct sum differently and we use the shorthand $(\varepsilon\, \boldsymbol{X}, \phi\, \boldsymbol{Y})_{st}$ for the multiplicative functional $\tilde{S}_{\varepsilon\phi}(\boldsymbol{X}, \boldsymbol{Y})_{st}$ where $S_{\varepsilon\phi}(\boldsymbol{v} + \boldsymbol{w}) = \varepsilon\, \boldsymbol{v} + \phi\, \boldsymbol{w}$.

Consider how this inhomogeneous scaling interacts with a control on the p-variation:

Lemma 4.1. *Let* $\boldsymbol{X} \in \Omega(V)^p$ *be controlled by* $\omega(s,t)$ *so that*

$$\|\boldsymbol{X}^j_{s,t}\| \le \frac{\omega(s,t)^{j/p}}{\beta\, (j/p)!} \tag{4.13}$$

and $(\boldsymbol{X}, \boldsymbol{Y}) \in \Omega(V \oplus W)^p$ *be an extension of* $\boldsymbol{X}$. *Suppose* $(\boldsymbol{X}, \boldsymbol{Y})_{st}$ *is controlled by* $K\, \omega(s,t)$. *Then* $(\boldsymbol{X}, \phi \boldsymbol{Y})_{st}$ *is controlled by*

$$\max\{1, \phi^{kp/j}\, K :\ 1 \le k \le j \le [p]\}\, \omega(s,t)\,. \tag{4.14}$$

In particular, if $\phi < K^{-[p]/p} < 1$ *then* $(\boldsymbol{X}, \phi\, \boldsymbol{Y})_{st}$ *is controlled by* $\omega(s,t)$.

PROOF. Let $\boldsymbol{Z}^j_{s,t} = (\boldsymbol{X}, \boldsymbol{Y})^j_{s,t}$ be the component of the multiplicative functional of degree j and let $\boldsymbol{Z}^{j-k,k}_{s,t}$ denote the component of this tensor of degree $j-k$ in V and k in W. Then by assumption

$$\|\boldsymbol{Z}_{s,t}^{j-k,k}\| \leq \frac{(K\,\omega(s,t))^{j/p}}{\beta\,(j-k/p)!\,(k/p)!}\,, \tag{4.15}$$

therefore

$$\|\tilde{S}_{1\phi}(\boldsymbol{Z}_{s,t}^{j-k,k})\| \leq \phi^k\,\frac{(K\,\omega(s,t))^{j/p}}{\beta\,(j-k/p)!\,(k/p)!}\,, \tag{4.16}$$

but $\boldsymbol{Z}_{s,t}^{j,0} = \boldsymbol{X}_{s,t}^{j}$ and so

$$\|\boldsymbol{Z}_{s,t}^{j,0}\| \leq \frac{\omega(s,t)^{j/p}}{\beta\,(j/p)!}\,, \tag{4.17}$$

without any constant. It follows that $(\boldsymbol{X},\phi\,\boldsymbol{Y})_{st}$ is controlled by

$$\max\{1,\phi^{kp/j}\,K:\ 1\leq k\leq j\leq [p]\}\,\omega(s,t) \tag{4.18}$$

as required.

Step 3. **The boundedness of the Picard integral operator.** As a simple application of the scaling lemma we have just established, we prove the following a priori bound.

Lemma 4.1.2. *Let $\boldsymbol{Z}^{(0)}$ be the initial multiplicative functional in the Picard iteration scheme defined recursively by* (4.10)*. Suppose $\boldsymbol{Z}^{(0)}$ is controlled by ω_0. Then all iterates $\boldsymbol{Z}^{(j)}$ are uniformly controlled by*

$$\omega = \max\{1, K\,(M,p,\gamma)^{[p]}\}\,\omega_0$$

on the time interval $J = \{u:\ \omega(0,u) < 1\}$.

Here M is the Lip$[\gamma-1]$ *norm of f on*

$$\left\{\omega:\ \|\omega - a\| \leq \frac{1}{\beta}\left(\frac{1}{p}\right)!\right\},$$

and K is the constant introduced in Theorem 3.2.1.

Proof. First we condition the problem. Suppose that the initial point $\boldsymbol{Z}_{s,t}^{(0)} = (\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ in our Picard iteration is of finite p-variation controlled by ω_0. For any $\varepsilon > 0$ we may choose a regular

$$\omega = \max\{\varepsilon^{-p}, 1\}\,\omega_0$$

so that $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ is controlled by ω, and a short interval depending on ω where $\omega < 1$. We choose $\varepsilon = K(M,p,\gamma)^{-[p]/p}$ where K is the function derived in Theorem 3.2.1, and M, is defined to be the Lip$[\gamma-1]$ norm of the one form $h(x,y)$ restricted to the domain

$$V \times \left\{ w : \ \|w-a\| < \frac{1}{\beta}\left(\frac{1}{p}\right)! \right\}.$$

We now proceed by induction. Suppose that $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ is controlled by ω where $\omega < 1$. The control on $(\boldsymbol{Y}^{(0)}_{0,t})^1$ ensures that its projection onto the path $Y^{(0)}_u$ starting at a remains in the ball of radius $(1/\beta)\,(1/p)!$ centred on a. Observe that the multiplicative functional

$$\int_{s<u<t} h((\varepsilon^{-1}X_u, Y^{(0)}_u))\,\delta(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(0)})\,(\varepsilon^{-1}X_0,Y^{(0)}_0)_t = (\varepsilon^{-1}\,b,a) \tag{4.19}$$

equals $(\varepsilon^{-1}\boldsymbol{X},\varepsilon^{-1}\boldsymbol{Y}^{(1)})_{s,t}$ where $(\boldsymbol{X},\boldsymbol{Y}^{(1)})_{s,t}$ is the Picard iterate of $(\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ defined in (4.10). By Theorem 3.2.1, it is controlled by $K(M,p,\gamma)\,\omega$ on the chosen time interval; here K depends only on the explicit variables (we have arranged that $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ is controlled by ω where $\omega < 1$).

The difference in homogeneity between $(\varepsilon^{-1}\boldsymbol{X},\varepsilon^{-1}\boldsymbol{Y}^{(1)})_{s,t}$ and our starting data $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(0)})_{s,t}$ is crucial to the analysis. If the reader finds the unfamiliar notation difficult then the equivalent formulation for smooth paths is

$$\begin{aligned} \varepsilon^{-1}Y^{(1)}_{st} &= \int_s^t f\,(Y^{(0)}_u)\,d\varepsilon^{-1}X_u\ , \\ \varepsilon^{-1}X_{st} &= \int_s^t d\varepsilon^{-1}X_u\ . \end{aligned} \tag{4.20}$$

By assumption $\varepsilon \le K\,(M,p,\gamma)^{-[p]/p}$, so we may apply Lemma 4.1.1 to prove that the rescaled functional $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(1)})_{s,t}$ is controlled by $\omega < 1$. This concludes the induction. We deduce that all the Picard iterates $(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(n)})_{s,t}$ are uniformly controlled by this same $\omega < 1$ on this fixed time interval.

An obvious extension of the same idea shows that

$$(\varepsilon^{-1}\boldsymbol{X},\boldsymbol{Y}^{(n)},\boldsymbol{Y}^{(n+1)})_{s,t}$$

is also uniformly bounded for a different fixed choice of ε, ω and the time interval. This observation will be useful to us later.

This result only requires the minimal smoothness condition required to make sense of the equation. It can be interpreted as a compactness result and can probably be used to deduce a Peano theorem in the general case although we have not pursued the matter.

The main existence result is a more subtle and complicated version of the same approach.

Step 4. **A division lemma.** Suppose that f is a Lip$[\gamma]$ vector field on W, then there exists a function g which is Lip$[\gamma - 1]$ on $W \times W$ and such that

$$f^i(x) - f^i(y) = \sum_j (x-y)^j \, g^{ij}(x,y) \,. \tag{4.21}$$

The function g is not uniquely defined, but for example the mean value of *df* along the ray from x to y [2] will do perfectly well. Thus we can rewrite the classical Picard iteration in the more useful form

$$\begin{aligned} &(Y_t^{(n+1)} - Y_t^{(n)}) \\ &\qquad = \int\limits_{0<u<t} (Y_u^{(n)} - Y_u^{(n-1)}) \, g(Y_u^{(n)}, Y_u^{(n-1)}) \, dX_u \,. \end{aligned} \tag{4.22}$$

The crucial difference between the earlier formulation of Picard iteration and the approach here is that we have introduced an expression which is quasi-linear in $(Y_n - Y_{n-1})$. We will really be able to take advantage of this and push the scaling arguments we introduced above.

Interpreting the integral (4.22) requires the extra smoothness we assume for our main theorem on the convergence of Picard's iterative scheme.

Step 5. **Defining the correct iteration.** In fact we consider recursively, a sequence containing a wider series of interrelated objects

$$(\boldsymbol{Z}^{(n)}, \boldsymbol{Y}^{(n)}, \boldsymbol{Y}^{(n-1)}, \boldsymbol{X})_{s,t} \,. \tag{4.23}$$

[2] Note that if the function f is only defined on a subset of $\mathbb{R}^d$ then one would need to apply the extension theorem for Lipschitz functions to use this approach.

For smooth paths the iteration is defined by

$$
\begin{aligned}
dZ_t^{(n+1)} &= Z_u^{(n)}\, g(Y_u^{(n)}, Y_u^{(n-1)})\, dX_u\ , \\
dY_t^{(n+1)} &= dY_u^{(n)} + dZ_u^{(n)}\ , \\
dY_t^{(n)} &= dY_t^{(n)}\ , \\
dX_t &= dX_t\ ,
\end{aligned}
\tag{4.24}
$$

where $dZ_t^{(1)} = f(a)\, dX_t$, $Z_0^{(n)} = 0$, $Y^{(0)} \equiv a$, and $Y_0^{(n)} = a$. Now (4.24) defines a one-form; we can use this to extend the iteration, in the now obvious way, to functionals $(Z^{(n)}, Y^{(n)}, Y^{(n-1)}, X)_{s,t}$ in $\Omega G(W \oplus W \oplus W \oplus V)^p$. The iteration step makes sense because g (and hence the full one-form) is $\mathrm{Lip}[\gamma - 1]$.

It is obvious for smooth driving paths X and smooth initial estimates for the solution, that projection onto the last two co-ordinates gives the Picard iteration we studied in Step 3. The continuity of the iteration procedure makes it clear that this identity extends to geometric functionals.

We must prove that the sequence of iterations converge as a multiplicative functional to a functional $(\boldsymbol{I}, \boldsymbol{Y}, \boldsymbol{Y}, \boldsymbol{X})$, the continuity will then show that this is a fixed point for the equation. The argument will rely on a careful exploitation of the homogeneity of the various components.

Step 6. **The conditioning.** The first step is to rescale the coordinates and condition the problem.

For any choice of $\beta > 1$, and $\varepsilon < 1$ there is a choice of ω (depending on both parameters) so that if

$$
\boldsymbol{U}^{(0)} = (\tilde{\boldsymbol{Z}}^{(1)}, \boldsymbol{Y}^{(1)}, \boldsymbol{Y}^{(0)}, \varepsilon^{-1}\boldsymbol{X})
\tag{4.25}
$$

where $\tilde{\boldsymbol{Z}}^{(1)} = \beta \boldsymbol{Z}^{(1)}$, then $\boldsymbol{U}^{(0)}$ is controlled by ω.

We now use our estimates to study what happens when we replace the top line in (4.24) by

$$
\tilde{\boldsymbol{Z}}_t^{(n+1)} = \varepsilon\, \beta \int_{0<u<t} \tilde{\boldsymbol{Z}}_u^{(n)}\, g(Y_u^{(n)}, Y_u^{(n-1)})\, \varepsilon^{-1}\, d\boldsymbol{X}_u
$$

and use the new one-form to define a changed recursion involving $\tilde{\boldsymbol{Z}}^{(n)} = \beta^n \boldsymbol{Z}^{(n)}$ etc. In other words we recursively define

$$
\boldsymbol{U}^{(n-1)} = (\tilde{\boldsymbol{Z}}^{(n)}, \boldsymbol{Y}^{(n)}, \boldsymbol{Y}^{(n-1)}, \varepsilon^{-1}\boldsymbol{X})\,.
\tag{4.26}
$$

We will prove by induction that, for any choice of $\beta > 1$ there is a suitably small choice of $\varepsilon < 1$, chosen to depend on $K(M, p, \gamma)^{-[p]/p}$ and β alone, so that the sequence of elements in the sequence $\boldsymbol{U}^{n-1}$ are uniformly controlled by our ω on our predetermined time interval. By rescaling, it will be clear that the increments in the original iteration converge to zero with a geometric rate giving the overall result.

Step 7. **The induction step.** First fix the time interval so that $\omega < 1$ and assume that

$$\varepsilon < K_1(M, p, \gamma)^{-[p]/p}, \tag{4.27}$$

where M will be chosen later, but only depends on the Lip norms of various one forms and will be independent of other parameters in this problem.

We assume as our induction hypotheses that $\boldsymbol{U}^{n-1}$ is controlled by ω. Consider the form we must integrate to go from

$$\boldsymbol{U}^{n-1} \qquad \text{to} \qquad ((\varepsilon\beta)^{-1}\tilde{\boldsymbol{Z}}^{n+1}, \boldsymbol{Y}^{n+1}, \boldsymbol{Y}^{n}, \varepsilon^{-1}\boldsymbol{X}) : \tag{4.28}$$

$$\begin{aligned} d(\varepsilon\,\beta)^{-1}\,\tilde{\boldsymbol{Z}}_t^{(n+1)} &= \tilde{\boldsymbol{Z}}_u^{(n)}\, g(Y_u^{(n)}, Y_u^{(n-1)})\, d\varepsilon^{-1}\boldsymbol{X}_u\,, \\ d\boldsymbol{Y}_t^{(n+1)} &= d\boldsymbol{Y}_u^{(n)} + \beta^{-n}\, d\tilde{\boldsymbol{Z}}_u^{(n)}\,, \\ d\boldsymbol{Y}_t^{(n)} &= d\boldsymbol{Y}_t^{(n)}\,, \\ d\varepsilon^{-1}\boldsymbol{X}_t &= d\varepsilon^{-1}\boldsymbol{X}_t\,. \end{aligned} \tag{4.29}$$

Although examination of the second line in the expression shows this form varies with n the effect of increasing n is to decrease the Lipschitz norm. Hence, and because g is $\mathrm{Lip}[\gamma - 1]$ there is a uniform bound M on the $\mathrm{Lip}[\gamma - 1]$ norms of the forms on the range of paths under $\boldsymbol{U}^{n-1}$. (Recall that the $\boldsymbol{U}^{n-1}$ are controlled by ω and this in turn is uniformly bounded by one).

Hence there exists $K(M, p, \gamma)$, independent of our particular multiplicative functionals, time interval, etc., so that

$$((\varepsilon\beta)^{-1}\tilde{\boldsymbol{Z}}^{n+1}, \boldsymbol{Y}^{n+1}, \boldsymbol{Y}^{n}, \varepsilon^{-1}\boldsymbol{X}) \tag{4.30}$$

is of finite p-variation controlled by $K(M, p, \gamma)\,\omega$. By Step 3 we observe that providing $\varepsilon < K_1(M, p, \gamma)^{-[p]/p}$ then $(Y^{(n)}, Y^{(n-1)}, \varepsilon^{-1}X)$ is controlled by ω on any interval where $\omega < 1$ without any sort of factor.

Therefore we can apply the rescaling lemma again. Choose ε so that $\beta\,\varepsilon < K\,(M,p,\gamma)^{-[p]/p}$ and $\varepsilon < K_1\,(M,p,\gamma)^{-[p]/p}$. Then

$$(\tilde{\boldsymbol{Z}}^{n+1}, \boldsymbol{Y}^{n+1}, \boldsymbol{Y}^{n}, \varepsilon^{-1}\boldsymbol{X}) \tag{4.31}$$

is also controlled by ω, without a constant. This establishes the induction step.

Step 8. **Convergence.** At the level of paths it is now trivial that we have convergence. Let

$$(\tilde{Z}^{n+1}, Y^{n+1}, Y^{n}, \varepsilon^{-1}X)$$

be the path under

$$(\tilde{\boldsymbol{Z}}^{n+1}, \boldsymbol{Y}^{n+1}, \boldsymbol{Y}^{n}, \varepsilon^{-1}\boldsymbol{X})$$

satisfying the initial condition

$$(\tilde{Z}_0^{n+1}, Y_0^{n+1}, Y_0^{n}, \varepsilon^{-1}X_0) = (0, a, a, 0)\,.$$

Then it is clear that for smooth paths, and by continuity, for elements of ΩG^p (and geometric multiplicative functionals are all that one will ever see) the algebraic identity

$$Y_t^{(n+1)} = Y_u^{(n)} + \beta^{-n}\,\tilde{Z}_u^{(n)} \tag{4.32}$$

holds. But $\beta > 1$ and we have just proved that the difference process $Z_t^{(n)}$ is bounded independently of n on our time interval and so we have uniform convergence. The convergence is in p-variation, and as the sequence $Y_t^{(n)}$, $Y_u^{(n+1)}$ is uniformly bounded in p-variation norm that bound goes over to the limit.

However, our real objective is not just to construct a path in W and call it the solution, we want to construct a multiplicative functional. In other words we want to show that the multiplicative functionals $(\boldsymbol{Y}^{(n)}, \boldsymbol{X})$ converge in $\Omega G(W \oplus V)^p$. This is essentially trivial as well. Consider the projection $(\tilde{Z}^{(n)}, Y^{(n)}, \varepsilon^{-1}X)$ of $\boldsymbol{U}^{(n-1)}$ and $(Y^{(n+1)}, \varepsilon^{-1}X)$ of $\boldsymbol{U}^{(n-1)}$. Let Π_n be the linear map $(z,\ y,\ x) \longrightarrow (\beta^{-n}z + y, x)$ then the induced map Π_n on the tensor algebra takes $(\tilde{\boldsymbol{Z}}^{(n)}, \boldsymbol{Y}^{(n)}, \varepsilon^{-1}\boldsymbol{X})$ to $(\boldsymbol{Y}^{(n+1)}, \varepsilon^{-1}\boldsymbol{X})$ (again this is obvious for smooth sequences, and algebraic identities hold on closed sets, and hence extend to geometric functionals). But now the convergence is clear. and uniformly controlled by the β. The uniform nature of the estimates here on the convergence of Picard iteration prove the Itô map is continuous since our earlier arguments demonstrate that the finite iterations are certainly continuous.

4.2. Uniqueness.

To see uniqueness is also relatively straightforward and we do not dwell on it. We did not need to start our new Picard iteration with the function that was constant at a and its integral. We could have started it at two of our "solutions", in this case our iteration would have compared the difference and shown that it went to zero.

Acknowledgements. This paper has not been easy for the author to write. So his first acknowledgement must go to those, like Leonard Gross, David Elworthy and Bruce Driver (and of course his family) who showed interest and provided encouragement on the way, and were silent and uncritical when the final mss with the details took well over a year to put together. I hope that the content in some small part makes up for the delay.

In addition, I owe an enormous debt to the students who attended a graduate course in the fall of 1994 where the details of the proofs were first presented in a public manner, particularly to Marc Joannides and Zhongmin Qian for taking good notes which could be typed up as rough drafts from which I prepared this mss. In addition Zhongmin Qian read an intermediate draft very carefully and his contribution has significantly improved the exposition (although of course there will still be plenty of errors and these are entirely my own resposibility). David R. E. Williams did a fantastic and fairly massive job in the final days converting the mss into a suitable form of TEX for submission.

I would also recall that it was in an open and ongoing discussion during 1992/3 with my then students Jessica Gaines and Eva Sippiläinen, that we all assimilated the Lie algebraic perspective of Chen on ordinay differential equations, which of course forms a quite essential foundation for the present paper.

Finally I acknowledge research funding from EPSRC and EEC. In particular, the support of the SERC via senior fellowship B/93/sf/445 and grants YYYY SERC 908 GR/J 55946, EEC grants SC1-784 and SC1-0062.

References.

[1] Burdzy, K., Variation of iterated Brownian motion. In: Workshop and Conference on Measure-valued Processes, Stochastic Partial Differential equations and Interacting Systems, (CRM Proceedings and Lecture Notes, 1993).

[2] Bachelier, L., *Calcul des probabilités.* Reprint of the 1912 original: Les Grands Classiques Gauthier-Villars, 1992, Editions Jacques Gabay, Sceaux, 1992.

[3] Ben Arous, G., Flots et Séries de Taylor Stochastique. *Probab. Theor. Relat. Fields* **81** (1989), 29-77.

[4] Castell, F., Gaines, J. G., The ordinary differential equation approach to asymptotically efficient schemes for solution of stochastic differential equations. *Ann. Inst. H. Poincaré* **32** (1996), 231-250.

[5] Chen, K.-T., Integration of Paths, Geometric Invariants and a Generalized Baker-Hausdorff Formula. *Ann. of Math.* **65** (1957), 163-178.

[6] Föllmer, H., Calcul d'Itô sans probabilités. *Seminar on Probability, XV.* Lecture Notes in Math. **850**, 143-150, Springer, 1981.

[7] Föllmer, H., Dirichlet processes, Stochastic integrals. Proceedings of the LMS Durham Symposium held at the University of Durham, Durham, July 7-17, 1980. Edited by David Williams. Lecture Notes in Math. **851** 476-478, Springer, 1981.

[8] Gaines, J. G., Lyons, T. J., Variable Step Size Control in the Numerical Solution of Stochastic Differential Equations. *SIAM J. Appl. Math.* **51** (1997), 1455-1484.

[9] Lévy, P., *Théorie de l'Addition des Variables Aléatoires.* Gauthier-Villars, 1937.

[10] Hambly, B. M., Lyons, T. J., Stochastic Area for Brownian motion on the Sierpinski gasket. *Ann. Probab.* **26** (1998), 132-148.

[11] Ikeda, N., Watanabe, S., *Stochastic Differential Equations and Diffusion Processes.* North-Holland, 1989.

[12] Itô, K., Stochastic Integral. *Proc. Imperial Acad. Tokyo* **20** (1944), 519-524.

[13] Lévy, P., *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, 1948.

[14] Lyons, T. J., *The Interpretation and Solution of Ordinary Differential Equations Driven by Rough Signals.* Proceedings of Symposia in Pure Mathematics. **57** (1995).

[15] Lyons, T., Differential equations driven by rough signals (I): an extension of an inequality of L. C. Young. *Mathematical Research Letters* **1** (1994), 451-464.

[16] Lyons, T., On the nonexistence of path integrals. *Proc. Roy. Soc. London Ser. A* **432** (1991), 281-290.

[17] Lyons, T. J., Qian, Z. M., Calculus of variation for multiplicative functionals. In: *New Trends in Stochastic Analysis.* World Scientific. Ed. Elworthy. (1997), 348-374.

[18] Lyons, T. J., Qian, Z. M., Flow of diffeomorphisms induced by a geometric multiplicative functional. To appear in *Probab. Theor. Relat. Fields.*

[19] Lyons, T. J., Qian, Z. M., Calculus for multiplicative functionals, Itô's formula, and differential equations. In: *Itô's Stochastic Calculus.* Ed. Ikeda. Springer. (1996), 233-250.

[20] Malliavin, P., Infinite-dimensional analysis. *Bull. Sci. Math.* **117** (1993), 63-90.

[21] Protter, P., On the existence, uniqueness, convergence and explosions of solutions of systems of stochastic differential equations. *Ann. Probab.* **5** (1977), 243-261.

[22] Reutenauer, C., *Free Lie Algebras.* London Mathematical Society Monographs, New Series 7, Oxford Science Publications, 1993.

[23] Sipiläinen, E.-M., A pathwise view of solutions of stochastic differential equations. Ph.D. Thesis, University of Edinburgh, 1993.

[24] Stein, E. M., *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, 1970.

[25] Sugita, H., Various topologies in the Wiener space and Lévy's stochastic area. *Probab. Theor. Relat. Fields* **91** (1992), 283-296.

[26] Doss, H., Liens entre équations différentielles stochastiques et ordinaires. *Ann. Inst. H. Poincaré* **13** (1977), 99-125.

[27] Sussmann, H. J., On the gap between deterministic and stochastic ordinary differential equations. *Ann. Probab.* **6** (1978), 19-41.

[28] Kunita, H., On the representation of solutions of stochastic differential equations. *Seminar on Probability, XIV.* Lecture Notes in Math. **784**, 282-304, Springer, 1980.

[29] Yamato, Y., Stochastic differential equations and nilpotent Lie Algebras, *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **47** (1979), 213-229.

[30] Gershkovich, V., Vershik, A., Nonholonomic manifolds and nilpotent analysis. *J. Geom. Phys.* **5** (1988), 407-452.

[31] Williams, D. R. E., Solutions of differential equations driven by Càdlàg paths of finite p-variation $(1 \leq p < 2)$. Thesis, Imperial College, 1998.

[32] Young, L. C., An inequality of Hölder type, connected with Stieltjes integration. *Acta Math.* **67** (1936), 251-282.

Recibido: 17 de septiembre de 1.996

Terry J. Lyons
Department of Mathematics
Imperial College
Huxley Building, 180 Queen's Gate
London, SW7 2BZ, UNITED KINGDOM
t.lyons@ic.ac.uk