Discretely sampled signals and the rough Hoff process

Guy Flint, Ben Hambly, Terry Lyons

Mathematical Institute, University of Oxford, Woodstock Road, OX2 6GG, UK

February 26, 2014

Abstract

Sampling a $d$-dimensional continuous semimartingale $X : [0, 1] \to \mathbb{R}^d$ at times $D = (t_i)$, we construct a lead-lag path; to be precise, a piecewise-linear, axis-directed process $X^D : [0, 1] \to \mathbb{R}^{2d}$ comprised of a past and future component. We call such an object the Hoff process and when lifting it to its natural rough path enhancement, we consider the question of convergence as the latency of our sampling becomes finer. We find that the Itô integral can be recovered from a sequence of random ODEs driven by the components of $X^D$. This is in contrast to the usual Stratonovich integral limit suggested by the classical Wong-Zakai theorem [18].

Keywords. Rough path theory, lead-lag path, Hoff process, Itô-Stratonovich correction

1 Introduction

A common statistical method in analysing high frequency financial data is to take a time series view and consider the data as a sequence of ordered vectors. However, from this perspective one may lose some of the important structure which characterises useful features within the data; in particular, we may overlook the order of events in different coordinates and their consequent effects on one another. Moreover, time series analysis is typically highly dependent on the underlying model chosen.

By taking an alternative philosophical view of the discrete points being part of a continuous path in a multidimensional space (say some Euclidean space $\mathbb{R}^d$), one may lift the path to its signature and apply the powerful theory of rough paths in order to analyse the original signal. The idea that the signature can be thought of as an informative transform of a multidimensional time series was first introduced by B. Hambly and T. Lyons [9]. We use the term transform in the sense that no information is lost (modulo the notion of tree-like equivalence) by taking the signature of a bounded variation path.

With this view in mind, D. Levin et al in [8] first introduced the possibility of using the signature of a time series to understand financial data. In this work the authors exploit the fact that any polynomial of iterated integrals can be represented by a linear combination of iterated integrals [15, Theorem 2.15]. This makes the signature a natural candidate for linear regression analysis, which is precisely what the authors do with numerical examples using classical time series models. Their transform of the time series to a signature of a bounded variation path closely resembles our definition of a Hoff process below.

The recent paper [7] also takes the viewpoint of defining a process (which is precisely our definition of the Hoff process) out of the discrete data points and applies standard regression analysis to extract features out of WTI crude oil futures market data among other asset classes.
In this paper we assume that the underlying signal is a continuous semimartingale \( X : [0,1] \to \mathbb{R}^d \) with canonical decomposition \( X = M + V \), \( (M \text{ being a local martingale and } V \text{ being a bounded variation path}) \). Since the signature is invariant under translation we make the simplifying assumption that \( X_0 = 0 \). Let us establish some notation first:

**Definition 1.1 (Partition).** A partition is a collection of distinct points in \( [0,1] \) always containing the endpoints \( \{0,1\} \). Let \( \mathcal{D}_{[0,1]} \) denote the set of all partitions of \( [0,1] \) and we set \( |D| := \max_{i \in D} |t_{i+1} - t_i| \) to be the mesh of \( D \).

As our original time series set up, suppose we are given the values of \( X \) at times given by a partition \( D = \{(t_i)_{i=0}^n \in \mathcal{D}_{[0,1]} \} \) that is we have a sequence of ordered pairs \( \{(t_i, X(t_i))\}_{t_i \in D} \).

At this point we follow the cited papers \([7, 8]\) and depart from traditional time series analysis by constructing a \((2d)\)-dimensional process \( X^D : [0,1] \to \mathbb{R}^{2d} \) comprised of the lead and lag components of the data. We call this the Hoff process. Using the notation \( t_i^\ast := 1/2(t_i + t_{i+1}) \), we give the following precise definition:

**Definition 1.2 (Hoff process).** Given a time series \( (X_t)_{t_i \in D} \) we construct the continuous axis process \( X^D : [0,1] \to \mathbb{R}^{2d} \),

\[
X^D = (X^{D,\theta,i})_{t_i \in \{1, \ldots, d\}, \theta \in \{b,f\}} = (X^{D,b,1}, \ldots, X^{D,b,d}, X^{D,f,1}, \ldots, X^{D,f,d}),
\]

which we define as follows: for \( D = (t_i)_{i=0}^n \in \mathcal{D}_{[0,1]} \);

\[
X^D_u = \begin{cases} 
(X_{t_i - \frac{u-t_i}{t_{i+1}-t_i}} X_{t_i+1} - X_{t_i-1}); X_{t_i+1} & \text{if } u \in [t_i, t_i^\ast); \quad \text{for } i = 1, \ldots, n-1, \\
(X_{t_i}; X_{t_i+1} + \frac{u-t_i^\ast}{t_{i+1}-t_i^\ast} X_{t_{i+1}} - X_{t_i}) & \text{if } u \in [t_i^\ast, t_{i+1}); \quad \text{for } i = 0, 1, \ldots, n-2,
\end{cases}
\]

and set the first and last intervals of \( X^D \) to be:

\[
X^D_0 = \begin{cases} 
(0; \frac{u}{0} X_{t_i}) & \text{if } u \in [0, t_0^\ast); \\
(X_{t_{n-1}} + \frac{u-t_{n-1}}{t_n-t_{n-1}} X_{t_n} - X_{t_{n-1}}); X_{t_n}) & \text{if } u \in [t_{n-1}^\ast, t_n].
\end{cases}
\]

We call the process \( X^D \) the **Hoff process** associated with the time stamped series \( (X_t)_{t_i \in D} \). We denote the Lévy area process associated with \( X^D \) by

\[
A^D_{s,t} : = \frac{1}{2} \int_s^t \left( X^D_{s,r} dX^D_{r}\lambda_{s,r} - X^D_{s,r} dX^D_{r}\theta_{s,r} \right) = \int_s^t X^D_{s,r} dX^D_{r}\lambda_{s,r} - \frac{1}{2} X^D_{s,t} X^D_{s,t}. \tag{1}
\]

Since \( X^D \) is piecewise linear we are able to switch between the Itô and Stratonovich integrals in the Stokes’ area formula (1) without incurring an additional drift term.

**Remark 1.3.** To our knowledge, B. Hoff was the first to consider the area between rough paths and their delay within the context of moving Brownian frames in his Ph.D. thesis [10]. Thus for this reason we use the nomenclature of a Hoff process for \( X^D \).

We call \( X^{D,b} \) and \( X^{D,f} \) the **lag** and **lead** components of the Hoff process respectively. Since lag and lead both share a common first letter we use \( b \) and \( f \) instead, which can be thought of as backward and forward respectively. At this point we find it helpful in building intuition about \( X^D \) to include a diagram of a possible trajectory (see Figure 1).

The first result of this paper concerns the convergence of the rough path lift \( X^D \) of the Hoff process \( X^D \) as our partition mesh \( |D| \to 0 \). The second result is a corollary of the first; we prove that certain natural random ODEs driven by \( X^D \) converge to the corresponding Itô integral limit and not the usual Stratonovich stochastic integral as predicted by classical Wong-Zakai
The theory of rough paths allows us to prove otherwise and refer to [6, Theorem 17.20] and [4] for similar results concerning lead-lag driven random ODE convergence.

Interest in recovering the Itô integral and the usual Itô formula in a rough path context is an active area of research. From a much more theoretical perspective, D. Yang and T. Lyons in their recent paper [17] proved that the Itô integral can be recovered in a pathwise sense. Their proof introduces the novel idea of concatenating a mean of Stratonovich solutions to add as a “polluting” noise to the driving martingale signal. The perturbed signal allows the precise recovery of the Itô integral, which is the main idea behind the proof of the recovery theorem in the present paper.

Throughout we denote the class of $L^p(P)$-bounded martingales by $\mathcal{M}^p$; that is

$$\|M\|_p := \sup_{t \in [0, 1]} |M_t|_{L^p(P)} = \mathbb{E} \left( \sup_{t \in [0, 1]} |M_t|^p \right)^{1/p} < \infty.$$  

We have attempted to make the paper as self-contained as possible. Before presenting the main results of the paper in Sections 3 and 4, we establish the necessary rough path theory and notation in the next section. Particular emphasis is placed on rough differential equations.

The remainder of the paper deals with establishing auxiliary regularity results required for the proof of the main convergence theorem; in particular, the pointwise convergence of the rough Hoff process lift and establishing corresponding maximal $p$-variation bounds. We also prove a tightness property on a given collection of rough Hoff processes.

**Remark 1.4.** Throughout the paper, $C_1, C_2, \ldots$ denote various deterministic constants (which may vary from line to line). Their dependence on $\|M\|_p$ and other similar quantities will be omitted for brevity within proofs but made explicit within theorem statements.
2 Rough path concepts and notation

In this section we provide a tailored overview of relevant rough path theory and take the opportunity to establish the notation we will need. For a more detailed overview of the theory we direct the reader to [3, 6, 13, 16, 14, 15] among many others.

Rough path analysis introduced by T. Lyons in the seminal article [14] provides a method of constructing solutions to differential equations driven by paths that are not of bounded variation but have controlled roughness. The measure of this roughness is given by the $p$-variation of the path (see (2) below).

2.1 Rough path overview

Since this paper deals with continuous $\mathbb{R}^d$-valued (and $\mathbb{R}^{2d}$) paths on $[0, 1]$ we restrict this brief overview to the finite dimensional case (mainly adopting the notation found in [6]). For the extension to rough paths over infinite dimensional Banach spaces we refer to [16]. We denote the space of such functions by $C([0, 1], \mathbb{R}^d)$. We write $x_{s,t} = x_t - x_s$ as a shorthand for the increments of the path when $x \in C([0, 1], \mathbb{R}^d)$. For $p \geq 1$ we define the following metrics:

$$|x|_\infty := \sup_{t \in [0,1]} |x_t|, \quad |x|_{p\text{-var}:[0,1]} := \left( \sup_{D=\{(t_j)\} \in D([0,1])} \sum_{t_j \in D} |x_{t_j, t_{j+1}}|^p \right)^{1/p},$$

which we call the uniform norm and $p$-variation semi-norm respectively. We denote by $C^{p\text{-var}}([0, 1], \mathbb{R}^d)$ the linear subspace of $C([0, 1], \mathbb{R}^d)$ consisting of paths of finite $p$-variation.

In the case of $x \in C^{2\text{-var}}([0, 1], \mathbb{R}^d)$ where $p \in [1, 2)$, the iterated integrals of $x$ are canonically defined via Young integration [19]. The collection of all these iterated integrals as an object in itself is called the signature of the path, given by

$$S(x)_{s,t} := 1 + \sum_{k=1}^{\infty} \int_{s < t_1 < t_2 < \ldots < t_k < t} dx_{t_1} \otimes dx_{t_2} \otimes \ldots \otimes dx_{t_k} \in \bigoplus_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k},$$

for all $(s, t) \in \Delta_{[0,1]} := \{(s, t) : 0 \leq s \leq t \leq 1\}$. With the convention that $(\mathbb{R}^d)^{\otimes 0} := \mathbb{R}$, we define the tensor algebras:

$$T^\infty(\mathbb{R}^d) := \bigoplus_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k}, \quad T^N(\mathbb{R}^d) := \bigoplus_{k=0}^{N} (\mathbb{R}^d)^{\otimes k}.$$

Thus the signature takes values in $T^\infty(\mathbb{R}^d)$. Defining the canonical projection mappings $\pi_N : T^\infty(\mathbb{R}^d) \to T^N(\mathbb{R}^d)$, we can also consider the truncated signature:

$$S_N(x)_{s,t} := \pi_N(S(x)_{s,t}) := 1 + \sum_{k=1}^{N} \int_{s < t_1 < t_2 < \ldots < t_k < t} dx_{t_1} \otimes dx_{t_2} \otimes \ldots \otimes dx_{t_k} \in T^N(\mathbb{R}^d),$$

and thus view $S_N$ as a continuous mapping from $\Delta_{[0,1]}$ into $T^N(\mathbb{R}^d)$. Throughout this paper we will also reserve the notation $\pi_N$ for the canonical projection of $T^M(\mathbb{R}^d)$ to $T^N(\mathbb{R}^d)$ when $M > N$.

It is a well-known fact that the path $S_N(x)$ takes values in the step-$N$ free nilpotent Lie group with $d$ generators, which we denote by $G^N(\mathbb{R}^d)$. Indeed, defining the free nilpotent step-$N$ Lie algebra $g^N(\mathbb{R}^d)$ by

$$g^N(\mathbb{R}^d) := \left[ \mathbb{R}^d, \ldots, \left[ \mathbb{R}^d, \mathbb{R}^d \right] \right] \left( N-1 \right) \text{brackets},$$
and the natural exponential map \( \exp_N : T^N(\mathbb{R}^d) \to T^N(\mathbb{R}^d) \) by

\[
\exp_N(a) = 1 + \sum_{k=1}^N \frac{a^k}{k!},
\]

we define \( G^N(\mathbb{R}^d) := \exp_N(p^N(\mathbb{R}^d)) \). The following characterization establishes the well-known fact (a proof can be found in [6, Theorem 7.30]).

**Theorem 2.1 (Chow’s Theorem).** We have

\[
G^N(\mathbb{R}^d) = \left\{ S_N(x)_{0,1} : x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d) \right\}.
\]

More generally, if \( p \geq 1 \) we can consider the set of such group-valued paths

\[
x_t = (1, x_t^1, \ldots, x_t^{|p|}) \in G^{|p|}(\mathbb{R}^d).
\]

Importantly, the group structure provides a natural notion of increment for the signature, namely that \( x_{s,t} := x_t^{-1} \otimes x_t \).

**Example 2.2.** Take a path \( x \in C^{1\text{-var}}([0, 1], \mathbb{R}^d) \) and set \( x := S_2(x) \in C^{0,1,G^2(\mathbb{R}^d)} \) for its level-2 signature, which we call the (level-2) rough path lift of \( x \). An exercise in algebra and calculus confirms that

\[
x_{s,t} = 1 + x_{s,t} + \frac{1}{2}x_{s,t} \otimes x_{s,t} + A_{s,t} = \exp_2(x_{s,t} + A_{s,t}),
\]

where \( A : \Delta_{0,1} \to [\mathbb{R}^d, \mathbb{R}^d] \) is the Lévy area process of \( x \). Note that we have used the truncated exponential notation defined above.

We can describe the set of norms on \( G^{|p|}(\mathbb{R}^d) \) which are homogeneous with respect to the natural dilation operation on the tensor algebra (see [6] for more definitions and details). The subset of these so-called homogeneous norms which are symmetric and sub-additive gives rise to genuine metrics on \( G^{|p|}(\mathbb{R}^d) \). We work with the Carnot-Carathéodory norm on \( G^N(\mathbb{R}^d) \) given by

\[
\|g\| := \inf \left\{ \int_0^1 |d\gamma| : \gamma \in C^{1\text{-var}}([0, 1], \mathbb{R}^d) \text{ and } S_N(\gamma)_{0,1} = g \right\},
\]

which is well-defined by Chow’s theorem. This in turn gives rise to the notion of a homogeneous metric on \( G^N(\mathbb{R}^d) \):

\[
d(g, h) = \|g^{-1} \otimes h\|.
\]

Moreover, this metric gives rise to the following \( p \)-variation and \( \alpha \)-Hölder metrics on the set of \( G^{|p|}(\mathbb{R}^d) \)-valued paths:

\[
d_{p\text{-var},[0,1]}(x, y) := \left( \sup_{D=(t_j)_{j \in \mathbb{N}}} \sum_{j \in D} d(x_{t_j, t_{j+1}}, y_{t_j, t_{j+1}})^p \right)^{1/p},
\]

\[
d_{\alpha\text{-Höld},[0,1]}(X, Y) := \sup_{0 \leq s < t \leq 1} \frac{d(x, y)}{|t - s|^{1/p}}.
\]

We also define the metrics

\[
d_0([0,1])(x, y) := \sup_{0 \leq s < t \leq 1} d(x_{s,t}, y_{s,t}), \quad d_{\infty,[0,1]}(x, y) := \sup_{t \in [0,1]} d(x_t, y_t).
\]

If no confusion may arise we will often drop the \([0, 1]\) appearance in these metrics. If (2) is finite, then \( \omega(s, t) := \|x\|^p_{p\text{-var},[s,t]} \) is a control in the following sense:
Definition 2.3. A mapping \( \omega : \Delta_{[0,1]} \to [0,\infty) \) is a control if it is continuous, bounded, vanishes on the diagonal and is super-additive; that is, for all \( s < t < u \) in \([0,1]\):

\[
\omega(s,t) + \omega(t,u) \leq \omega(s,u).
\]

The space of weakly geometric \( p \)-rough paths, denoted by \( W\Omega_p(\mathbb{R}^d) \), is the set of paths with values in \( G^{[p]}(\mathbb{R}^d) \) such that (2) is finite. A refinement of this notion is the space of geometric \( p \)-rough paths, denoted \( G\Omega_p(\mathbb{R}^d) \), which is the closure of

\[
\{ S_p(x) : x \in C^{1\text{-var}}([0,1],\mathbb{R}^d) \},
\]

with respect to the topology induced by the rough path metric \( d_{p\text{-var};[0,1]} \). Certainly we have the inclusion \( G\Omega_p(\mathbb{R}^d) \subset W\Omega_p(\mathbb{R}^d) \) and it turns out that this inclusion is strict (see [15, §3.2.2]).

This paper is concerned with semimartingales (which almost surely have finite \( p \)-variation for all \( p \in (2,3) \) [6, Theorem 14.9]). Thus \( |p| = 2 \) and so we are dealing with \( p \)-rough paths in the step-2 group \( G^2(\mathbb{R}^d) \). Given a stochastic process \( X \) we denote its corresponding rough path lift as \( X \) (as opposed to the rough path lift of \( x \) denoted by \( x \)).

2.2 Rough differential equations

In this subsection we introduce rough differential and integral equations. For the more technical topic of rough differential equations with drift we refer to the exhaustive [6, Chapter 12]. The theory quoted here can be found in greater detail in [6, Chapter 10].

For now let \( x \in C^{1\text{-var}}([0,1],\mathbb{R}^d) \). We denote the solution \( y \) of the (controlled) ordinary differential equation (ODE)

\[
dy = V(y) \, dx := \sum_{i=1}^{d} V_i(y) \, dx^i, \quad y_0 \in \mathbb{R}^c,
\]

by \( \pi(V) (0; y_0, x) \). The notation \( \pi(V) (s, y_s; x) \) stands for solutions of (3) started at time \( s \) from a point \( y_s \in \mathbb{R}^c \).

Definition 2.4 (RDE). Let \( x \in W\Omega_p(\mathbb{R}^d) \) for some \( p \geq 1 \). We say that \( y \in C([0,1],\mathbb{R}^c) \) is a solution to the rough differential equation (shorthand: RDE solution) driven by \( x \) along the collection of \( \mathbb{R}^c \)-vector fields \( V = (V_i)_{i=1,\ldots,d} \) and started at \( y_0 \) if there exists a sequence \( (x_n)_n \) in \( C^{1\text{-var}}([0,1],\mathbb{R}^d) \) such that:

(i) \( \lim_{n \to \infty} d_{0,[0,1]}(S_p(x_n), x) = 0 \);

(ii) \( \sup_n \| S_p(x_n) \|_{p\text{-var};[0,1]} < \infty \);

and ordinary differential equations (ODE) solutions \( y_n = \pi(V) (0, y_0; x_n) \) such that

\( y_n \to y \) uniformly on \([0,1]\) as \( n \to \infty \).

We denote this situation with the (formal) equation:

\[
dy = V(y) \, dx, \quad y_0 \in \mathbb{R}^c,
\]

which we refer to as a rough differential equation.

Remark 2.5. The RDE solution map

\[
x \in C^{p\text{-var}}([0,1], G^{[p]}(\mathbb{R}^d)) \mapsto \pi(V) (0, y_0; x) \in C \left( [0,1], G^{[p]}(\mathbb{R}^c) \right)
\]

is known in the rough path literature as the \textbf{Itô-Lyons map}. 

6
Armed with the definition of a rough differential equation, it is natural to ask what a rough integration definition would be. To this end we follow Lyons’ original approach in \cite{14}. To be precise, we wish to make sense of the (formal) integral equation
\[
\int_0^1 \varphi(z) \, dz \text{ with } z = \pi_1(x),
\]
where \(z \in W^{1,1}[0,1] \). Certainly in the classical case of \(z = S_{[\xi]}(z)\) for some \(z \in C^{1,\text{var}}([0,1], \mathbb{R}^d)\), we want \(4\) to coincide with \(S_{[\xi]}(\xi)\), where \(\xi\) is the classical Riemann-Stieltjes integral \(\int_0^1 \varphi(z) \, dz\). We take the following definition straight from \cite[Definition 10.44]{6}:

**Definition 2.6 (Rough integration).** Let \(x \in W^{1,1}[0,1] \) and \(\varphi = (\varphi_i)_{i=1}^d\) be a collection of maps from \(\mathbb{R}^d\) to \(\mathbb{R}^d\). We say that \(y \in C \left([0,1], G^{[\gamma\text{-var}]}(\mathbb{R}^d)\right)\) is a rough path integral of \(\varphi\) along \(x\) if there exists a sequence \((x^n) \subset C^{1,\text{var}}([0,1], \mathbb{R}^d)\) such that:

1. \(x_0^n = \pi_1(x_0)\) for all \(n\);
2. \(\lim_{n \to \infty} d_{[0,1]} \left(S_{[\xi]}(x^n), x\right) = 0\);
3. \(\sup_n \|S_{[\xi]}(x^n)\|_{\text{p-var}, [0,1]} < \infty\);
4. \(\lim_{n \to \infty} d_{\infty} \left(S_{[\xi]} \left(\int_0^1 \varphi(x^n) \, dx^n\right), y\right) = 0\).

We write \(\int \varphi(x) \, dx\) for the set of rough path integrals of \(\varphi\) along \(x\) (without additional regularity assumptions there could be more than one suitable candidate).

From RDE solutions we can construct full RDE solutions; in particular, we construct a “full” solution as a weak geometric \(p\)-rough path in its own right. This in turn will allow us to use a solution to a first RDE as the driving signal for a second RDE. Correspondingly, RDE solutions can then be used as “integrators” in rough integrals. This has more practical implications than mere aesthetics of the theory. In Lyons’ original work \cite{14} the existence and uniqueness of RDEs was established via Picard iteration which critically relied on the notion of “full” RDE solutions.

**Definition 2.7 (Full RDEs).** As usual, let \(x \in W^{1,1}[0,1] \) for some \(p \geq 1\). We define \(y \in C \left([0,1], G^{[\gamma\text{-var}]}(\mathbb{R}^d)\right)\) to be the solution of the full RDE driven by \(x\) along the vector fields \(V = (V_i)_{i=1}^d\) and started at \(y_0 \in C^{1,\text{var}}(\mathbb{R}^d)\) if there exists a sequence \((x^n) \subset C^{1,\text{var}}([0,1], \mathbb{R}^d)\) such that (ii) and (iii) hold, and ODE solutions \(y_n \in \pi_1(V) (0, \pi_1(y_0); x^n)\) such that \(y_0 \odot S_{[\xi]}(y^n)\) converges uniformly to \(y\) as \(n \to \infty\).

We write the corresponding (formal) equation as \(dy = V(y) \, dx\).

**Remark 2.8.** A note on regularity: consider RDEs driven by a path \(x\) in \(W^{1,1}[0,1] \) along a collection of vector fields \(V = (V^i)_{i=1}^d\) on \(\mathbb{R}^d\). Considerations of existence and uniqueness of such RDEs show that the appropriate way to measure the regularity of the collection \(V\) turns out to be notion of \(\gamma\)-Lipschitzness (denoted \(\text{Lip}^\gamma\)) in the sense of Stein. See \cite{6, 15} to note the contrast with the classical notion of Lipschitzness.

3 Rough Hoff process convergence

The next theorem presents the first main result of the paper, namely the convergence of the rough path lift \(X^0\) of the Hoff process \(X^0\) under the \(p\)-variation topology. Define the process \(\bar{X}_t = (X_1, X_t)\) and use the standard notation of \(\langle M \rangle := \left(\langle M^i, M^j \rangle : i, j = 1, \ldots, d\right)\) for the quadratic variation of \(M\) as a \(\mathbb{R}^{d \times d}\)-valued process.
Theorem 3.1 (Rough Hoff process convergence). Let \( X = M + V : [0, 1] \to \mathbb{R}^d \) be a continuous semimartingale with associated Hoff process \( X^D \) for a given partition \( D \in \mathcal{D}_{[0,1]} \). Then for all \( p > 2 \) we have

\[
d_{p,\text{var};[0,1]}(X^D, X^\infty) \to 0 \text{ in probability as } |D| \to 0,
\]

where \( X^\infty \in C([0,1], C^2(\mathbb{R}^d)) \) is defined by

\[
X^\infty_{s,t} = \exp_2 \left( \bar{X}_{s,t} + \begin{bmatrix} A^1_{s,t} & A^2_{s,t} - 1/2(X)_{s,t} \\ A^2_{s,t} + 1/2(X)_{s,t} & A^1_{s,t} \end{bmatrix} \right).
\]

If in addition the finite variation process \( V \) is bounded in \( L^q(\mathbb{P}) \) (\( q \geq 1 \)) and \( M \in \mathcal{M}^p \), then the convergence also holds in \( L^{p\wedge n}(\mathbb{P}) \).

Proof. First we prove the theorem for \( V = 0 \). Assuming \( M \in \mathcal{M}^p \) it follows from Proposition 6.1 that the rough path collection \( \{M^D : D \in \mathcal{D}_{[0,1]}\} \) satisfies uniform \( p \)-variation bounds in \( L^p(\mathbb{P}) \). Since \( |D| \to 0 \), the collection also satisfies the tightness property of Corollary 7.2. Thus the second convergence statement follows from an application of [3, Corollary 50]. In the case that \( M \notin \mathcal{M}^p \) we obtain convergence in probability by a simple localization argument (as done in [6, Theorem 14.16]).

We now turn to the general case where \( V \neq 0 \) and define \( \bar{V} : [0,1] \to \mathbb{R}^d \) by \( \bar{V}_t = (V_t, V_t) \). In contrast to the martingale rough path convergence, the corresponding convergence proof for the finite variation process is straightforward:

Lemma 3.2. For any \( \delta > 0 \):

\[
d_{1+\delta,\text{var};[0,1]}(V^D, \bar{V}) \to 0 \text{ in probability as } |D| \to 0.
\]

If in addition \( V \) is bounded in \( L^q(\mathbb{P}) \) for some \( q \geq 1 \) then the convergence also holds in \( L^q(\mathbb{P}) \).

Proof. The first convergence follows readily from [6, Proposition 1.28] and interpolation (see [6, §14.1] for details). The stronger convergence in \( L^q(\mathbb{P}) \) is a consequence of the uniform \( L^q(\mathbb{P}) \)-bound assumption [1, Theorem 4.14].

Returning to the proof of Theorem 3.1, it is a straight-forward exercise in algebra to confirm that \( X^\infty_{s,t} = T_{\bar{V}}(X^\infty_{s,t}) \) using the translation operator defined in [6, §9.4.6]. The theorem then follows immediately from the continuity of the translation operator [6, Corollary 9.35].

Remark 3.3. The quadratic variation of \( M \) naturally appears as we consider the limiting area of the rough path lift \( X^D \). This non-linear phenomena was first observed by Hoff in [10].

Remark 3.4. A natural question to ask is whether the convergence result of Theorem 3.1 will hold using the stronger \( \alpha \)-Hölder variation metric (\( \alpha < 1/2 \), rather than the \( p \)-variation metric. The answer is no in general; a consequence of the behaviour of the Hoff process \( X^D \) over consecutive intervals. In particular, over a given interval \([t_i, t_{i+1}] \subset D \in \mathcal{D}_{[0,1]}\), \( X^D \) is defined by the values of \( X \) at the present, previous and future times: \( \{X_{t_{i-1}}, X_{t_i}, X_{t_{i+1}}, X_{t_{i+2}}\} \).

Thus for particular sequences of partitions \( (D_n)_{n \in \mathbb{N}} \) with \( |D_n| \to 0 \) but

\[
\sup_{n \in \mathbb{N}} |D_n| = \infty,
\]

(5)

(where \( D_n \) has \( n \) non-zero points), the resulting Hoff process sequence may not be uniformly bounded in Hölder norm (under expectation in \( L^p(\mathbb{P}) \)). Thus we cannot expect its rough path lift \( X^D \) to converge to \( X^\infty \) in \( \alpha \)-Hölder norm (\( \alpha < 1/2 \)), especially since the latter process can be shown to satisfy \( \mathbb{E} \left[ \|X^\infty\|_{\alpha, \text{Hölder};[0,1]}^p \right] < \infty \) for all \( p \geq 1 \).
It remains to give an example where (5) holds and the Hölder-norm of \( X^{D_n} \) explodes as \( n \to \infty \). To this end suppose that \( X \) is just a Brownian motion. Denoting the \( n \)th level dyadic partition by \( D_n = (t_k^n = k/2^n)_{k=0}^n \), we set
\[
D_n := \left( [0, 1/2] \cap D^{2^n} \right) \cup ([1/2, 1] \cap D_n).
\]
We find that for \( \alpha < 1/2 \):
\[
\mathbb{E} \left[ \left\| X^{D_n} \right\|_{\alpha; \text{Hoël}([0,1])}^p \right] \geq \mathbb{E} \left[ \left| X_{1/2 + 2^{-n}} - X_{1/2} \right|^p \right] \to \infty,
\]
as \( n \to \infty \). It is well known in rough path circles that in order to use the Hölder metric on discrete rough path sequences, one must insist that the quantity (5) is finite (see the recent paper [11]).

4 Recovery of the Itô integral

We now consider the second main result of the paper:

**Theorem 4.1 (Recovery of the Itô integral).** Let \( X : [0, 1] \to \mathbb{R}^d \) be a continuous semimartingale with canonical decomposition \( X = M + V \). Let \( f = (f_i)^d_{i=1} \) be a collection of \( \text{Lip}^q(\mathbb{R}^d, \mathbb{R}^e) \) mappings where \( \gamma > 2 \). For each \( n \) let \( Y^n \) be the solution to the random ODE:
\[
dY^n_t = f \left( X^n_{t-} \right) \, dX^n_t = \sum_{i=1}^d f_i \left( X^n_{t-} \right) \, dX^i_t, \quad Y^n_0 = y_0 \in \mathbb{R}^e,
\]
and let \( Y \) be the standard Itô integral
\[
dY_t = f \left( X_t \right) \, dX_t = \sum_{i=1}^d f_i \left( X_t \right) \, dX^i_t, \quad Y_0 = y_0 \in \mathbb{R}^e.
\]

Then for all \( p \in (2, \gamma) \),
\[
d_{\text{p-var};[0,1]} (Y^n, Y) \to 0 \text{ in probability as } n \to \infty.
\]

If in addition the finite variation process \( V \) of \( X \) is bounded in \( L^q(\mathbb{P}) \) \( (q \geq 1) \) and \( M \in \mathcal{M}^p \), then the convergence also holds in \( L^p(\mathbb{P}) \).

Before presenting the proof of Theorem 4.1, we set up some additional notation. Denote the rough path lift of \( X_t = (X_t, X_t) \) by \( \Xi_t \). Thus \( \Xi : \Delta_{[0,1]} \to G^2(\mathbb{R}^{2d}) \) with
\[
\Xi_{s,t} : = \left[ \begin{array}{c|c} X_{s,t} & X_{s,t} \\ \hline X_{s,t} & X_{s,t} \end{array} \right] + \frac{1}{2} \left[ \begin{array}{c|c} X_{s,t} & X_{s,t} \\ \hline X_{s,t} & X_{s,t} \end{array} \right] \otimes \left[ \begin{array}{c} A_{s,t} \\ A_{s,t} \end{array} \right] + \left[ \begin{array}{c} A_{s,t} \\ A_{s,t} \end{array} \right],
\]
where \( A : \Delta_{[0,1]} \to [\mathbb{R}^d, \mathbb{R}^d] \) is the Lévy area process of \( X \). Importantly we have \( \Xi = \Xi/\Xi \); that is \( \Xi = \Xi/\Xi \) is the corollary from perturbing \( \Xi \) by the map \( \psi : \Delta_{[0,1]} \to [\mathbb{R}^{2d}, \mathbb{R}^{2d}] \):
\[
\Xi^\infty = \Xi + \psi = \exp_2 \left( \Xi + \left[ \begin{array}{c|c} A_{s,t} & A_{s,t} \\ \hline A_{s,t} & A_{s,t} \end{array} \right] \right) + \psi = \exp_2 \left( \Xi + \left[ \begin{array}{c} A_{s,t} \\ A_{s,t} \end{array} \right] + \psi \right),
\]
where
\[
\psi : \Xi_{s,t} = \left[ \begin{array}{c} 0 \\ \frac{1}{2} \langle X \rangle_{s,t} \end{array} \right] \in [\mathbb{R}^{2d}, \mathbb{R}^{2d}].
\]
The appearance of the covariation term \( \psi \) in the limit \( \Xi^\infty \) of Theorem 3.1 allows us to recover the Itô integral from the Stratonovich integral limit; the drift terms cancel out. We use the notation \( \rho : \mathbb{R}^d \oplus \mathbb{R}^e \to \mathbb{R}^d \) for the canonical projection.
Proof of Theorem 4.1. We consider a more complicated SDE on the larger space $\mathbb{R}^d \oplus \mathbb{R}^e$ and denote the standard bases of $\mathbb{R}^d$ and $\mathbb{R}^e$ by $(\bar{e}_i)_{i=1}^d$ and $(\bar{\tau}_j)_{j=1}^e$ respectively. In particular, set $z = (\bar{z}, \bar{\zeta}) \in \mathbb{R}^d \oplus \mathbb{R}^e$ and define the vector fields on $\mathbb{R}^d \oplus \mathbb{R}^e$ by

$$Q_i(z) = (\bar{e}_i, 0), \quad W_k(z) = \sum_{j=1}^e (0, \bar{\tau}_j) f_k^j(\bar{z}),$$

for $i, k = 1, \ldots, d$ (we have used the standard notation convention: $f_i(x) = (f_i^1(x), \ldots, f_i^e(x)) \in \mathbb{R}^e$ for $x \in \mathbb{R}^d$). Let $z^n = (\bar{z}^{n,1}, \ldots, \bar{z}^{n,d}, \bar{\zeta}^{n,1}, \ldots, \bar{\zeta}^{n,e})$ be the solution to the SDE

$$dz^n_t = \sum_{i=1}^d Q_i(z^n_t) dX^n_t^{i, \bar{z}^{i,k}} + W_1(z^n_t) dX_t^{1, f_1^k} + \ldots + W_d(z^n_t) dX_t^{d, f_d^k},$$

Since the axis path $X^n$ is piecewise linear, we could equivalently formulate the above SDE into its Stratonovich form without incurring a corrective drift term. It follows that the projection of $z^n$ onto $\mathbb{R}^e$, $\bar{z}^n := \rho(z^n)$, satisfies the SDE:

$$d\bar{z}^n_t = \sum_{i=1}^d W_i^j (\bar{z}^n_t) dX_t^{i, f_i^j} = \sum_{i=1}^d f_i^j (\bar{z}^n_t) dX_t^{i, f_i^j}.$$  

Moreover, [6, Theorem 17.3] tells us that $z^n = \pi_{(Q, W)} (0, (0, y_0); X^n)$ almost surely.

A simple computation confirms that the only non-zero Lie brackets between the vector fields $Q_{i_1}, Q_{i_2}, W_{k_1}$ and $W_{k_2}$ are of the form

$$[Q_{i_1}, W_{k_2}] (\bar{z}) = Q_{i_2} W_{k_1} (\bar{z}) = \sum_{j=1}^e (0, \bar{\tau}_j) \partial_i f_k^j(\bar{z}),$$

for all $i, k = 1, \ldots, d$. We denote this collection of vector fields by $L := (Q_i W_k : i, k \in \{1, \ldots, d\})$.

Consider the SDE:

$$dz_t = (Q + W)(z_t) dX_t = (Q + W)(z_t) \circ dX_t - \frac{1}{2} \sum_{i,k=1}^d Q_i W_k (z_t) d\langle X^i, X^k \rangle_t$$

$$= (Q + W)(z_t) \circ dX_t - \frac{1}{2} \sum_{i,k=1}^d [Q_i, W_k] (z_t) d\langle X^i, X^k \rangle_t,$$

with $z_0 = (0, y_0) \in \mathbb{R}^d \oplus \mathbb{R}^e$. It follows from [6, Theorem 17.3] that almost surely:

$$z = \pi_{(Q, W, L)} (0, (0, y_0); (X, \psi)).$$

As noted above, $\bar{X}^\psi = X^\infty$ and so the perturbation theorem of [6, Theorem 12.14] gives the equality:

$$z = \pi_{(Q, W, L)} (0, (0, y_0); (X, \psi)) = \pi_{(Q, W)} (0, (0, y_0); \bar{X}^\psi) = \pi_{(Q, W)} (0, (0, y_0); X^\infty).$$

In light of the convergence result of Theorem 3.1, we conclude that

$$\forall p > 2, \quad d_{p, \text{var}; [0,1]} (z^n, z) \to 0 \text{ in probability (or in } L^{p, \text{var}}(\mathbb{P}) \text{ as } n \to \infty).$$

Indeed, as noted in the above remark, almost sure equality and convergence in probability and $L^p(\mathbb{P})$ are all preserved under continuous mappings. Thus the Itô-Lyons map is continuous with respect to the driving signal under the $p$-variation topology. The proof is then finished by noting that $Y^n = \rho(z^n)$ and $Y = \rho(z)$. \hfill $\Box$
Remark 4.2. Such random ODEs have a natural interpretation in the context of financial mathematics and similar lines of research have been considered in the recent papers [2, 7, 8]. Indeed, we could consider $X$ to be a d-dimensional stream of continuous financial data (which we assume is in the form of a semimartingale indexed over the unit interval $[0,1]$). As noted by Friz in [2], the ODE (6) could be specifically interpreted as the investment strategy of an agent with delayed access to a discrete sampling of the market data. An interesting question to ask then is what the limiting behaviour of this SDE is when our agent learns more and more information on smaller time scales? This reduced latency means the agent would have access to increasingly more data points, thus reducing the delay between the backward and forward processes of $X^D$. Theorem 4.1 tells us that the limiting solution is the Itô integral and not the Stratonovich formulation suggested by the classical Wong-Zakai theorem [18].

The remainder of the paper is devoted to establishing the technical results referred to in the proof of Theorem 3.1.

5 Pointwise convergence of $M_D$ in $L^p(\mathbb{P})$

In this section we prove the pointwise convergence of $M_D$ to $M_\infty$ in $L^p(\mathbb{P})$ for $p > 2$ as $|D| \to 0$. We also provide the rate of this convergence. We first prove the pointwise convergence in $L^p(\mathbb{P})$ of the increment process $M^D$ to $M$ as $|D| \to 0$, where $M_t = (M_t, M_{s,t}) \in \mathbb{R}^{2d}$. But before this we give a small technical lemma which we will apply throughout this section.

Lemma 5.1. Let $X \in M_{0,loc}^D([0,1], \mathbb{R})$ and $D = (t_i) \in \mathcal{D}_{[0,1]}$. For all $p > 2$

\[
E \left[ \sum_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \right] \leq \|X\|_{L^{p/2}(\mathbb{P})} \sup_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \tag{7}
\]

Proof. Applying the Hölder inequality with $q = p/2$ and $r = p/(p-2)$ (so that $1/q + 1/r = 1$) gives:

\[
E \left[ \sum_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \right] \leq E \left[ \sup_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2-1} \right] \leq \|X\|_{L^{p/2}(\mathbb{P})} \sup_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \frac{p-2}{r} \leq \|X\|_{L^{p/2}(\mathbb{P})} \sup_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \left( \frac{p-2}{r} \right)^{1/r} \leq \|X\|_{L^{p/2}(\mathbb{P})} \sup_{t_i \in D} \langle X \rangle_{t_i,t_{i+1}}^{p/2} \left( \frac{p-2}{r} \right)^{1/r} 
\]

Recalling that $E \left[ \langle X \rangle_{0,1}^{p/2} \right]^{1/p} = \|M\|_p$, the proof is complete. \hfill \Box

Remark 5.2. If $p = 2$ in (7) then the lemma reduces to a trivial equality. For our purposes of proving pointwise convergence we need the right-hand side of (7) to converge to 0 as $|D| \to 0$ and so it would appear that we really do need $M \in M^p([0,1], \mathbb{R}^d)$.

Lemma 5.3 (Pointwise convergence of $M^D$). For all $(s,t) \in \Delta_{[0,1]}$, $M^D_{s,t} \to M_{s,t}$ in $L^p(\mathbb{P})$ as $|D| \to 0$. Moreover, there exists a constant $C = C(p, \|M\|_p)$ such that the rate of convergence is given by

\[
\left| M^D_{s,t} \right|_{L^p(\mathbb{P})} \leq CE \left[ \sup_{t_i \in D} \langle M \rangle_{t_i,t_{i+1}}^{p/2} \right] \left( \frac{p-2}{r} \right)^{1/r} 
\]
Proof. The proof is quick: suppose $D = (t_i)$ and select a coordinate $j \in \{1, \ldots, d\}$. For $t \in [t_i, t_{i+1})$ we find that Lemma 5.1 gives

$$E \left[ \left| M_{t}^{D,b,j} - M_{t}^{i,j} \right|^p \right] \leq 2^{p-1} E \left[ \sup_{t \in [t_i, t_{i+1})} \left| M_{t}^{D} - M_{t}^{i} \right|^p + \sup_{s \in [t_i, t_{i+1})} \left| M_{s}^{D} - M_{s}^{i} \right|^p \right] \leq C_1 \sum_{t_i \in D} E \left[ \sup_{s \in [t_i, t_{i+1})} \left| M_{s}^{D} - M_{s}^{i} \right|^p \right] = C_1 E \left[ \sum_{t_i \in D} (M_{t_i}^{D})^{p/2} \right] \leq C_2 E \left[ \sup_{t_i \in D} (M_{t_i}^{D})^{p/2} \right]^{p-2}.$$  

The identical bound holds for the forward process. The lemma follows. □

The next proposition establishes the pointwise convergence of the area process of $\mathbb{M}^D$.

**Proposition 5.4 (Pointwise convergence of $A^D$).** Fix $\theta, \lambda \in \{b, f\}$ and $i, j \in \{1, \ldots, d\}$. For all $(s, t) \in \Delta_{[0,1]}$ we have the following convergence in $L^{p/2}(\mathbb{P})$ as $|D| \to 0$:

$$A_{s,t}^{D,\theta,\lambda,i,j} \to A_{s,t}^{\theta,\lambda,i,j} := \begin{cases} A_{s,t}^{D,i,j} & \text{if } \theta = \lambda; \\ A_{s,t}^{D,i,j} - \frac{1}{2} (M_{s}, M_{t})_{s,t} & \text{if } \theta = b, \lambda = f; \\ A_{s,t}^{D,i,j} + \frac{1}{2} (M_{s}, M_{t})_{s,t} & \text{if } \theta = f, \lambda = b. \end{cases}$$  

Moreover, there exists a constant $C = C(p, \|M\|_p)$ such that the rate of convergence is given by

$$\left| A_{s,t}^{D} - \widehat{A}_{s,t} \right|^{1/2} \leq C E \left[ \sup_{t_i \in D} (M_{t_i}^{D})^{p/2} \right]^{\frac{p-2}{2p}}.$$

The main result of this section is the following:

**Corollary 5.5 (Pointwise convergence in $L^p(\mathbb{P})$ of $\mathbb{M}^D$).** For all $(s, t) \in \Delta_{[0,1]}$ we have

$$d \left( (\mathbb{M}_{s,t}^{D}, \mathbb{M}_{s,t}^{\infty}) \right) \to 0 \text{ in } L^p(\mathbb{P}) \text{ as } |D| \to 0.$$

Moreover, there exists a constant $C = C(p, \|M\|_p)$ such that

$$\left| d \left( (\mathbb{M}_{s,t}^{D}, \mathbb{M}_{s,t}^{\infty}) \right) \right| \leq C \left( E \left[ \sup_{t_i \in D} (M_{t_i}^{D})^{p/2} \right]^{\frac{p-2}{2p}} + E \left[ \sup_{t_i \in D} (M_{t_i}^{D})^{p/2} \right]^{\frac{p-2}{2p}} \right).$$  

**Proof.** The claim follows immediately from combining Lemma 3.3 and Proposition 5.4. Indeed,

$$d \left( (\mathbb{M}_{s,t}^{D}, \mathbb{M}_{s,t}^{\infty}) \right) \simeq \left| \pi_1 \left( (\mathbb{M}_{s,t}^{D} - \mathbb{M}_{s,t}^{\infty}) \right) \right| \sqrt{\left| \pi_2 \left( (\mathbb{M}_{s,t}^{D}, \mathbb{M}_{s,t}^{\infty}) \right) \right|^2} \simeq \left| M_{s,t}^{D} - M_{s,t}^{\infty} \right| \vee \left| A_{s,t}^{D} - \widehat{A}_{s,t} \right|^{1/2}.$$

The proof follows. □

It remains to prove Proposition 5.4 and we devote the remainder of this section to its proof. To this end we now consider two small technical lemmas.

**Lemma 5.6.** Let $X, Y \in \mathcal{M}_{0,\text{loc}}^c([0,1], \mathbb{R})$ and fix a partition $D = (t_i)_{i=0}^n \in \mathcal{D}_{[0,1]}$. Then there exists a constant $C = C(p, \|X\|_p, \|Y\|_p)$ for all $p > 2$,

$$E \left[ \left| \sum_{i=0}^{n-2} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+2}} - Y_{t_{i+1}}) \right|^{p/2} \right] \leq C E \left[ \sup_{t_i \in D} (X_{t_i}^{p/2}) \right]^{\frac{p-2}{p}}.$$
Proof. The claimed inequality is a consequence of the Burkholder-Gundy-Davis and Cauchy-Schwarz inequalities. Define the previsible process

\[ Z_r = \sum_{i=0}^{n-2} (X_{t_{i+1}} - X_{t_i}) 1 \{ r \in (t_{i+1}, t_{i+2}] \}. \]

We have

\[ \sum_{i=0}^{n-2} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+2}} - Y_{t_{i+1}}) = \int_0^1 Z_r \, dY_r, \]

which is a well defined local martingale since \( Z \) is previsible. It follows from the aforementioned inequalities that

\[
E \left[ \int_0^1 Z_r \, dY_r \right]^ {p/2} \leq C_1 E \left[ \int_0^1 |Z_r|^2 \, d(Y_r)^{p/4} \right] \leq C_1 E \left[ (Y_r)^{p/4} \sup_{r \in [0,1]} |Z_r|^{p/2} \right]
\leq C_1 E \left[ (Y_r)^{p/2} \right]^{1/2} E \left[ \sup_{r \in [0,1]} |Z_r|^p \right]^{1/2}.
\]

Since \( \sup_{r \in [0,1]} |Z_r|^p \leq \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}|^p \), applying Lemma 5.1 gives

\[
E \left[ \sup_{r \in [0,1]} |Z_r|^p \right] \leq E \left[ \sum_{i=0}^{n-1} |X_{t_{i+1}} - X_{t_i}|^p \right] \leq C_2 E \left[ \sum_{i=0}^{n-1} |X(Y_{t_{i+2}} - Y_{t_{i+1}})|^{p/2} \right]
\leq C_2 \|X_{[0,1]}\|_{L^p/2(\mathbb{P})} E \left[ \sup_{\tau \in \mathcal{D}} |X(\tau)^{p/2} (\tau_{t_{i+1}} + 1) \right]^{p-2} \frac{p}{p}.
\]

The result follows at once.

The next lemma uses the same techniques to derive the same rate of convergence in \( L^{p/2}(\mathbb{P}) \) for discrete approximations to the Itô integral.

Lemma 5.7. Suppose \( X, Y \in \mathcal{M}_{0, \text{loc}}([0,1], \mathbb{R}) \). There exists a constant \( C = C(p, \|X\|_p, \|Y\|_p) \) such that for any given partition \( D = (t_i) \in \mathcal{D}_{[0,1]} \), we have

\[
E \left[ \sum_{t_i \in D} X_{t_i} (Y_{t_{i+1}} - Y_{t_i}) - \int_0^1 X_r \, dY_r \right]^{p/2} \leq C E \left[ \sup_{\tau \in \mathcal{D}} |X(\tau)^{p/2} (\tau_{t_{i+1}} + 1) \right]^{p-2} \frac{p}{p}.
\]

Proof. Defining \( Z_r := \sum_{t_i \in D} X_{t_i} 1 \{ r \in (t_i, t_{i+1}] \} \), the standard argument gives

\[
E \left[ \sum_{t_i \in D} X_{t_i} (Y_{t_{i+1}} - Y_{t_i}) - \int_0^1 X_r \, dY_r \right]^{p/2} = E \left[ \int_0^1 (Z_r - X_r) \, dY_r \right]^{p/2} \]
\leq C_1 E \left[ \int_0^1 |Z_r - X_r|^2 \, d(Y_r)^{p/4} \right] \leq C_1 E \left[ |Z_r - X_r|^{p/2} \right]^{1/2}
\leq C_2 E \left[ \left\| X(Y_{t_{i+2}} - Y_{t_{i+1}}) \right\|_{L^p/2(\mathbb{P})} \right]^{1/2}
= C_2 \left[ \sum_{r \in [0,1]} \sum_{t_i \in D} |X_{t_i} - X_r|^p 1 \{ r \in (t_i, t_{i+1}] \} \right]^{1/2}.
\]
Noting the inequality
\[
\sup_{r \in [0,1]} \sum_{t_i \in D} |X_{t_i} - X_r|^p 1 \{ r \in (t_i, t_{i+1}] \} \leq \sum_{t_i \in D} \sup_{r \in [t_i, t_{i+1}]^c} |X_{t_i} - X_r|^p,
\]
we have
\[
E \left[ \sup_{r \in [0,1]} \sum_{t_i \in D} |X_{t_i} - X_r|^p 1 \{ r \in (t_i, t_{i+1}] \} \right] \leq \sum_{t_i \in D} E \left[ \sup_{r \in [t_i, t_{i+1}]^c} |X_{t_i} - X_r|^p \right] \leq C_3 \sum_{t_i \in D} E \left[ \langle X \rangle_{t_i, t_{i+1}}^{p/2} \right].
\]
The proof then follows from Lemma 5.1. \( \square \)

**Proof of Proposition 5.4.** Recall the standard area decomposition formula; for \( s < t < u \) we have
\[
A_{s,u} = A_{s,t} + A_{t,u} + [M_{s,t} \wedge M_{u,t}] / 2,
\]
where for vectors \( a, b \in \mathbb{R}^d \), \( a \wedge b \) denotes the \( d \times d \) antisymmetric matrix with entries \( (a \wedge b)_{i,j} = a_i b_j - a_j b_i \). From this formula and the fact that the increment process \( M^D \) converges pointwise in \( L^p(\mathbb{P}) \) to \( \bar{M} \) (Lemma 5.3), it suffices to only consider the convergence of \( A_{D,t_k}^D \) for \( t_k \in D, \) as \( |D| \to 0 \). For ease of notation we select a pair of coordinates \( i, j \in \{1, \ldots, d\} \) and set \( X = M^i \) and \( Y = M^j \) (this also frees up \( i, j \) for indices notation).

We first consider the situation when \( \theta = \lambda = b \). For \( D = (t_i) \in D_{[0,1]} \) we have
\[
A_{0,t_k}^{D,b,b;i,j} = \frac{1}{2} \sum_{i=1}^{k-1} (X_{t_{i-1}} (Y_{t_{i-1}} - Y_{t_{i-1}}) - Y_{t_{i-1}} (X_{t_{i-1}} - Y_{t_{i-1}})),
\]
and thus Lemma 5.7 gives
\[
A_{0,t_k}^{D,b,b;i,j} \to \frac{1}{2} \int_0^t (X_r dY_r - Y_r dX_r) = A_{0,t}^{M;i,j} \text{ in } L^{p/2}(\mathbb{P}) \text{ as } |D| \to 0.
\]
The case of \( \theta = \lambda = f \) is almost identical and thus omitted for brevity.

Next we consider the case of \( \theta = b \neq f = \lambda \). Since \( A_{D,f,b,i,j}^{D,b,b;i,j} = -A_{D,b,f,i,j}^{D,b,b;i,j} \) we need only prove the convergence for \( A_{D,b,f,i,j}^{D,b,b;i,j} \) to establish corresponding results for the case of \( \theta = f \neq b = \lambda \).

To this end, using the previous partition \( D = (t_i) \) we find
\[
A_{0,t_k}^{D,b,f;i,j} = \sum_{i=0}^{k-1} X_{t_i} (Y_{t_{i+2}} - Y_{t_{i+1}}) - \frac{1}{2} A_{0,t_k}^{D,b,b;i,j}. \]

We have the decomposition:
\[
\sum_{i=0}^{k-1} X_{t_i} (Y_{t_{i+2}} - Y_{t_{i+1}}) = \sum_{i=0}^{k-1} X_{t_{i+1}} (Y_{t_{i+2}} - Y_{t_{i+1}}) - \sum_{i=0}^{k-1} (X_{t_{i+1}} - X_{t_i}) (Y_{t_{i+2}} - Y_{t_{i+1}}).
\]
To deal with the second sum, we refer to the inequality of Lemma 5.6 which gives
\[
E \left[ \left( \sum_{i=0}^{k-1} (X_{t_{i+1}} - X_{t_i}) (Y_{t_{i+2}} - Y_{t_{i+1}}) \right)^{p/2} \right] \leq C_1 E \left[ \sup_{j=0, \ldots, k-1} \langle X \rangle_{t_i, t_{i+1}}^{p/2} \right], \quad (8)
\]
From our regularity assumptions on the martingale \( M \) we have \( \langle M \rangle_{t_1} \in L^{p/2}(\mathbb{P}) < \infty \), and thus dominated convergence combined with uniform continuity ensure that
\[
E \left[ \sup_{t_i \in D} \langle X \rangle_{t_i, t_{i+1}}^{p/2} \right] \to 0 \text{ as } |D| \to 0.
\]
Therefore, in light of Lemma 5.7, we deduce that

\[ A_{0,t}^{D,b,f,i,j} \to \int_0^t X_r \, dY_r - \frac{1}{2} X_t Y_t \quad \text{in } L^{p/2}(\mathbb{P}) \quad \text{as } |D| \to 0. \]

Rearranging this last term using the integration by parts formula\(^1\) we arrive at the identity

\[ \int_0^t X_r \, dY_r - \frac{1}{2} X_t Y_t = \int_0^t X_r \, dY_r - \frac{1}{2} \left( \int_0^t X_r \, dY_r + \int_0^t Y_r \, dX_r + \langle X,Y \rangle_t \right) \]

\[ = \frac{1}{2} \int_0^t (X_r \, dY_r - Y_r \, dX_r) - \frac{1}{2} \langle X,Y \rangle_t = A_{0,t}^{D,b,f,i,j} - \frac{1}{2} \langle X,Y \rangle_t. \]

For the rate of convergence claim we simply combine (8) with Lemma 5.7 concerning the convergence of the discrete Itô approximation sums. The proof is complete.

\[ \square \]

6 Maximal \(p\)-variation norm for \( \{ M^D : D \in D_{[0,1]} \} \)

The objective of this section is to prove the following maximal \(p\)-variation bound:

**Proposition 6.1 (Maximal \(p\)-variation bounds).** For all \( p > 2 \), there exists a constant \( C = C(p, \|M\|_p) \) such that

\[ \sup_{D \in D_{[0,1]}} \mathbb{E} \left[ \left\| M^D \right\|_{p\text{-var};[0,1]}^p \right] = C < \infty. \]

We split the proof into two propositions concerning \( M^{D,\theta,\lambda} \); the first dealing with \( \theta = \lambda \) and the second \( \theta \neq \lambda \), where \( \theta, \lambda \in \{b,f\} \).

**Proposition 6.2.** For all \( p > 2 \) and \( \theta \in \{b,f\} \), there exists a finite constant \( C = C(p, \|M\|_p) \) such that

\[ \sup_{D \in D_{[0,1]}} \mathbb{E} \left[ \left\| S_2 \left( M^{D,\theta,\lambda} \right) \right\|_{p\text{-var};[0,1]}^p \right] \leq C. \]

**Proof.** For all \( \theta \in \{b,f\} \), \( M^{D,\theta} \) is a reparameterization of the standard piecewise linear approximation \( \tilde{M}^D \) of \( M \) based on the partition \( D \). Since \( p\)-variation is invariant under reparameterization, it follows from the Burkholder-Gundy-Davis rough path result of [6, Theorem 14.15] (using the moderate function \( F(x) = x^p \)) that

\[ \mathbb{E} \left[ \left\| S_2 \left( M^{D,\theta} \right) \right\|_{p\text{-var};[0,1]}^p \right] = \mathbb{E} \left[ \left\| S_2 \left( \tilde{M}^D \right) \right\|_{p\text{-var};[0,1]}^p \right] \leq C \mathbb{E} \left[ \left\| (M)_{0,1}^{p/2} \right\| \right]. \]

The proof is complete. \[ \square \]

**Remark 6.3.** The use of [6, Theorem 14.15] in the proof of Proposition 6.2 critically requires that \( p > 2 \).

We now consider the case of \( \theta \neq \lambda \). As in the proof of Proposition 5.4, by symmetry it suffices to only consider the case of \( \theta = b \neq \lambda = f \).

**Proposition 6.4.** For all \( p > 2 \), there exists a constant \( C = C(p, \|M\|_p) \) such that

\[ \sup_{D \in D_{[0,1]}} \mathbb{E} \left[ \left\| S_2 \left( M^{D,b,f} \right) \right\|_{p\text{-var};[0,1]}^p \right] \leq C. \]

\[ \text{For continuous semimartingales } X, Y: d(XY) = XdY + YdX + d\langle X,Y \rangle \quad [6, \text{Proposition IV.3.1}]. \]
Unlike the case of $\theta = \lambda$, the proof is slightly more involved and requires a technical result for its proof. This is presented in the next lemma which is purely deterministic (that is it involves no probability).

Suppose we are given the points $(x_i, y_i)_{i=0}^n \subset \mathbb{R}^2$ and for simplicity we assume that $(x_0, y_0) = (0, 0)$. From these points we construct the piecewise-linear, axis-directed path $z : [0, 2n] \to \mathbb{R}^2$ such that $z_{2k} = (x_k, y_k)$ and $z_{2k+1} = (x_{k+1}, y_k)$ with linear interpolation between these times. Moreover, we also construct the path $w : [0, 2n] \to \mathbb{R}^2$ as the standard piecewise linear interpolation between these points; that is, $w_{2k} = (x_k, y_k)$ and $w_{2k+2} = (x_{k+1}, y_{k+1})$ with lines between these points. We include an example of $z$ and $w$ in Figure 6. The level-2 rough path lifts of these paths are given by

$$z := S_2(z), \quad w := S_2(w) \in C([0, 2n], G^2(\mathbb{R}^2)).$$

**Lemma 6.5.** For all $p > 2$ there exists a constant $C = C(p)$ such that

$$\|z\|_{p\text{-var}; [0,2n]}^p \leq \|w\|_{p\text{-var}; [0,2n]}^p + \left( \sum_{i=0}^{n-1} \left\{ (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right\} \right)^{p/2}.$$

**Proof.** Fix a partition $R = (s_k) \in \mathcal{D}_{[0,2n]}$ and set

$$s_{k+1}^* := \begin{cases} 
2m & \text{if there exists an integer } m \text{ such that } s_k \leq 2m \leq s_{k+1}; \\
s_{k+1} & \text{otherwise};
\end{cases}$$

Figure 2: An example of the paths $z$ and $w$ (dashed)
Certainly \( s_k \leq s_{k+1}^* \leq s_{k+1} \) for all \( k \). By the subadditivity of the Carnot-Caratheodory norm,

\[
2^{1-p} \sum_{s_k \in R} \| z_{s_{k+1}-s_{k+1}} \|^p \leq \sum_{s_k \in R} \left( \| w_{s_k-s_{k+1}} \|^p + \| w_{s_k-s_{k+1}}^{-1} \otimes z_{s_k-s_{k+1}} \|^p \right) 
\]

\[ (9) \]

\[
\leq 2^{p-1} \sum_{s_k \in R} \left( \| w_{s_k-s_{k+1}} \|^p + \| w_{s_k-s_{k+1}}^{-1} \|^p + \| z_{s_{k+1}-s_{k+1}} \|^p + \| w_{s_k-s_{k+1}}^{-1} \otimes z_{s_k-s_{k+1}} \|^p \right).
\]

Certainly we have

\[
\sum_{s_k \in R} \| w_{s_k-s_{k+1}} \|^p, \sum_{s_k \in R} \| w_{s_k-s_{k+1}}^{-1} \|^p \leq \| w \|^p_{p\text{-var};[0,2n]},
\]

so we focus on the two remaining terms of (10).

To this end, denoting the Lévy area process of \( z \) by \( A_z \), it follows immediately that

\[
\sum_{s_k \in R} \| z_{s_{k+1}-s_{k+1}} \|^p \\
\approx \sum_{s_k \in R} \left( \| z_{s_{k+1}-s_{k+1}} \|^{1/2} \right)^p \leq \sum_{s_k \in R} \left( \| z_{s_{k+1}-s_{k+1}} \|^p + \| A_z \|^{p/2} \right) \\
\leq \sum_{s_k \in R} \| z_{s_{k+1}-s_{k+1}} \|^p + \left( \sum_{s_k \in R} \| A_z \|^p \right)^{p/2} \leq |z|^{p}_{p\text{-var};[0,2n]} + \left( \sum_{s_k \in R} \| A_z \|^p \right)^{p/2}.
\]

Each coordinate \( z^i \) of \( z \) is simply a reparameterization of \( w^i \) and thus the invariance of \( p \)-variation under reparameterization implies that

\[
|z|_{p\text{-var};[0,2n]} = |w|_{p\text{-var};[0,2n]} \leq \| w \|^p_{p\text{-var};[0,2n]}.
\]

By definition \( s_{k+1}^*, s_{k+1} \subset [2m, 2m + 2] \) for some integer \( m \in \{0, \ldots, n-1\} \). Thus the path \( z \) restricted to \( [s_{k+1}^*, s_{k+1}] \) is contained within an open right-angled triangle. Each corner-point of \( z \) can be contained in at most one of these intervals; therefore,

\[
\sum_{s_k \in R} \| A_z \|_{s_{k+1}-s_{k+1}} \leq \frac{1}{2} \sum_{i=0}^{n-1} |x_{i+1} - x_i| |y_{i+1} - y_i| \leq \frac{1}{2} \sum_{i=0}^{n-1} \left( |x_{i+1} - x_i|^2 + |y_{i+1} - y_i|^2 \right),
\]

using the inequality \( ab \leq 1/2(a^2 + b^2) \) for \( a, b \geq 0 \). It remains to bound the third term of (10).

We have:

\[
\sum_{s_k \in R} \left( \| z_{s_{k+1}-s_{k+1}} - w_{s_{k+1}^*} \|^{p/2} \right) \leq \sum_{s_k \in R} \left( \| z_{s_{k+1}-s_{k+1}} - w_{s_{k+1}^*} \|^p + \| A^\phi \|_{s_{k+1}^*}^{p/2} \right),
\]

where \( A^\phi \) is the absolute area enclosed by the path \( \phi \), where \( \phi \) is the path given by the concatenation of \( (z_t)_{t \in [s_{k+1}^*]} \) and the path \( (w_{s_{k+1}^*} - t)_{t \in [0, s_{k+1} - s_{k+1}]} \). Evidently,

\[
\sum_{s_k \in R} \| z_{s_{k+1}^*} - w_{s_{k+1}^*} \|^p \leq 2^{p-1} \left( |z|^{p}_{p\text{-var};[0,2n]} + |w|^{p}_{p\text{-var};[0,2n]} \right).
\]
The area $A_{\phi}^{k+k+1}$ is precisely the absolute sum of the triangles enclosed by the path $z$ along its corner points plus a partial triangle area. Excluding these partial triangles, every triangle is enclosed precisely once by $\phi$. Thus it follows that

$$\sum_{x_k \in R} |A_{\phi}^{k+k+1}|^{p/2} \leq \left( \sum_{x_k \in R} |A_{\phi}^{k+k+1}| \right)^{p/2} \leq \left( \frac{1}{2} \sum_{i=0}^{n-1} |x_{i+1} - x_i| |y_{i+1} - y_i| \right)^{p/2} \leq C \left( \sum_{i=0}^{n-1} (|x_{i+1} - x_i|^2 + |y_{i+1} - y_i|^2) \right)^{p/2},$$

again using the inequality $ab \leq 1/2(a^2 + b^2)$.

Putting the above inequalities together and noting that the resultant bound is independent of the chosen partition $R \in D_{[0,2n]}$ completes the proof.

We now apply the previous lemma to our process $M^D$:

**Proposition 6.6.** For all $p > 2$, there exists a constant $C = C(p, \|M\|_p)$ such that

$$\sup_{D \in D_{[0,1]}} \mathbb{E} \left[ \|S_2(M^{D,b,f})\|_{p-\text{var};[0,2n]}^p \right] \leq C.$$

**Proof.** It suffices to prove the claim for $S_2(M^{D,b,f,i,j})$, where $i,j \in \{1, \ldots, d\}$. First fix a partition $D = (t_k)_{k=0}^n \in D_{[0,1]}$. We set the values $x_0 = 0, x_n = M_t^n,$

$$x_k = M_{t_{k-1}}^i \quad \text{for } k = 1, \ldots, n-1,$$

and $y_0 = 0, y_n = M_t^n$ with

$$y_k = M_{t_{k+1}}^j \quad \text{for } k = 0, \ldots, n-1.$$

Based on this discrete data $(x_k, y_k)_{k=0}^n$ we define $z$ and $w$ as above and note that $z$ is precisely a reparameterization of the process $M^{D,b,f,i,j}$ but instead defined over the interval $[0,2n]$. Similarly, $w$ is precisely a reparameterization of the standard piecewise linear approximation of $(M^t, M^\theta)$ with respect to the partition $D$, which is itself a reparameterization of $M^{D,\theta,b,i,j}$ for any $\theta \in \{b, f\}$. Thus Lemma 6.2 informs us that

$$\mathbb{E} \left[ \|w\|_{p-\text{var};[0,2n]}^p \right] = \mathbb{E} \left[ \|S_2(M^{D,b,f,i,j})\|_{p-\text{var};[0,1]}^p \right] \leq C_1.$$

It follows from Lemma 6.5 that

$$\mathbb{E} \left[ \|S_2(M^{D,b,f,i,j})\|_{p-\text{var};[0,1]}^p \right] = \mathbb{E} \left[ \|z\|_{p-\text{var};[0,2n]}^p \right] \leq C_1 \left( \mathbb{E} \left[ \|w\|_{p-\text{var};[0,2n]}^p \right] + \mathbb{E} \left[ \sum_{k=0}^{n-1} \left( M_{t_{k+1}}^i - M_{t_k}^i \right)^2 + \left( M_{t_{k+1}}^j - M_{t_k}^j \right)^2 \right] \right)^{p/2} \right) \leq C_2 + C_3 \mathbb{E} \left[ \sum_{k=0}^{n-1} \left( M_{t_{k+1}}^i - M_{t_k}^i \right)^2 + \left( M_{t_{k+1}}^j - M_{t_k}^j \right)^2 \right]^{p/2} \right].$$
From the discrete Burkholder-Davis-Gundy (BDG) inequality (c.f. [6, Chapter 14]) followed by both sides of the continuous BDG inequality, we have
\[
\mathbb{E} \left[ \left( \sum_{k=0}^{n-1} \left( M_{t_{k+1}}^i - M_{t_k}^i \right)^2 \right)^{p/2} \right] \leq C_4 \mathbb{E} \left[ \sup_{t_k \in D} \left| M_{t_k}^i - M_{t_0}^i \right|^p \right] \leq C_5 \mathbb{E} \left[ (M_{t_{0.1}}^{i,p/2})^2 \right] \leq C_5 \|M\|^p_p.
\]
Thus there exists some constant \( C = C \left( p, \|M\|_p \right) \) such that
\[
\mathbb{E} \left[ \left\| S_2 \left( M^{D,b,i} \right) \right\|_{p\text{-var} \times [0,2n]}^p \right] \leq C.
\]
As this bound is independent of the original partition choice of \( D \), the proof is complete. \( \square \)

**Remark 6.7.** Initially it was thought that the dyadic trick of bounding the \( p \)-variation of a level-2 rough path by a weighted sum of its moments and areas over every dyadic interval in \([0,1]\) (see [12]) could be employed. To the present author it appears that this technique enforces restrictions on the rate of convergence of the partition mesh \( |D| \) and demands that the partitions be nested. Our algebraic proof above is much longer and involved but places no restrictions on the sequence of partitions. Indeed, the maximal bound holds for any partition \( D \in D_{[0,1]} \).

## 7 A tightness property

We arrive at a tightness result on a countable rough path collection \( \{M_{D,n}^i\}_{n=1}^{\infty} \) satisfying \( |D_n| \to 0 \) as \( n \to \infty \). But first a simple lemma.

**Lemma 7.1.** Fix \( D \in D_{[0,1]} \). If \( \delta \geq |D| \) then
\[
\sup_{|t-s| \leq \delta} \left| M_{s,t}^D \right| \leq 3 \sup_{|t-s| \leq 2\delta} \left| M_{s,t} \right|.
\]

**Proof.** Denoting the standard piecewise-linear approximation of \( M \) based on the partition \( D \) by \( M^D \), we certainly have
\[
\sup_{|t-s| \leq \delta} \left| M_{s,t}^D \right| \leq \sup_{|t-s| \leq \delta} \left| M_{s,t} \right|.
\]
Suppose \( D = (t_i) \). The process \( M^D \) is a reparameterization of \( \tilde{M}^D \). Indeed, \( M^D \) runs at twice the speed of \( \tilde{M}^D \) over \([t_i, t_{i+1}] \) then stops over \([t_i^*, t_{i+1}^*] \). Considering the endpoints it follow that
\[
\sup_{|t-s| \leq \delta} \left| M_{s,t}^D \right| \leq 2 \sup_{|t-s| \leq |D|} \left| \tilde{M}_{s,t}^D \right| + \sup_{|t-s| \leq 2\delta} \left| \tilde{M}_{s,t}^D \right|.
\]
The proof is complete. \( \square \)

**Corollary 7.2 (Tightness property).** Suppose \( M \in \mathcal{M}^p \left( [0,1], \mathbb{R}^d \right) \) for some \( p > 2 \) and let \( (D_n)^\infty_{n=1} \) be a sequence of partitions with \( |D_n| \to 0 \) as \( n \to \infty \). Then
\[
\limsup_{\delta \downarrow 0} \mathbb{E} \left[ \sup_{|t-s| \leq \delta} \left| M_{s,t}^{D_n} \right|^p \right] = 0. \tag{11}
\]

**Proof.** Fix \( \theta, \lambda \in \{b,f\} \), \( i, j \in \{1, \ldots, d\} \) and consider the two-dimensional path \( t \in [0,1] \mapsto (M_{t}^{D,\theta,i}, M_{t}^{D,\theta,j}) \) with Lévy area \( A^{D,\theta,\lambda,i,j} : \Delta_{[0,1]} \to \mathbb{R}^2 \). Since this path is piecewise linear and axis-directed, basic isoperimetry theory tells us that for \( \delta \leq |D| \):
\[
\sup_{|t-s| \leq \delta} \left| A_{s,t}^{D,\theta,\lambda,i,j} \right| \leq \frac{1}{2} \left( \sup_{|t-s| \leq \delta} \left| M_{s,t}^{D,\theta,i} \right| + \sup_{|v-u| \leq \delta} \left| M_{u,v}^{D,\lambda,j} \right| \right)^2.
\]
That is, the area must be bounded above by the area of the right-angled triangle with side-lengths given by sum of the supremum of each increment processes. Using the two classical inequalities: \((a + b)^q \leq 2^{q-1}(a^q + b^q),\) \(q \geq 1\) and \(ab \leq 1/2(a^2 + b^2)\) for \(a, b \geq 0,\) a quick computation confirms that for the case of \(\delta \leq |D|:\)

\[
E \left[ \sup_{|t-s| \leq \delta} \left| A_{s,t}^{D,\theta,\lambda;i,j} \right|^{p/2} \right] \leq C \left( E \left[ \sup_{|t-s| \leq \delta} \left| M_{s,t}^{D,\theta;i} \right|^{p} \right] + E \left[ \sup_{|v-u| \leq \delta} \left| M_{u,v}^{D,\lambda;j} \right|^{p} \right] \right)
\]

for some universal constant \(C\) depending only on \(p.\) In the case of \(\delta \geq |D|,\) the proof of Lemma 6.5 gives

\[
\sup_{|t-s| \leq \delta} \left| A_{s,t}^{D,\theta,\lambda;i,j} \right|^{p/2} \leq \sum_{t_k \in D} \left( \left| M_{t_k,t_{k+1}}^{D,\theta;i} \right|^{p} + \left| M_{t_k,t_{k+1}}^{D,\lambda;j} \right|^{p} \right) \]

Appealing to Lemma 5.1, we have:

\[
E \left[ \sup_{|t-s| \leq \delta} \left| A_{s,t}^{D,\theta,\lambda;i,j} \right|^{p/2} \right] \leq C_1 \left( E \left[ \sup_{t_k \in D} \left( \left| M_{t_k,t_{k+1}}^{D,\theta;i} \right| + \left| M_{t_k,t_{k+1}}^{D,\lambda;j} \right| \right)^{p} \right] \right)
\]

\[\leq C_1 \sup_{l=1,\ldots,d} E \left[ \sup_{t_k \in D} \left( \left| X_{t_k}^{l} \right|^{p/2} \right)^{p/2} \right] \leq C_1 \sup_{l=1,\ldots,d} E \left[ \sup_{|t-s| \leq \delta} \left| X_{s,t}^{l} \right|^{p/2} \right],\]

which certainly converges to zero as \(\delta \rightarrow 0.\)

Thus it suffices to prove the limit (11) only for the increment process \(M^D_\square\) and not the corresponding rough path lift \(\tilde{M}^D.\) For the purpose of creating a contradiction, suppose that the conclusion of the corollary is false; that is, there exist some constant \(\epsilon > 0\) and an increasing sequence \((n_k)_{n=1}^\infty \subset \mathbb{Z}_+\) such that for all \(n:\)

\[
E \left[ \sup_{|t-s| \leq |D_{n_k}|} \left| M_{s,t}^{D_{n_k}} \right|^{p} \right] > \epsilon. \quad (12)
\]

It follows that there exists an infinite subsequence \((a_m)_{m=1}^\infty\) of \((n_k)_{n=1}^\infty\) such that \(n_j \leq m_j.\)

Indeed, if it were otherwise we could find an infinite subsequence \((a_m)_{m=1}^\infty\) of \((n_k)_{n=1}^\infty\) such that \(m_j > k_m.\)

\[
E \left[ \sup_{|t-s| \leq |D_{m_j}|} \left| M_{s,t}^{D_{m_j}} \right|^{p} \right] > \epsilon.
\]

Since \(m_j > k_m,\) \(|D_{m_j}| \leq |D_{k_m}|\) and so by Lemma 7.1:

\[
\epsilon < E \left[ \sup_{|t-s| \leq |D_{m_j}|} \left| M_{s,t}^{D_{m_j}} \right|^{p} \right] \leq E \left[ \sup_{|t-s| \leq |D_{k_m}|} \left| M_{s,t}^{D_{k_m}} \right|^{p} \right] \leq 3^p E \left[ \sup_{|t-s| \leq |D_{k_m}|} \left| M_{s,t} \right|^{p} \right].
\]

but this last quantity converges to zero as \(j \rightarrow \infty\) by the dominated convergence theorem. Indeed, since we have \(M \in \mathcal{M}^p\) for some \(p > 2,\) the classical BDG inequality gives us the necessary domination in order to apply the dominated convergence theorem..
Thus there exists such a subsequence \((k_n)_{j=1}^\infty\) of \((k_n)_{n=1}^\infty\) such that \(n_j \leq k_n\) for each \(j\).

This means \(|D_{n_j}| \geq |D_{k_n}|\) and so appealing to Lemma 7.1 again, it follows that

\[
\sup_{|t-s| \leq D_{n_j}} \left| M_{s,t}^{D_{k_{n_j}}} \right|^p \leq 3^p \sup_{|t-s| \leq 2D_{n_j}} |M_{s,t}|^p.
\]

Combined with (12), we conclude:

\[
\epsilon < \mathbb{E} \left[ \sup_{|t-s| \leq D_{n_j}} \left| M_{s,t}^{D_{k_{n_j}}} \right|^p \right] \leq 3^p \mathbb{E} \left[ \sup_{|t-s| \leq 2D_{n_j}} |M_{s,t}|^p \right] \to 0 \text{ as } j \to \infty,
\]

and so the desired contradiction is established.

**Remark 7.3.** Corollary 7.2 will not hold for an arbitrary sequence of partitions \((D_n)_{n=1}^\infty \subset D_{[0,1]}\). The condition \(|D_n| \to 0\) as \(n \to \infty\) is necessary. To show this we provide a simple counterexample: suppose we have a one dimensional Brownian motion \(B\) on \([0,1]\) and set the partition sequence \(D_n = \{0, 1/n, 1\}\). Then \(B_{1/n}^{D_{n_j}} = (0, X_1)\) for all \(n\), and thus for all \(k\):

\[
\sup_n \mathbb{E} \left[ \sup_{|t-s| \leq D_{n_j}} \left| M_{s,t}^{D_{n_j}} \right|^p \right] \geq 1.
\]

**Acknowledgements**

The first author would like to thank Dr. Weijun Xu for his helpful comments and Drs. Zhongmin Qian and Danyu Yang for their careful reading of earlier drafts. The research of all authors is supported by the European Research Council under the European Union’s Seventh Framework Programme (FP7-IDEAS-ERC, ERC grant agreement nr. 291244). The research of Terry Lyons is supported by EPSRC grant EP/H000100/1. The authors are grateful for the support of the Oxford-Man Institute.

**References**


