CANONICAL RDES AND GENERAL SEMIMARTINGALES AS ROUGH PATHS

By Ilya Chevyrev[*,‡] and Peter K. Friz[†,§,¶]

University of Oxford[‡], TU Berlin[§] and WIAS[¶]

In the spirit of Marcus canonical stochastic differential equations, we study a similar notion of rough differential equations (RDEs), notably dropping the assumption of continuity prevalent in the rough path literature. A new metric is exhibited in which the solution map is a continuous function of the driving rough path *and* a so-called path function, which directly models the effect of the jump on the system. In a second part, we show that general multidimensional semimartingales admit canonically defined rough path lifts. An extension of Lépingle's BDG inequality to this setting is given, and in turn leads to a number of novel limit theorems for semimartingale driven differential equations, both in law and in probability, conveniently phrased via Kurtz–Protter's uniformly-controlled-variations (UCV) condition. A number of examples illustrate the scope of our results.

1. Introduction. Itô stochastic integrals are well-known to violate a first order chain rule of Newton–Leibniz type, as is manifest from Itô's formula. In a number of applications, is is important to have a chain rule which, in the context of *continuous* semimartingales, was achieved in a satisfactory way by Stratonovich stochastic integration, which - loosely speaking - replaces left-point evaluation (in Itô–Riemann sums) by a symmetric midpoint evaluation. In the case of stochastic integration against *general* semimartingales (Lévy processes as an important special case), one can check that the Stratonovich integral no longer gives a chain rule – a more sophisticated approach is necessary to take care of jumps and the mechanism for doing this was developed by Marcus [39, 40]. The resulting "Marcus canonical integration" and "Marcus canonical (stochastic differential) equations" (in the terminology of [1]) was then investigated in a number of works, in-

cluding [23, 29, 22, 3, 2, 31, 30, 24], see also [1, 8] and the references therein.

On the other hand, continuous stochastic integration has been understood for some time in the context of rough path theory, see [38, 12, 17], [20, Ch. 14]. Loosely speaking, given a multidimensional continuous semimartingale X, the Stratonovich integral $\int f(X) \circ dX$ can be given a robust (pathwise) meaning in terms of $\mathbf{X} = (X, \int X \otimes \circ dX)$, a.e. realization of which constitutes a geometric rough path of finite p-variation for any $p > 2$. In contrast to the popular class of Hölder rough paths (usually sufficient to deal with Brownian motion, see e.g. [18]), p-variation has the advantage that it immediately allows for jumps. This also prompts the remark that Young theory, somewhat the origin of Lyons' rough paths, by no means requires continuity. Extensions of rough path theory to a general p-variation setting (for possibly discontinuous paths) were then explored in [47], [19] and [9]. However, none of these works provided a proper extension of Lyons' main result in rough path analysis: continuity of the solution map as a function of the driving rough path.

The first contribution of this paper is exactly that. We introduce a new metric on the space of càdlàg rough paths, and a type of (Marcus) canonical rough differential equation, for which one has the desired stability result. (Experts in *la théorie générale des processus* will recognize our topology as a p-variation rough paths variant of Skorokhod's strong M_1 topology.) In fact, we reserve the prefix "Marcus" to situations in which jumps only arise in the d-dimensional driving signal, and are handled (in the spirit of Marcus) by connecting X_{t-} and X_t by a straight line. (As a straight line has no area, this creates no jump in the area.) A "general" rough path (level N, over $\mathbb{R}^d$), however, can have jumps of arbitrary value $\mathbf{X}_{t-}^{-1} \otimes \mathbf{X}_t \in G^N(\mathbb{R}^d)$, and there are many (different) ways to implement Marcus's idea of continuously (parametrized over a fictitious time interval) connecting $\mathbf{X}_{t-}$ and $\mathbf{X}_t$. This is really a modelling choice, no different than choosing the driving signal and/or the driving vector fields. The notion of a "path function" ϕ helps us to formalize this, and indeed one may view $(\mathbf{X}, \phi)$ as the correct/extended rough driver.

In the second part of the paper, we show how general (càdlàg) semimartingales fit into the theory. In particular, we show that the canoncial lift of a semimartingale indeed is a.s. a (geometric) rough path of finite p-variation for any $p > 2$ (several special cases, including Lévy processes, were discussed in [47, 19] but the general case remained open). Our result is further made quantitative by establishing a BDG inequality for general local martingale rough paths. (We thus extend simultaneously the classical

p-variation BDG inequality [36], and its version for continuous local martingale rough paths [17]). This BDG inequality turns out to be a powerful tool, especially in conjunction with *uniform tightness* (UT) and *uniformly controlled variation* (UCV) of semimartingale sequences. (Introduced by [26] and [32] respectively, these conditions are at the heart of basic convergence theorems for stochastic integrals in Skorokhod topology; we work only with UCV in this article, but note that UCV and UT are equivalent under extra assumptions, e.g., convergence in law [33, Thm 7.6]. See Section 4.2 for the definition of UCV, and also [12] for some links to continuous semimartingale rough paths.)

As an example of an application to general semimartingale theory, we are able to state a criterion for convergence in law (resp. in probability) of Marcus SDEs, which is an analogue of the celebrated criterion for Itô SDEs due to Kurtz–Protter [32, Thm. 5.4] (we emphasize however that neither criterion is a simple consequence of the other). Loosely speaking, the result asserts that if $X, (X^n)_{n \geq 1}$ are $\mathbb{R}^d$-valued semimartingales such that $(X^n)_{n \geq 1}$ satisfies UCV and $X^n \to X$ in law (resp. in probability) for the Skorokhod topology, then the solutions to Marcus SDEs driven by X^n (along fixed vector fields) converge in law (resp. in probability) to the Marcus SDE driven by X (see Theorem 4.18 for a precise formulation). Our theorem (which crucially involves rough paths in the proof, but not in the statement) entails a pleasantly elegant approach to the Wong–Zakai theorem for SDEs with jumps (Kurtz–Protter–Pardoux [31], with novel interest from physics [8]) and in fact gives a number of novel limit theorems for Marcus canonical SDEs (see Theorem 4.29). We remark further that homeomorphism and diffeomorphism properties of solution flows are straightforward, in contrast to rather lengthy and technical considerations required in a classical setting (see, e.g., [1, p. 423] and the references therein). At last, we discuss the impact of more general path functions, noting that the "Marcus choice" really corresponds to the special case of the linear path function.

The paper is organized as follows. In Section 2 we collect some necessary preparatory material, including basic properties of path functions. In Section 3 we give meaning to canonical RDEs, for which drivers are (rough path, path function) pairs $(\mathbf{X}, \phi)$, and introduce the metric $\alpha_{p\text{-var}}$ for which the direct analogue of Lyons' universal limit theorem holds. Section 4 is then devoted to applications to càdlàg semimartingale theory, particularly in connection with the UCV condition and Wong–Zakai type approximations. We briefly comment in Section 5 on the further scope of the theory.

2. Preparatory material.

2.1. *Wiener and Skorokhod space.*   Throughout the paper, we denote by $C([s,t], E)$ and $D([s,t], E)$ the space of continuous and càdlàg functions (paths) respectively from an interval $[s,t]$ into a metric space (E, d).

Unless otherwise stated, we equip $C([s,t], E)$ and $D([s,t], E)$ respectively with the uniform metric

$$d_{\infty;[s,t]}(\mathbf{x}, \bar{\mathbf{x}}) = \sup_{u \in [s,t]} d(\mathbf{x}_u, \bar{\mathbf{x}}_u)$$

and the Skorokhod metric

$$\sigma_{\infty;[s,t]}(\mathbf{x}, \bar{\mathbf{x}}) = \inf_{\lambda \in \Lambda_{[s,t]}} |\lambda| \vee d_{\infty;[s,t]}(\mathbf{x} \circ \lambda, \bar{\mathbf{x}}),$$

where $\Lambda_{[s,t]}$ denotes the set of all strictly increasing bijections of $[s,t]$ to itself, and $|\lambda| := \sup_{u \in [s,t]} |\lambda(u) - u|$. When we omit the interval $[s,t]$ from our notation, we will always assume it is $[0,T]$. We recall that if E is Polish, then so is the Skorokhod space $D([0,T], E)$ with topology induced by σ_∞, also known as the J_1-topology.

We always let $\mathcal{D} = (t_0 = s < t_1 < \ldots < t_{k-1} < t_k = t)$ denote a partition of $[s,t]$, and notation such as $\sum_{t_i \in \mathcal{D}}$ denotes summation over all points in $\mathcal{D}$ (possibly without the initial/final point depending on the indexing). We let $|\mathcal{D}| = \max_{t_i \in \mathcal{D}} |t_{i+1} - t_i|$ denote the mesh-size of a partition.

For $p > 0$, we define the p-variation of a path $\mathbf{x} \in D([s,t], E)$ by

$$\|\mathbf{x}\|_{p\text{-var};[s,t]} := \sup_{\mathcal{D} \subset [s,t]} \left(\sum_{t_i \in \mathcal{D}} d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p \right)^{1/p}.$$

We use superscript notation such as $D^{p\text{-var}}([s,t], E)$ to denote subspaces of paths of finite p-variation. For continuous $\mathbf{x}$ only $p \geq 1$ is interesting, for otherwise $\mathbf{x}$ is constant.

REMARK 2.1.   At least when $E = \mathbb{R}^d$ (or $G^N(\mathbb{R}^d)$, see below) there is an immediate p-variation metric and topology. Due to the fact that convergence in J_1 topology to a continuous limit is equivalent to uniform convergence, a discontinuous path cannot be approximated by a sequence of continuous paths in the metric σ_∞. The same will be true for a J_1/p-variation (rough path) metric $\sigma_{p\text{-var}}$ below. That said, we will propose below a useful SM_1/p-variation (rough path) metric $\alpha_{p\text{-var}}$ under which the space of continuous rough paths is not closed.

2.2. *Rough paths.* For $N \geq 1$, we let $G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d) \equiv \sum_{k=0}^{N}(\mathbb{R}^d)^{\otimes k}$ denote the step-N free nilpotent Lie group over $\mathbb{R}^d$, embedded into the truncated tensor algebra $(T^N(\mathbb{R}^d), +, \otimes)$, which we equip with the Carnot-Carathéodory norm $\|\cdot\|$ and the induced (left-invariant) metric d. Recall that the step-N free nilpotent Lie algebra $\mathfrak{g}^N(\mathbb{R}^d) = \log G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d)$ is the Lie algebra of $G^N(\mathbb{R}^d)$. The space $C^{p\text{-var}}([0,T], G^N(\mathbb{R}^d))$, with $N = \lfloor p \rfloor$, is the classical space of (continuous, weakly) geometric p-rough paths as introduced by Lyons.

Unless otherwise stated, we always suppose a path $\mathbf{x} : [s,t] \to G^N(\mathbb{R}^d)$ starts from the identity $\mathbf{x}_s = 1_{G^N(\mathbb{R}^d)}$. We denote the increments of a path by $\mathbf{x}_{s,t} = \mathbf{x}_s^{-1}\mathbf{x}_t$. We consider on $D^{p\text{-var}}([s,t], G^N(\mathbb{R}^d))$ the inhomogeneous p-variation metric

$$(1) \qquad \rho_{p\text{-var};[s,t]}(\mathbf{x}, \bar{\mathbf{x}}) = \max_{1 \leq k \leq N} \sup_{\mathcal{D} \subset [s,t]} \Big(\sum_{t_i \in \mathcal{D}} \Big| \mathbf{x}_{t_i,t_{i+1}}^k - \bar{\mathbf{x}}_{t_i,t_{i+1}}^k \Big|^{p/k} \Big)^{k/p}.$$

Unless otherwise stated, we shall always assume that p and N satisfy $\lfloor p \rfloor \leq N$.

We let $C^{0,p\text{-var}}([s,t], G^N(\mathbb{R}^d))$ denote the closure in $C^{p\text{-var}}([s,t], G^N(\mathbb{R}^d))$ under the metric $\rho_{p\text{-var}}$ of the lifts of smooth paths $C^\infty([s,t], \mathbb{R}^d)$. Recall in particular that $C^{0,1\text{-var}}([s,t], G^N(\mathbb{R}^d))$ is precisely the space of absolutely continuous paths.

We let $V = (V_1, \ldots, V_d)$ denote a collection of vector fields in $\text{Lip}^{\gamma+m-1}(\mathbb{R}^e)$ with $\gamma > p$ and $m \geq 1$. For a geometric p-rough path $\mathbf{x} \in C^{p\text{-var}}([s,t], G^N(\mathbb{R}^d))$, we let $\pi_{(V)}(s, y_s; \mathbf{x}) \in C^{p\text{-var}}([s,t], \mathbb{R}^e)$ denote the solution to the RDE

$$dy_t = V(y_t)d\mathbf{x}_t, \quad y_s \in \mathbb{R}^e.$$

We let $U_{t \leftarrow s}^{\mathbf{x}} : \mathbb{R}^e \to \mathbb{R}^e$ denote the associated flow map $y \mapsto \pi_{(V)}(s, y; \mathbf{x})_t$, which we recall is an element of $\text{Diff}^m(\mathbb{R}^e)$. For further details on the theory of (continuous) rough paths theory, we refer to [20].

For the purpose of his paper we have (cf. [19])

DEFINITION 2.2. *Let $1 \leq p < N + 1$. Any $\mathbf{x} \in D^{p\text{-var}}([0,T], G^N(\mathbb{R}^d))$ is called a general (càdlàg, weakly) geometric p-rough path over $\mathbb{R}^d$. Define $\Delta\mathbf{x}_t := \mathbf{x}_{t_-}^{-1} \otimes \mathbf{x}_t$ and say $\mathbf{x}$ is Marcus-like if for all $t \in [0,T]$*

$$\log \Delta\mathbf{x}_t \in \mathbb{R}^d \oplus \{0\} \oplus \ldots \oplus \{0\} \subset \mathfrak{g}^N(\mathbb{R}^d),$$

where $\log$ is taken in $T^N(\mathbb{R}^d)$.

As we will see later, any canonical lift of a general d-dimensional semi-martingale X (with area given by $\frac{1}{2}\int[X^-,dX]$) gives rise to a Marcus-like general geometric p-rough path for $p > 2$. The model case of Lévy processes was studied in [46, 19].

2.3. *Path functions.* We briefly review and elaborate on the concept of a path function introduced in [9]. Let (E,d) be a metric space.

DEFINITION 2.3.   *A path function on E is a map $\phi : J \to C([0,1],E)$ defined on a subset $J \subseteq E \times E$ for which $\phi(x,y)_0 = x$ and $\phi(x,y)_1 = y$ for all $(x,y) \in J$.*

For a path $\mathbf{x} \in D([0,T],E)$, we say that $t \in [0,T]$ is a jump time of $\mathbf{x}$ if $\mathbf{x}_{t-} \neq \mathbf{x}_t$. We call the pair $(\mathbf{x},\phi)$ admissible if $(\mathbf{x}_{t-},\mathbf{x}_t) \in J$ for all jumps times t of $\mathbf{x}$. We say that two admissible pairs $(\mathbf{x},\phi)$ and $(\bar{\mathbf{x}},\bar{\phi})$ are equivalent, and write $(\mathbf{x},\phi) \sim (\bar{\mathbf{x}},\bar{\phi})$, if $\mathbf{x} = \bar{\mathbf{x}}$ and $\phi(\mathbf{x}_{t-},\mathbf{x}_t)$ is a reparametrization of $\bar{\phi}(\mathbf{x}_{t-},\mathbf{x}_t)$.

We denote by $\bar{\mathscr{D}}([0,T],E)$ the set of all admissible pairs $(\mathbf{x},\phi)$, and by $\mathscr{D}([0,T],E) = \bar{\mathscr{D}}([0,T],E)/\sim$ the set of all equivalence classes of admissible pairs. For a fixed path function ϕ, let $\mathscr{D}_\phi([0,T],E)$ denote the set of all $\mathbf{x} \in D([0,T],E)$ such that $(\mathbf{x},\phi)$ is admissible.

We will often simply say that ϕ is a path function on E and keep implicit the fact there is an underlying domain of definition J. We point out that situations where $J \neq E \times E$ arise naturally when studying solution maps of canonical càdlàg RDEs, see Theorem 3.13 and the discussion before it.

In the case that E is a Lie group with identity element 1_E (taken in this article to always be $G^N(\mathbb{R}^d)$), we shall often assume that ϕ is *left-invariant*, which is to say that there exists a subset $B \subseteq E$ such that $J = \left\{(x,y) \in E \times E \,\middle|\, x^{-1}y \in B\right\}$ and

$$\phi(x,y)_t = x\phi(1_E, x^{-1}y)_t, \ \forall(x,y) \in J, \ \forall t \in [0,1].$$

In this case, it is equivalent to view ϕ as a map $\phi : B \to C([0,1],E)$ such that $\phi(x)_0 = 1_E$ and $\phi(x)_1 = x$ for all $x \in B$, for which $\phi(x,y)_t = x\phi(x^{-1}y)_t$. Whenever we write ϕ with only one argument as $\phi(x)$, we shall always mean that it is left-invariant.

EXAMPLE 2.4 (log-linear and Marcus path function).   The prototypical example of a (left-invariant) path function ϕ on $G^N(\mathbb{R}^d)$, which we shall often refer to in the paper, is the *log-linear* path function

$$\phi(x)_t = e^{t\log x}, \ \ \forall x \in G^N(\mathbb{R}^d), \ \forall t \in [0,1],$$
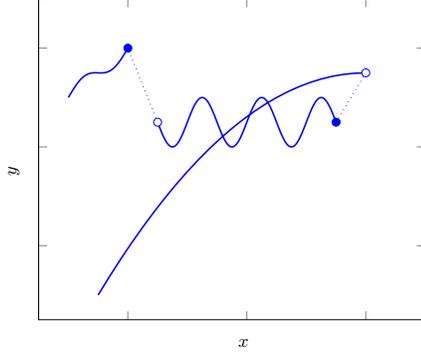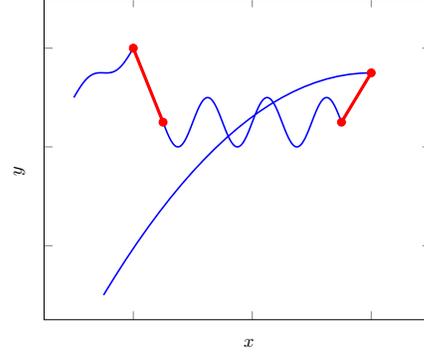
Fig 1: 2-dimensional càdlàg path

Fig 2: Marcus interpolation

where log is taken in $T^N(\mathbb{R}^d)$. Since $J = G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$, we have $(\mathbf{x}, \phi)$ is admissible for all $\mathbf{x} \in D([0,T], G^N(\mathbb{R}^d))$. When $N = 1$ we see a familiar special case: since $G^1(\mathbb{R}^d) \cong \mathbb{R}^d$, one has $\phi(x)_t = tx$, and then $\phi(x, y) = x + t(y - x)$. This is precisely the "Marcus interpolation" of a càdlàg path before and after its jump. See parametric plots in Figures 1 and 2.

For $(\mathbf{x}, \phi) \in \bar{\mathscr{D}}([0,T], E)$ we now construct a continuous path $\mathbf{x}^\phi \in C([0,T], E)$ as follows. Fix a convergent series of strictly positive numbers $\sum_{k=1}^\infty r_k$. Let $t_1, t_2, \ldots$ be the jump times of $\mathbf{x}$ ordered so that $d(\mathbf{x}_{t_1-}, \mathbf{x}_{t_1}) \geq d(\mathbf{x}_{t_2-}, \mathbf{x}_{t_2}) \geq \ldots$, and $t_j < t_{j+1}$ if $d(\mathbf{x}_{t_j-}, \mathbf{x}_{t_j}) = d(\mathbf{x}_{t_{j+1}-}, \mathbf{x}_{t_{j+1}})$. Let $0 \leq m \leq \infty$ be the number of jumps of $\mathbf{x}$.

Let $r = \sum_{k=1}^m r_k$ and define the strictly increasing (càdlàg) function

$$\tau : [0, T] \to [0, T + r], \quad \tau(t) = t + \sum_{k=1}^m r_k \mathbf{1}_{\{t_k \leq t\}}.$$

Note that $\tau(t-) < \tau(t)$ if and only if $t = t_k$ for some $1 \leq k < m + 1$. Moreover, note that the interval $[\tau(t_k-), \tau(t_k))$ is of length r_k.

Define $\widehat{\mathbf{x}} \in C([0, T + r], E)$ by

$$\widehat{\mathbf{x}}_s = \begin{cases} \mathbf{x}_t & \text{if } s = \tau(t) \text{ for some } t \in [0, T] \\ \phi(\mathbf{x}_{t_k-}, \mathbf{x}_{t_k})_{(s-\tau(t_k-))/r_k} & \text{if } s \in [\tau(t_k-), \tau(t_k)) \text{ for some } 1 \leq k < m + 1. \end{cases}$$

Denote by $\tau_r(t) = t(T + r)/T$ the increasing linear bijection from $[0, T]$ to $[0, T + r]$. We finally define

$$\mathbf{x}^\phi = \widehat{\mathbf{x}} \circ \tau_r \in C([0, T], E).$$

Fig 3: Hoff interpolation $\mathbf{x}^\phi$, with ϕ given as in Example 4.27
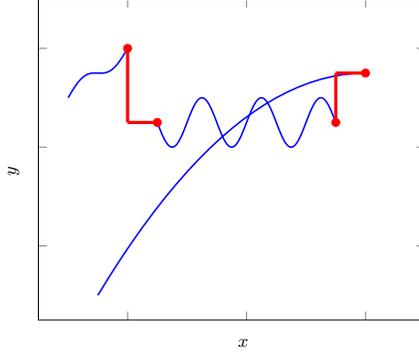
We note that one can recover $\mathbf{x} = \mathbf{x}^\phi \circ \tau_{\mathbf{x}}$ via the time change

$$\tag{2} \tau_{\mathbf{x}} := \tau_r^{-1} \circ \tau,$$

for which it holds that

$$\sup_{t\in[0,T]} |\tau_{\mathbf{x}}(t) - t| \leq \sum_{k=1}^{\infty} r_k.$$

REMARK 2.5 (Intrinsic definition of $\mathbf{x}^\phi$).   The construction of $\mathbf{x}^\phi$ involves an ad-hoc choice, namely the sequence (r_n) and the increasing bijection τ_r. If $\bar{\mathbf{x}}^\phi$ is constructed similarly, but via a sequence $(\bar{r}_n)$, followed by another reparametrisation given by $\bar{\tau}_{\bar{r}}$, then $\mathbf{x}^\phi$ and $\bar{\mathbf{x}}^\phi$ are reparametrizations of one another.

REMARK 2.6.   The construction above is similar to ones appearing in [19, 47], and is a simplification of the construction in [9]. The primary difference is that in [9] the added fictitious time r_k for the jump t_k depended further on the size of the jump $d(\mathbf{x}_{t_k-}, \mathbf{x}_{t_k})$. This extra dependence was used to show continuity of the map $\mathbf{x} \mapsto \mathbf{x}^\phi$ from $D([0,T], E) \to C([0,T], E)$, which we will not require here.

2.4. *A generalisation of Skorokhod's SM_1 topology.*   For $(\mathbf{x}, \phi) \in \bar{\mathscr{D}}([0,T], E)$ and $\delta > 0$, let $\mathbf{x}^{\phi,\delta} \in C([0,T], E)$ be constructed in the same procedure as $\mathbf{x}^\phi$ but using the series $\sum_{k=1}^{\infty} \delta r_k$ instead of $\sum_{k=1}^{\infty} r_k$.

LEMMA 2.7. *For all* $(\mathbf{x}, \phi), (\bar{\mathbf{x}}, \bar{\phi}) \in \bar{\mathscr{D}}([0,T], E)$*, the limit*

$$(3) \qquad \lim_{\delta \to 0} \sigma_{\infty;[0,T]}(\mathbf{x}^{\phi,\delta}, \bar{\mathbf{x}}^{\bar{\phi},\delta})$$

exists, is independent of the choice of series $\sum_{k=1}^{\infty} r_k$*, and induces a pseudometric on the set of equivalence classes* $\mathscr{D}([0,T], E)$*.*

PROOF. To show that the limit exists, note that for every $\delta, \bar{\delta} > 0$, there exists $\lambda \in \Lambda$ such that $|\lambda| < 2(\delta + \bar{\delta}) \sum r_k$ and $\mathbf{x}^{\phi,\bar{\delta}} = \mathbf{x}^{\phi,\delta} \circ \lambda$. Since $|\lambda \circ \bar{\lambda}| \leq |\lambda| + |\bar{\lambda}|$, it follows that for every $\delta, \bar{\delta} > 0$

$$|\sigma_{\infty}(\mathbf{x}^{\phi,\delta}, \bar{\mathbf{x}}^{\bar{\phi},\delta}) - \sigma_{\infty}(\mathbf{x}^{\phi,\bar{\delta}}, \bar{\mathbf{x}}^{\bar{\phi},\bar{\delta}})| < 4(\bar{\delta} + \delta) \sum r_k,$$

from which the existence of the limit follows. The fact that (3) is independent of the series $\sum r_k$ and is zero if $(\mathbf{x}, \phi) \sim (\bar{\mathbf{x}}, \bar{\phi})$ is straightforward. $\square$

DEFINITION 2.8. *Define the pseudometric* α_{∞} *on* $\mathscr{D}([0,T], E)$ *by*

$$(4) \qquad \alpha_{\infty}(\mathbf{x}, \bar{\mathbf{x}}) := \alpha_{\infty;[0,T]}((\mathbf{x}, \phi), (\bar{\mathbf{x}}, \bar{\phi})) := \lim_{\delta \to 0} \sigma_{\infty;[0,T]}(\mathbf{x}^{\phi,\delta}, \bar{\mathbf{x}}^{\bar{\phi},\delta}).$$

(Usually no confusion will arise by using the abusive notation on the left-hand side.)

Note that for any fixed ϕ, α_{∞} induces a genuine metric on the space $\mathscr{D}_{\phi}([0,T], E) \subseteq D([0,T], E)$. We note also that the strong M_1 (i.e., SM_1) topology on the space $D([0,T], \mathbb{R}^d)$ is a special case of the topology induced by the metric α_{∞}, as demonstrated by the following result.

PROPOSITION 2.9. *For* $E = \mathbb{R}^d$ *and* ϕ *the linear path function, it holds that* α_{∞} *induces the* SM_1 *topology on the space* $D([0,T], \mathbb{R}^d)$*.*

PROOF. It is straightforward to verify that α_{∞} in this case is equivalent to the metric d_s (see [45, Sec. 12.3.1]) which induces the SM_1 topology. $\square$

REMARK 2.10. The reader may wonder if convergence in the (Skorokhod J_1) metric σ_{∞} implies, as in the classical setting, convergence in the (Skorokhod SM_1-type) pseudometric α_{∞}. In essence, the answer is yes, however this requires "reasonable" path functions, see Lemma 2.20.

REMARK 2.11 (Restriction of time interval). It is trivial to see that uniform convergence of paths on $[0,T]$ implies convergence on any subinterval of $[0,T]$, while this fails for both Skorokhod J_1 and $(S)M_1$ metrics. The

observation generalizes to our setting and, in particular, the α_∞ metric does not behave well under restriction. Indeed, while for any $(\mathbf{x}, \phi) \in \mathscr{D}([0, T], E)$ the jumps of $\mathbf{x}|_{[s,t]}$ still belong to J, so that $(\mathbf{x}|_{[s,t]})^\phi$ is well-defined, it does not hold that $\alpha_\infty(\mathbf{x}^n, \mathbf{x}) \to 0$ implies that $\alpha_\infty(\mathbf{x}^n|_{[s,t]}, \mathbf{x}|_{[s,t]}) \to 0$.

We now collect several useful definitions and lemmas concerning path functions.

LEMMA 2.12.   *Let $(\mathbf{x}, \phi) \in \mathscr{D}([0, T], E)$ for which*

(5) $$\lim_{n \to \infty} \sup_{s \in [0,1]} d(\mathbf{x}_{t_n-}, \phi(\mathbf{x}_{t_n-}, \mathbf{x}_{t_n})_s) = 0,$$

where the limit is taken over some enumeration of jump times of $\mathbf{x}$. Let $(\mathbf{x}^k, \phi^k)_{k \geq 1}$ be a sequence in $\mathscr{D}([0, T], E)$ such that $\alpha_\infty(\mathbf{x}^k, \mathbf{x}) \to 0$. Then $\mathbf{x}_t^k \to \mathbf{x}_t$ for every continuity point of $\mathbf{x}$.

REMARK 2.13.   Note that condition (5) is satisfied whenever ϕ is either endpoint continuous or $(\mathbf{x}, \phi)$ has finite p-variation (see Definitions 2.14 and 2.17 below). In particular, since càdlàg paths are uniquely determined by their continuity points, it follows from Lemma 2.12 that α_∞ is a genuine metric on the space $\mathscr{D}^{p\text{-var}}([0, T], E)$ introduced in Definition 2.14.

PROOF.   Suppose t is a continuity point of $\mathbf{x}$. Then using (5) and the definition of α_∞, it holds that for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all k sufficiently large and δ_k sufficiently small we have

$$\sup_{s \in [t-\delta, t+\delta]} d(\mathbf{x}_t, (\mathbf{x}^k)_s^{\phi_k, \delta_k}) < \varepsilon,$$

from which the conclusion follows. $\square$

DEFINITION 2.14.   *For $p \geq 1$, we define the p-variation of $(\mathbf{x}, \phi) \in \mathscr{D}([0, T], E)$ as*

$$\|(\mathbf{x}, \phi)\|_{p\text{-var};[0,T]} := \|\mathbf{x}^\phi\|_{p\text{-var};[0,T]}$$

and let $\mathscr{D}^{p\text{-var}}([0, T], E)$ denote all $(\mathbf{x}, \phi) \in \mathscr{D}([0, T], E)$ of finite p-variation.

Moreover, a path function $\phi : J \to C([0, 1], E)$ is called p-approximating if there exists a function $\eta_{p\text{-var}} : [0, \infty) \to [1, \infty)$ such that for all $r \in [0, \infty)$

$$\sup_{(x,y) \in J; d(x,y) \leq r} \|\phi(x, y)\|_{p\text{-var};[0,1]} \leq \eta_{p\text{-var}}(r) d(x, y).$$

We say that $\eta_{p\text{-var}}$ is a p-variation modulus of ϕ.

REMARK 2.15. Due to the invariance of p-variation norms under reparametrizations, and our previous Remark 2.5, we see that there is no ambiguity in the definition of $\|(\mathbf{x}, \phi)\|_{p\text{-var};[0,T]}$ and that $\mathscr{D}^{p\text{-var}}([0,T], E)$ is well-defined.

The following lemma gives a simple criterion for a pair $(\mathbf{x}, \phi) \in \mathscr{D}([0,T], E)$ to have finite p-variation.

LEMMA 2.16 ([9] Lemma A.5). Let $p \geq 1$ and set $R = 1 + 2^p + 3^{p-1}$. Then for every $(\mathbf{x}, \phi) \in \mathscr{D}([0,T], E)$, it holds that

$$\|\mathbf{x}\|^p_{p\text{-var};[0,T]} \vee \Big(\sum_t \|\phi(\mathbf{x}_{t-}, \mathbf{x}_t)\|^p_{p\text{-var};[0,1]} \Big) \leq \|\mathbf{x}^\phi\|^p_{p\text{-var};[0,T]}$$

$$\leq R \|\mathbf{x}\|^p_{p\text{-var};[0,T]} + (R + 3^{p-1}) \sum_t \|\phi(\mathbf{x}_{t-}, \mathbf{x}_t)\|^p_{p\text{-var};[0,1]},$$

where the summations are over the jump times of $\mathbf{x}$.

In particular, if ϕ has a p-variation modulus $\eta_{p\text{-var}}$, then for all $\mathbf{x} \in \mathscr{D}_\phi([0,T], E)$,

$$\|\mathbf{x}^\phi\|^p_{p\text{-var};[0,T]} \leq \big[R + \eta_{p\text{-var}}(r)^p(R + 3^{p-1})\big] \|\mathbf{x}\|^p_{p\text{-var};[0,T]},$$

where $r = \sup_{t \in [0,T]} d(\mathbf{x}_{t-}, \mathbf{x}_t)$.

DEFINITION 2.17. A path function $\phi : J \to C([0,1], E)$ is called endpoint continuous if

1. $(x, x) \in J$ whenever $(x, y) \in J$,
2. $\phi(x, x) \equiv x$ for all $(x, x) \in J$, and
3. ϕ is continuous with $C([0,1], E)$ equipped with the uniform topology.

Moreover, we say that a function $\eta_\infty : [0, \infty) \to [0, \infty)$ is a uniform modulus of ϕ if $\eta_\infty(r) \geq r$ for all $r \geq 0$, $\lim_{r \to 0} \eta_\infty(r) = \eta_\infty(0) = 0$, and for all $(x, y), (\bar{x}, \bar{y}) \in J$

$$d_{\infty;[0,1]}(\phi(x, y), \phi(\bar{x}, \bar{y})) \leq \eta_\infty(\max\{d(x, \bar{x}), d(y, \bar{y})\}).$$

REMARK 2.18. In general, it is hard to find an explicit uniform modulus of a path function (or even show that one exists). But evidently if ϕ is restricted to $J \cap (K \times K)$ for a compact $K \subseteq E$, then a uniform modulus exists whenever ϕ is endpoint continuous.

EXAMPLE 2.19. Let ϕ be the log-linear path function on $G^N(\mathbb{R}^d)$. Then clearly ϕ is endpoint continuous and there exists a constant $C \geq 1$ such that for all $p \geq N$ and $x, y \in G^N(\mathbb{R}^d)$

$$\|\phi(x, y)\|_{p\text{-var};[0,1]} \leq Cd(x, y),$$

so that the constant C is a p-variation modulus of ϕ.

LEMMA 2.20.   *Suppose ϕ has a uniform modulus η_∞. Then for all $\mathbf{x}, \bar{\mathbf{x}} \in \mathscr{D}_\phi([0,T], E)$, it holds that $\alpha_\infty(\mathbf{x}, \bar{\mathbf{x}}) \leq \eta_\infty(\sigma_\infty(\mathbf{x}, \bar{\mathbf{x}}))$.*

PROOF.   Suppose there exists $\lambda \in \Lambda$ such that $|\lambda| < r$ and $d_\infty(\mathbf{x}, \bar{\mathbf{x}} \circ \lambda) < r$. Then for all $\delta > 0$ sufficiently small there exists $\lambda_\delta \in \Lambda$ such that $|\lambda_\delta| < r$ and $d_\infty(\mathbf{x}^{\phi,\delta}, \bar{\mathbf{x}}^{\phi,\delta} \circ \lambda_\delta) < \eta_\infty(r)$, and the conclusion follows.   □

3. Canonical RDEs driven by general rough paths.   To ease notation, we assume throughout this section that all path spaces, unless otherwise stated, are defined on the interval $[0,T]$ and take values in $G^N(\mathbb{R}^d)$. For example $\mathscr{D}^{p\text{-var}}$ will be shorthand for $\mathscr{D}^{p\text{-var}}([0,T], G^N(\mathbb{R}^d))$.

3.1. *Notion of solution.*   Following the notation of Section 2.2, let $1 \leq p < N+1$ and fix a family of vector fields $V = (V_1, \ldots, V_d)$ in $\mathrm{Lip}^{\gamma+m-1}(\mathbb{R}^e)$ for some $\gamma > p$ and $m \geq 1$. For $\mathbf{x} \in D^{p\text{-var}}$, we would like to solve the RDE

$$\text{``}dy_t = V(y_t)d\mathbf{x}_t\text{''}.$$

Our notion of solution to this equation will depend on a path function ϕ defined on a subset $J \subseteq G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$, and therefore the fundamental input to an RDE will be a pair $(\mathbf{x}, \phi) \in \mathscr{D}^{p\text{-var}}$.

DEFINITION 3.1 (Canonical RDE).   *Consider $(\mathbf{x}, \phi) \in \mathscr{D}^{p\text{-var}}$ and let $\tilde{y} \in C^{p\text{-var}}([0,T], \mathbb{R}^e)$ be the solution to the continuous RDE*

$$d\tilde{y}_t = V(\tilde{y}_t)d\mathbf{x}_t^\phi, \ \ \tilde{y}_0 = y_0 \in \mathbb{R}^e.$$

We define the solution $y \in D^{p\text{-var}}([0,T], \mathbb{R}^e)$ to the canonical RDE

(6) $$dy_t = V(y_t) \diamond d(\mathbf{x}_t, \phi), \ \ y_0 \in \mathbb{R}^e,$$

by $y = \tilde{y} \circ \tau_\mathbf{x}$ (where $\tau_\mathbf{x}$ is given by (2)).
    In the particular case that ϕ is the log-linear path function from Example 2.4, we denote the RDE simply by

$$dy_t = V(y_t) \diamond d\mathbf{x}_t, \ \ y_0 \in \mathbb{R}^e.$$

REMARK 3.2.   While the continuous RDE solution $\tilde{y}$ clearly depends (up to reparametrization) on the choice of representative $(\mathbf{x}, \phi) \in \mathscr{D}^{p\text{-var}}$ as well as the choice of (r_k), it is easy to see that y is independent of these choices, and is therefore well-defined on $\mathscr{D}^{p\text{-var}}$.

REMARK 3.3. Observe that every continuity point t of $\mathbf{x}$ is also a continuity point $\tau_{\mathbf{x}}$, and is therefore also a continuity point of y.

REMARK 3.4. For the log-linear path function ϕ, the solution y agrees precisely with the solution to the rough canonical equation considered in [19, Def. 37, Thm. 38]. Furthermore, we shall see in Section 4 that all semimartingales admit a canonical lift to a càdlàg geometric p-rough path, and that, for the log-linear path function, the solution y agrees with the Marcus solution of the associated SDE (see Proposition 4.16 below).

3.2. *Skorokhod-type p-variation metric.* We now introduce a metric $\alpha_{p\text{-var}}$ on $\mathscr{D}^{p\text{-var}}$ for which the RDE solution map is locally Lipschitz continuous. We first define an auxiliary metric $\sigma_{p\text{-var}}$ on $D^{p\text{-var}}$ which is independent of any path function. Recall the inhomogeneous p-variation metric $\rho_{p\text{-var}}$ from Section 2.2.

DEFINITION 3.5. *For $p \geq 1$ and $\mathbf{x}, \bar{\mathbf{x}} \in D^{p\text{-var}}$, define*

$$\sigma_{p\text{-var}}(\mathbf{x}, \bar{\mathbf{x}}) = \inf_{\lambda \in \Lambda} \max\{|\lambda|, \rho_{p\text{-var}}(\mathbf{x} \circ \lambda, \bar{\mathbf{x}})\}.$$

REMARK 3.6 (Topologies induced by $\sigma_{p\text{-var}}$ and $\rho_{p\text{-var}}$). Note that $\sigma_{1\text{-var}}$ and $\rho_{1\text{-var}}$ induce the same topology on $C^{0,1\text{-var}}$. Indeed, it is sufficient to show that $\rho_{1\text{-var}}(\mathbf{x}, \mathbf{x} \circ \lambda^n) \to 0$ for all $\mathbf{x} \in C^{0,1\text{-var}}$ and $|\lambda^n| \to 0$, which follows from writing $\mathbf{x}_t = \int_0^t \dot{\mathbf{x}}_s ds$ and applying dominated convergence.

Furthermore, for $p' > p \geq 1$, $\sigma_{p'\text{-var}}$ and $\rho_{p'\text{-var}}$ induce the same topology on $C^{p\text{-var}}$. Indeed, it again suffices to show that $\rho_{p'\text{-var}}(\mathbf{x}, \mathbf{x} \circ \lambda^n) \to 0$ for all $\mathbf{x} \in C^{p\text{-var}}$ and $|\lambda^n| \to 0$, which follows from $d_{\infty;[0,T]}(\mathbf{x}, \mathbf{x} \circ \lambda^n) \to 0$ and interpolation [20, Lem. 8.16].

However, note that $\sigma_{1\text{-var}}$ and $\rho_{1\text{-var}}$ do not induce the same topology on $C^{1\text{-var}}$. This can be seen from the fact that $C^{0,1\text{-var}}$ is dense in $C^{1\text{-var}}$ under $\sigma_{1\text{-var}}$ (see Proposition 3.10 part (iii)), or from the following direct example: consider the $\mathbb{R}$-valued Cantor function $\mathbf{x}_t = \mu([0,t])$, where μ is the Cantor distribution, and shifts $\mathbf{x}_t^n = \mathbf{x}_{t-\alpha_n}$ (with $\mathbf{x}_t^n = 0$ for $t \in [0, \alpha_n]$), where $\alpha_n \downarrow 0$. Clearly $\sigma_{1\text{-var}}(\mathbf{x}, \mathbf{x}^n) \leq \alpha_n \vee \|\mathbf{x}\|_{1\text{-var};[0,\alpha_n]} \to 0$. However, choosing α_n irrational, one can show that μ and $\mu(\cdot - \alpha_n)$ are mutually singular measures (see, e.g., [13]), so that $\rho_{1\text{-var}}(\mathbf{x}, \mathbf{x}^n) = \|\mathbf{x}^n\|_{1\text{-var};[0,1]} + \|\mathbf{x}\|_{1\text{-var};[0,1]} \to 2$.

We note that for the case $1 = N \leq p < 2$, the metric $\sigma_{p\text{-var}}$ already appears in the works of Simon [44] and Williams [46] where in particular a continuity statement for RDE solutions in the Young regime appears in terms of $\sigma_{p\text{-var}}$ (cf. Remark 3.19).

A drawback of the metric $\sigma_{p\text{-var}}$ is that the space of continuous rough paths $C^{p\text{-var}}$ is closed under $\sigma_{p\text{-var}}$. In particular, this implies that $\sigma_{p\text{-var}}$ is unable to describe situations in which continuous drivers approximate a discontinuous one (e.g. the Wong–Zakai theorem in [31]). We are thus motivated to introduce the following metric whose relation with $\sigma_{p\text{-var}}$ is analogous to that of α_∞ with σ_∞.

DEFINITION 3.7.   For $(\mathbf{x}, \phi), (\bar{\mathbf{x}}, \bar{\phi}) \in \mathscr{D}^{p\text{-var}}$, define the metric

$$\text{(7)} \qquad \alpha_{p\text{-var}}(\mathbf{x}, \bar{\mathbf{x}}) = \lim_{\delta \to 0} \sigma_{p\text{-var}}(\mathbf{x}^{\phi,\delta}, \bar{\mathbf{x}}^{\bar{\phi},\delta}),$$

where $\mathbf{x}^{\phi,\delta}$ is defined as at the start of Section 2.4.

REMARK 3.8.   Note that the limit (7) exists, is independent of the choice of series $\sum_{k=1}^{\infty} r_k$, and induces a well-defined metric on $\mathscr{D}^{p\text{-var}}$, all of which follows from the same argument as in Lemma 2.7 and Remark 2.13.

REMARK 3.9.   In light of Remark 3.6, it may seem possible to define an equivalent topology as that induced by $\alpha_{p\text{-var}}$ (at least on $C^{0,1\text{-var}}$) by replacing $\sigma_{p\text{-var}}$ by $\rho_{p\text{-var}}$ in (4) for the definition of $\alpha_{p\text{-var}}$ (and thus avoid introducing $\sigma_{p\text{-var}}$ altogether). However one can readily check that doing so will induce a completely different topology even on $C^{0,1\text{-var}}$ (in fact the same remark applies to replacing σ_∞ by d_∞ when defining α_∞ in Definition 2.8).

We record several basic properties of the metric space $(\mathscr{D}^{p\text{-var}}, \alpha_{p\text{-var}})$. For a path function ϕ, let $\mathscr{D}_\phi^{0,p\text{-var}} = \mathscr{D}_\phi^{0,p\text{-var}}([0, T], G^N(\mathbb{R}^d))$ denote the closure of $C^{0,1\text{-var}}$ in the metric space $(\mathscr{D}_\phi^{p\text{-var}}, \alpha_{p\text{-var}})$.

PROPOSITION 3.10.   Let $p \geq 1$ and ϕ a path function defined on a subset $J \subseteq G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$.

(i)   The space $(\mathscr{D}_\phi^{0,p\text{-var}}, \alpha_{p\text{-var}})$ is a separable metric space.

(ii)   It holds that $C^{p\text{-var}}$ is dense in $(\mathscr{D}_\phi^{p\text{-var}}, \alpha_{p\text{-var}})$.

(iii)   It holds that $\mathscr{D}_\phi^{0,1\text{-var}} = \mathscr{D}_\phi^{1\text{-var}}$.

(iv)   If $p > 1$, the closure of $C^{0,1\text{-var}}$ in $(C^{p\text{-var}}, \sigma_{p\text{-var}})$ is precisely $C^{0,p\text{-var}}$. In particular, $\mathscr{D}_\phi^{0,p\text{-var}} \subsetneq \mathscr{D}_\phi^{p\text{-var}}$.

(v)   For every $p' > p$, $\mathscr{D}_\phi^{p\text{-var}} \subsetneq \mathscr{D}_\phi^{0,p'\text{-var}}$.

PROOF. (i) Recall that $(C^{0,1\text{-var}}, \rho_{p\text{-var}})$ is a separable space, and therefore so is $(C^{0,1\text{-var}}, \sigma_{p\text{-var}})$ (see Remark 3.6). Since the metrics $\sigma_{p\text{-var}}$ and $\alpha_{p\text{-var}}$ coincide on $C^{0,1\text{-var}}$, it follows that $\mathscr{D}_\phi^{0,p\text{-var}}$ is separable.

(ii) For every $\mathbf{x} \in \mathscr{D}_\phi^{p\text{-var}}$ and $\delta > 0$, it holds that $\mathbf{x}^{\phi,\delta} \in C^{p\text{-var}}([0, T], G^N(\mathbb{R}^d))$. One can readily see that $\lim_{\delta \to 0} \alpha_{p\text{-var}}(\mathbf{x}, \mathbf{x}^{\phi,\delta}) = 0$, from which the claim follows.

(iii) By point (ii), it suffices to show that $C^{0,1\text{-var}}$ is dense in $C^{1\text{-var}}$ under $\sigma_{1\text{-var}}$. This in turn follows from the fact that any $\mathbf{x} \in C^{1\text{-var}}([s, t], G^N(\mathbb{R}^d))$ can be reparametrized to be in $C^{1\text{-Höl}}([s, t], G^N(\mathbb{R}^d)) \cong L^\infty([s, t], \mathbb{R}^d)$ (where the isometry is via the weak derivative) and thus lies in $C^{0,1\text{-var}}([s, t], G^N(\mathbb{R}^d)) \cong L^1([s, t], \mathbb{R}^d)$.

(iv) Recall by Wiener's characterization [20, Thm. 8.22] (which relies on $p > 1$) that $\mathbf{x} \in C^{0,p\text{-var}}$ if and only if

$$\lim_{\delta \to 0} \sup_{|\mathcal{D}| < \delta} \sum_{t_i \in \mathcal{D}} \|\mathbf{x}\|_{p\text{-var};[t_i,t_{i+1}]}^p = 0.$$

As a consequence, for any $\lambda \in \Lambda$, it holds that $\mathbf{x} \circ \lambda \in C^{0,p\text{-var}}$ if and only if $\mathbf{x} \in C^{0,p\text{-var}}$. Therefore, for any sequences $(\mathbf{x}^n)_{n \geq 1}$ in $C^{0,p\text{-var}}$ and $(\lambda^n)_{n \geq 1}$ in Λ for which $\rho_{p\text{-var}}(\mathbf{x}, \mathbf{x}^n \circ \lambda_n) \to 0$, it holds that $\mathbf{x} \in C^{0,p\text{-var}}$, from which the conclusion follows.

(v) Since $C^{p\text{-var}} \subsetneq C^{0,p'\text{-var}}$, the conclusion follows from (ii). $\qquad \square$

We now record an interpolation estimate which will be helpful later. It turns out to be simpler to state in terms of a homogeneous version of the distance $\alpha_{p\text{-var}}$. See [20, Ch. 8] for the definition and basic properties of the metrics d_0 and $d_{p\text{-var}}$. For $(\mathbf{x}, \phi), (\bar{\mathbf{x}}, \bar{\phi}) \in \mathscr{D}^{p\text{-var}}$ we define

$$\beta_{p\text{-var}}(\mathbf{x}, \bar{\mathbf{x}}) = \lim_{\delta \to 0} \inf_{\lambda \in \Lambda} |\lambda| \vee d_{p\text{-var}}(\mathbf{x}^{\phi,\delta} \circ \lambda, \bar{\mathbf{x}}^{\bar{\phi},\delta})$$

as well as

$$\alpha_0(\mathbf{x}, \bar{\mathbf{x}}) = \lim_{\delta \to 0} \inf_{\lambda \in \Lambda} |\lambda| \vee d_0(\mathbf{x}^{\phi,\delta} \circ \lambda, \bar{\mathbf{x}}^{\bar{\phi},\delta})$$

(which are well-defined metrics on $\mathscr{D}^{p\text{-var}}$ by the same argument as in Lemma 2.7 and Remark 2.13). Observe that the d_0/d_∞ estimate [20, Prop. 8.15] implies

$$\alpha_\infty(\mathbf{x}, \bar{\mathbf{x}}) \leq \alpha_0(\mathbf{x}, \bar{\mathbf{x}}) \leq C \max \left\{ \alpha_\infty(\mathbf{x}, \bar{\mathbf{x}}), \alpha_\infty(\mathbf{x}, \bar{\mathbf{x}})^{1/N} \left(\|\mathbf{x}^\phi\|_\infty + \|\bar{\mathbf{x}}^{\bar{\phi}}\|_\infty \right)^{1-1/N} \right\}.$$

Moreover, to move from $\beta_{p\text{-var}}$ to $\alpha_{p\text{-var}}$, it holds that the identity map

$$\text{id} : (\mathscr{D}^{p\text{-var}}, \beta_{p\text{-var}}) \leftrightarrow (\mathscr{D}^{p\text{-var}}, \alpha_{p\text{-var}})$$

is Lipschitz on bounded sets in the $\to$ direction, and $1/N$-Hölder on bounded sets in the $\leftarrow$ direction [20, Thm. 8.10].

Finally, the following result now follows directly from the usual interpolation estimate for the homogeneous metric $d_{p\text{-var}}$ [20, Lem. 8.16].

LEMMA 3.11. *Let $1 \le p < p'$. For all $(\mathbf{x}, \phi), (\bar{\mathbf{x}}, \bar{\phi}) \in \mathscr{D}^{p\text{-var}}$ it holds that*

$$\beta_{p'\text{-var}}(\mathbf{x}, \bar{\mathbf{x}}) \le \left(\|\mathbf{x}^\phi\|_{p\text{-var}} + \|\bar{\mathbf{x}}^{\bar{\phi}}\|_{p\text{-var}} \right)^{p'/p} \alpha_0(\mathbf{x}, \bar{\mathbf{x}})^{1 - p'/p}.$$

As a consequence, we obtain the following useful embedding result.

PROPOSITION 3.12. *Let $1 \le p < p'$ and ϕ a p-approximating, endpoint continuous path function defined on a subset $J \subset G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$. Then the identity map*

$$\mathrm{id} : (\mathscr{D}^{p\text{-var}}_\phi, \sigma_\infty) \to (\mathscr{D}^{p\text{-var}}_\phi, \alpha_{p'\text{-var}})$$

is uniformly continuous on sets of bounded p-variation.

PROOF. This is a combination of Remark 2.18 and Lemmas 2.16, 2.20, and 3.11. □

3.3. *Continuity of the solution map.* An advantage of the metric $\alpha_{p\text{-var}}$ is it allows us to directly carry over continuity statements about the classical (continuous) RDE solution map to the discontinuous setting. Recall the RDE (6)

$$dy_t = V(y_t) \diamond d(\mathbf{x}_t, \phi), \quad y_0 \in \mathbb{R}^e,$$

which is well-defined for any admissible pair $(\mathbf{x}, \phi) \in \mathscr{D}^{p\text{-var}}$.

Consider the driver-solution space $E := G^N(\mathbb{R}^d) \times \mathbb{R}^e$. Given that ϕ is defined on $J \subset G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$, we obtain a natural path function $\phi_{(V)}$ defined on the following (necessarily strict)subset of $E \times E$

$$W := \{((x, y), (\bar{x}, \bar{y})) \mid (x, \bar{x}) \in J, \pi_{(V)}(0, y; \phi(x, \bar{x}))_1 = \bar{y}\}$$

and which is given by

$$\phi_{(V)}((x, y), (\bar{x}, \bar{y}))_t = (\phi(x, \bar{x})_t, \pi_{(V)}(0, y; \phi(x, \bar{x}))_t),$$

where we recall $\pi_{(V)}$ from Section 2.2 is the solution map for the (continuous) RDE driven along V. Therefore, an admissible pair $(\mathbf{x}, \phi) \in \mathscr{D}^{p\text{-var}}$ yields an admissible pair $((\mathbf{x}, y), \phi_{(V)}) \in \mathscr{D}^{p\text{-var}}([0, T], E)$ as the solution to the RDE. The following is now a consequence of Lyons' classical rough path universal limit theorem.

THEOREM 3.13 (Continuity of solution map). *For vector fields $V = (V_1, \ldots, V_d)$ in $\mathrm{Lip}^{\gamma+m-1}(\mathbb{R}^e)$ with $\gamma > p$ and $m \ge 1$, the solution map of the RDE (6)*

$$\mathbb{R}^e \times (\mathscr{D}^{p\text{-var}}, \alpha_{p\text{-var}}) \to (\mathscr{D}^{p\text{-var}}([0, T], E), \alpha_{p\text{-var}})$$

$$(y_0, (\mathbf{x}, \phi)) \mapsto ((\mathbf{x}, y), \phi_{(V)})$$

is locally Lipschitz. In particular,

$$\lim_{n\to\infty} |y_0^n - y_0| + \alpha_{p\text{-var}}(\mathbf{x}^n, \mathbf{x}) = 0 \tag{8}$$

implies that

$$\sup_n \|y^n\|_{p\text{-var}} < \infty, \quad \text{and} \quad \lim_{n\to\infty} y_t^n = y_t \quad \text{for all continuity points } t \text{ of } \mathbf{x}.$$

Furthermore the flow map

$$\left(\mathscr{D}_\phi^{p\text{-var}}, \alpha_{p\text{-var}}\right) \to \mathrm{Diff}^m(\mathbb{R}^e)$$

$$\mathbf{x} \mapsto U_{T\leftarrow 0}^{\mathbf{x}} \equiv U_{T\leftarrow 0}^{\mathbf{x}^\phi}$$

is uniformly continuous on sets of bounded p-variation (see Section 2.2 for the definition of $U_{T\leftarrow 0}^{\mathbf{x}^\phi}$).

REMARK 3.14. Note that one cannot replace $U_{T\leftarrow 0}^{\mathbf{x}^n}$ by $U_{t\leftarrow 0}^{\mathbf{x}^n}$ for any (fixed) $t \in [0, T]$ in the final statement of Theorem 3.13. This is a manifestation of the fact that $\alpha_{p\text{-var}}$ does not behave well under restrictions to subintervals $[0, t] \subset [0, T]$ (cf. Remark 2.11).

REMARK 3.15. Though we don't address this here, the Lipschitz constant appearing in Theorem 3.13 can be made to depend explicitly on V and the p-variation of $\mathbf{x}^\phi$.

REMARK 3.16. Note that in the second statement of Theorem 3.13, one cannot replace $\mathbf{x}$ by y in "for all continuity points t of $\mathbf{x}$". Note also that this type of convergence is the one considered in the Wong–Zakai theorem of [31].

PROOF. The claim that the solution map is locally Lipschitz and that the associated flows converge follows from the corresponding result for continuous rough paths (see, e.g., [20, Thm. 10.26]). To make this explicit, consider $\mathbf{x}$ with path function ϕ and then $z = (\mathbf{x}, y)$ with path function $\phi_{(V)}$. Write also $z^\delta = z^{\phi_{(V)}, \delta}$ and $\mathbf{x}^\delta = \mathbf{x}^{\phi, \delta}$. By definition

$$
\begin{aligned}
\alpha_{p\text{-var}}(z_1, z_2) &= \lim_{\delta\to 0} \inf_{\lambda\in\Lambda} \max\left\{ |\lambda|, \rho_{p\text{-var}}\left(z_1^\delta \circ \lambda, z_2^\delta\right) \right\} \\
&= \lim_{\delta\to 0} \inf_{\lambda\in\Lambda} \max\left\{ |\lambda|, \rho_{p\text{-var}}\left(\mathbf{x}_1^\delta \circ \lambda, \mathbf{x}_2^\delta\right) + \left| y_1^\delta \circ \lambda - y_2^\delta \right| \right\}.
\end{aligned}
$$

On the other hand, for $i = 1, 2$, we have $\mathbf{x}_i^\delta = \mathbf{x}_i^{\phi_i, \delta} \in C^{p\text{-var}}$. Write $\tilde{y}_i = \pi_{(V)}(0, y_{i,0}, \mathbf{x}_i^\delta)$ for the unique RDE solution to $dy = V(y)\, d\mathbf{x}_i^\delta$ and note that $\tilde{y}_i \equiv y_i^\delta$, by the very definition of the path function $\phi_{(V)}$. Note also that

$$y_i^\delta \circ \lambda \equiv \tilde{y}^i \circ \lambda = \pi_{(V)}(0, y_0^i, \mathbf{x}_i^\delta \circ \lambda)$$

for every time change $\lambda \in \Lambda$. It is then a direct consequence of the (local Lipschitz) continuity of the Itô-Lyons map, in the setting of continuous rough paths, that

$$\left| y_1^\delta \circ \lambda - y_2^\delta \right|_{p\text{-var};[0,T]} \lesssim \rho_{p\text{-var}} \left(\mathbf{x}_1^\delta \circ \lambda, \mathbf{x}_2^\delta \right) + |y_{1,0} - y_{2,0}| .$$

As a consequence,

$$\rho_{p\text{-var}} \left(z_1^\delta \circ \lambda, z_2^\delta \right) \lesssim \rho_{p\text{-var}} \left(\mathbf{x}_1^\delta \circ \lambda, \mathbf{x}_2^\delta \right) + |y_{1,0} - y_{2,0}| .$$

Finally, take $\lim_{\delta \to 0} \inf_{\lambda \in \Lambda} \max(|\lambda|, \cdot)$ on both sides to see that

$$\alpha_{p\text{-var}} (z_1, z_2) \lesssim \alpha_{p\text{-var}} (\mathbf{x}_1, \mathbf{x}_2) + |y_{1,0} - y_{2,0}| .$$

The claim of a.e. pointwise convergence follows from Lemma 2.12, while uniform continuity of $(\mathbf{x}, \phi) \mapsto U_{T \leftarrow 0}^{\mathbf{x}}$ follows as above (cf. [20, Thm. 11.12]).
$\square$

REMARK 3.17.   An important feature of solutions to continuous RDEs (6) is that they can be canonically treated as geometric p-rough paths $\mathbf{y} \in C^{p\text{-var}}([0, T], G^N(\mathbb{R}^e))$, and the solution map $(\mathbf{y}_0, \mathbf{x}) \mapsto \mathbf{y}$ remains locally Lipschitz for the metric $\rho_{p\text{-var}}$ (see [20, Sec. 10.4]). This allows one to use the solution $\mathbf{y}$ as the driving signal in a secondary RDE

$$(9) \qquad\qquad dz_t = W(z_t) d\mathbf{y}_t, \;\; z_0 \in \mathbb{R}^n.$$

At least with the notion of canonical RDEs considered in this article, we cannot expect this functorial nature to be completely preserved due to the fact that in general no path function ψ can be defined on $G^N(\mathbb{R}^e)$ to capture the information of how $\mathbf{y}$ traversed a jump $(\mathbf{y}_{t-}, \mathbf{y}_t)$ (which, in our context, clearly impacts the solution of (9)).

However, at the cost of retaining the original driving signal $\mathbf{x}$, we can readily solve the secondary RDE (9). Indeed, for Lip^γ vector fields $V = (V_1, \ldots, V_d)$ on $\mathbb{R}^e$ and $W = (W_1, \ldots, W_e)$ on $\mathbb{R}^n$, we consider the Lip^γ vector fields $U = (U_1, \ldots, U_d)$ on $\mathbb{R}^e \oplus \mathbb{R}^n$ defined by $U_i(y, z) = V_i(y) +$

$\sum_{k=1}^{e} W_k(z)V_i(y)^k$, where $V_i(y)^k$ is the k-th coordinate of $V_i(y) \in \mathbb{R}^e$. Then the natural solution to (9) is given by the larger RDE

$$(10) \qquad d(y_t, z_t) = U(y_t, z_t) \diamond d(\mathbf{x}_t, \phi), \quad (y_0, z_0) \in \mathbb{R}^e \oplus \mathbb{R}^n.$$

As a consistency check, one can readily see that, in the continuous setting, the second component of the solution to (10) coincides with the solution z_t of (9).

If we further assume that drivers converge in Skorokhod topology, then more can be said about convergence of RDE solutions.

PROPOSITION 3.18. *Let notation be as in Theorem 3.13.*

(i) *Suppose that (8) holds and that $\lim_{n\to\infty} \mathbf{x}^n = \mathbf{x}$ in the Skorokhod topology. Then $\lim_{n\to\infty} y^n = y$ in the Skorokhod topology.*
(ii) *Suppose that ϕ is an endpoint continuous, p-approximating path function defined on (a subset of) $G^N(\mathbb{R}^d) \times G^N(\mathbb{R}^d)$. Then on sets of bounded p-variation, the solution map*

$$\mathbb{R}^e \times \left(\mathscr{D}_\phi^{p\text{-var}}, \sigma_\infty \right) \to (D^{p\text{-var}}([0,T], \mathbb{R}^e), \sigma_\infty)$$

$$(y_0, \mathbf{x}) \mapsto y$$

is continuous.

PROOF. (i) By Theorem 3.13, it suffices to show that $(y^n)_{n\geq 1}$ is compact in the Skorokhod space $D([0,T], \mathbb{R}^e)$. Recall that for a Polish space E, a subset $A \subset D([0,T], E)$ is compact if and only if $\{\mathbf{y}_t \mid \mathbf{y} \in A, t \in [0,T]\}$ is compact and

$$\lim_{\varepsilon \to 0} \sup_{\mathbf{y} \in A} \omega_{\mathbf{y}}'(\varepsilon) = 0,$$

where

$$\omega_{\mathbf{y}}'(\varepsilon) := \inf_{|\mathcal{D}|_{\min} > \varepsilon} \max_{t_i \in \mathcal{D}} \sup_{s,t \in [t_i, t_{i+1})} d(\mathbf{y}_s, \mathbf{y}_t), \quad |\mathcal{D}|_{\min} := \min_{t_i \in \mathcal{D}} |t_{i+1} - t_i|,$$

see, e.g., [4, Thm. 12.3]. In particular, using that $\|\mathbf{x}^\phi\|_{p\text{-var}} < \infty$ and applying the first inequality in Lemma 2.16 to $\mathbf{x}$, we see that $\lim_{\varepsilon \to 0} \lim_{\delta \to 0} \omega_{\mathbf{x}^{\phi,\delta}}^{p\text{-var}}(\varepsilon) = 0$, where

$$\omega_{\mathbf{y}}^{p\text{-var}}(\varepsilon) := \inf_{|\mathcal{D}|_{\min} > \varepsilon} \max_{t_i \in \mathcal{D}} \|\mathbf{y}\|_{p\text{-var};[t_i, t_{i+1}]}.$$

It now follows from (8) that

$$\lim_{\varepsilon \to 0} \sup_n \lim_{\delta \to 0} \omega_{(\mathbf{x}^n)^{\phi^n, \delta}}^{p\text{-var}}(\varepsilon) = 0,$$

from which we see that $\lim_{\varepsilon \to 0} \sup_n \omega_{y^n}^{p\text{-var}}(\varepsilon) = 0$. Since $\omega_y'(\varepsilon) \leq \omega_y^{p\text{-var}}(\varepsilon)$, it follows that $(y^n)_{n \geq 1}$ is compact in $D([0,T], \mathbb{R}^e)$ as required.

(ii) This follows directly from taking $p < p' < \gamma$ and applying (i) and Proposition 3.12. $\qquad\square$

REMARK 3.19. We suspect that under suitable conditions on a fixed path function ϕ, the solution map $\mathscr{D}_\phi^{p\text{-var}} \to D^{p\text{-var}}([0,T], \mathbb{R}^e)$ remains locally uniformly (or even Lipschitz) continuous in the classical p-variation metric $\rho_{p\text{-var}}$ defined by (1) (and thus trivially for the metric $\sigma_{p\text{-var}}$). However there seems to be no easy way to derive this as a consequence of our main Theorem 3.13. We suspect that this could be done by carefully relating canonical with "Itô-type" *non-canonical* RDEs (see [19, Def. 37]), followed by a proper stability analysis of the latter. The study of such non-canonical equations was recently carried out in [21].

3.4. *Young pairing and translation operator.* In this section we extend the Young pairing $S_N(\mathbf{x}, h)$ and rough path translation operator $T_h(\mathbf{x})$ to the càdlàg setting.

Given a path (x, h) in $\mathbb{R}^{d+d'}$ (smooth), its canonical level-2 rough path lift is given by $((x, h), (\int x \otimes dx, \int x \otimes dh, \int h \otimes dx, \int h \otimes dh))$. This extends immediately to a p-rough path $\mathbf{x}$, for $p \in (2, 3)$, with $\int x \otimes dx$ replaced by the a priori level-2 information $\mathbf{x}^2$. For h of finite q-variation with $1/p + 1/q > 1$, all other cross-integrals remain defined. This is the Young pairing of a p-rough path $\mathbf{x}$ with a q-variation path h called $S_2(\mathbf{x}, h)$. The mapping $(\mathbf{x}, h) \mapsto S_2(\mathbf{x}, h)$ is continuous, and the general construction, for continuous (rough) paths, is found in [20, Sec. 9.4]. The important operation of translating rough paths in some h-direction, formally $\int (\mathbf{x} + h) \otimes d(\mathbf{x} + h)$, can be algebraically formulated in terms of the Young pairing and has found many applications [18].

We first illustrate the difficulty of *continuously* extending $S_N(\mathbf{x}, h)$ and $T_h(\mathbf{x})$ to càdlàg paths by showing that the addition map $(\mathbf{x}, h) \mapsto \mathbf{x} + h$ is not continuous as a map $\mathscr{D}^{1\text{-var}}([0,T], \mathbb{R}^d) \times \mathscr{D}^{1\text{-var}}([0,T], \mathbb{R}^d) \to \mathscr{D}([0,T], \mathbb{R}^d)$ where we equip $\mathscr{D}^{1\text{-var}}([0,T], \mathbb{R}^d)$ and $\mathscr{D}([0,T], \mathbb{R}^d)$ with $\alpha_{1\text{-var}}$ and α_∞ respectively (which is reminiscent of the fact that $D([0,T], \mathbb{R}^d)$ is not a topological vector space).

EXAMPLE 3.20. Consider the sequences $(\mathbf{x}^n)_{n \geq 1}$, $(h^n)_{n \geq 1}$, and $(\bar{h}^n)_{n \geq 1}$, of continuous piecewise linear paths from $[0, 2]$ to $\mathbb{R}^2$, where $\mathbf{x}^n$ is constant on $[0, 1 - 1/n]$, moves linearly from 0 to e_1 over $[1 - 1/n, 1]$ and remains constant on $[1, 2]$, while h^n and $\bar{h}^n$ do the same, except from 0 to e_2 and over the intervals $[1 - 2/n, 1 - 1/n]$ and $[1, 1 + 1/n]$ respectively.

One can see that $\mathbf{x}^n$ converges to $(\mathbf{1}_{\{t\geq 1\}}e_1, \phi)$ in $\alpha_{1\text{-var}}$, and h^n and $\bar{h}^n$ both converge to $(\mathbf{1}_{\{t\geq 1\}}e_2, \phi)$ in $\alpha_{1\text{-var}}$, where ϕ is the linear path function on $\mathbb{R}^2$. However $\mathbf{x}^n + h^n$ converges to $(\mathbf{1}_{\{t\geq 1\}}(e_1+e_2), \phi_{2,1})$ while $\mathbf{x}^n + \bar{h}^n$ converges to $(\mathbf{1}_{\{t\geq 1\}}(e_1+e_2), \phi_{1,2})$ in α_∞, where $\phi_{i,j}$ is the "Hoff" path function which moves first in the i-th coordinate, and then in the j-th coordinate.

Note that the limiting path in the above example is unambiguously defined (namely $\mathbf{1}_{\{t\geq 1\}}(e_1 + e_2)$) whereas the corresponding path function is not. In the following definition we circumvent this problem by choosing a priori the (log-)linear path function.

DEFINITION 3.21. *Let $p \geq 1$ and $1 \leq q < 2$ such that $1/p + 1/q > 1$. For integers $d, d' \geq 1$, let ϕ and $\bar{\phi}$ be the log-linear path functions on $G^N(\mathbb{R}^{d+d'})$ and $G^N(\mathbb{R}^d) \times \mathbb{R}^{d'}$ respectively.*

For $\mathbf{x} \in D^{p\text{-var}}([0,T], G^N(\mathbb{R}^d))$ and $h \in D^{q\text{-var}}([0,T], \mathbb{R}^{d'})$, consider the continuous $G^N(\mathbb{R}^d) \times \mathbb{R}^{d'}$-valued path $(\hat{\mathbf{x}}, \hat{h}) = (\mathbf{x}, h)^{\bar{\phi}}$. We define $(S_N(\mathbf{x}, h), \phi) \in \mathscr{D}^{p\text{-var}}([0,T], G^N(\mathbb{R}^{d+d'}))$ by

$$(11) \qquad S_N(\mathbf{x}, h) = S_N(\hat{\mathbf{x}}, \hat{h}) \circ \tau_{(\mathbf{x},h)},$$

where we recall that $\tau_{(\mathbf{x},h)}$ is defined by (2) for the càdlàg path $(\mathbf{x}, h)$. In the case that $d = d'$, we let $\hat{\phi}$ denote the log-linear path function on $G^N(\mathbb{R}^d)$ and define $(T_h(\mathbf{x}), \hat{\phi}) \in \mathscr{D}^{p\text{-var}}([0,T], G^N(\mathbb{R}^d))$ by $T_h(\mathbf{x}) = T_{\hat{h}}(\hat{\mathbf{x}}) \circ \tau_{(\mathbf{x},h)}$.

REMARK 3.22. Due to the choice of log-linear path function, note that $\hat{\mathbf{x}}$ and $\hat{h}$ have finite p- and q-variation respectively, so that $S_N(\hat{\mathbf{x}}, \hat{h})$ and $T_{\hat{h}}(\hat{\mathbf{x}})$ are well-defined as continuous rough paths. Moreover, we have the relation $T_h(\mathbf{x}) = \mathbf{plus}\,(S_N(\mathbf{x}, h))$ (see [20, Thm. 9.33]).

We now record a simple result in the case $N = 2$ which will be helpful in the remainder of the paper. Recall from Definition 2.2 that $\mathbf{x} \in D^{p\text{-var}}([0,T], G^2(\mathbb{R}^d))$ is called Marcus-like if $\log(\Delta \mathbf{x}_t) \in \mathbb{R}^d \oplus \{0\} \subset \mathfrak{g}^2(\mathbb{R}^d)$, i.e., $\mathbf{x} = \exp(x, A)$ where $\exp$ is taken in $T^2(\mathbb{R}^d)$ and $A_{t-,t} = 0$ for all $t \in [0,T]$.

PROPOSITION 3.23. *Let notation be as in Definition 3.21. Suppose that $\mathbf{x} \in D^{p\text{-var}}([0,T], G^2(\mathbb{R}^d))$ is Marcus-like. Then for every $h \in D^{q\text{-var}}([0,T], \mathbb{R}^{d'})$ it holds that $S_2(\mathbf{x}, h)$ is a weakly geometric Marcus-like p-rough path with the anti-symmetric part of its second level determined for $i \in \{1, \ldots, d\}$ and $i' \in \{1 \ldots, d'\}$ by*

$$(12) \qquad S_2(\mathbf{x}, h)_{s,t}^{i,d+i'} - S_2(\mathbf{x}, h)_{s,t}^{d+i',i} = \int_{(s,t]} x_{s,u-}^i \otimes dh_u^{i'} - h_{s,u-}^{i'} \otimes dx_u^i,$$

where the integrals are well-defined Young integrals (all other components of the anti-symmetric part are given canonically by $\mathbf{x}$ and h). Furthermore

$$(13) \quad \|S_2(\mathbf{x}, h)^\phi\|_{p\text{-var}} \vee \|(T_h(\mathbf{x})^{\hat{\phi}}\|_{p\text{-var}} \lesssim \|(\mathbf{x}, h)^{\bar{\phi}}\|_{p\text{-var}} \lesssim \|\mathbf{x}\|_{p\text{-var}} + \|h\|_{q\text{-var}}.$$

PROOF. By definition of $\mathbf{Z} := (Z, \mathbb{Z}) := S_2(\mathbf{x}, h)$ and the choice of log-linear path functions, we see that $\mathbf{Z}$ is indeed Marcus-like. Observe that $\tilde{\mathbf{Z}} := S_2(\hat{\mathbf{x}}, \hat{h})$ satisfies the linear RDE in $T^N(\mathbb{R}^{d+d'})$ driven by itself $d\tilde{\mathbf{Z}}_t = \tilde{\mathbf{Z}}_t \otimes d\tilde{\mathbf{Z}}$. It follows by [19, Thm. 38] and the definition of $\tau_{(\mathbf{x}, h)}$ that $\mathbf{Z}$ satisfies the canonical rough equation $d\mathbf{Z}_t = \mathbf{Z}_t \otimes \diamond d\mathbf{Z}_t$ (in the sense of [19, Def. 37]), so that

$$\mathbf{Z}_t = \mathbf{Z}_0 + \int_0^t \mathbf{Z}_{s-} \otimes d\mathbf{Z}_s + \sum_{0 < s \leq t} \mathbf{Z}_{s-} \otimes \exp(\Delta Z_s) - \mathbf{Z}_{s-} - \mathbf{Z}_{s-} \otimes \Delta Z_s$$

$$= 1_{G^N(\mathbb{R}^{d+d'})} + \int_0^t \mathbf{Z}_{s-} \otimes d\mathbf{Z}_s + \sum_{0 < s \leq t} \frac{1}{2} (\Delta Z_s)^{\otimes 2},$$

where we have used that $\mathbf{Z}$ is Marcus-like. In particular, the last term on the RHS only contributes to the symmetric part of $\mathbb{Z}$, so that (12) follows by identifying $\int_0^t \mathbf{Z}_{s-} \otimes d\mathbf{Z}_s$ in the appropriate components with well-defined Young integrals. The inequality (13) follows from standard (continuous) rough path estimates [20, Thm. 9.26, 9.33] along with the fact that $\mathbf{x}$ is Marcus-like (so that upon appropriately restricting domains, we may assume ϕ, $\bar{\phi}$, and $\hat{\phi}$ are 1-approximating). □

4. General multidimensional semimartingales as rough paths.

4.1. *Enhanced p-variation BDG inequality.* The main result of this subsection is the BDG inequality Theorem 4.7 for enhanced càdlàg local martingales. The proof largely follows a classical argument found in [36, 20] with the exception of Lemma 4.2 which constitutes our main novel input.

Let X be an $\mathbb{R}^d$-valued semimartingale with $X_0 = 0$ and $(\mathbb{R}^d)^{\otimes 2}$-valued bracket $[X]$. Consider the $\mathfrak{g}^{(2)}(\mathbb{R}^d) = \mathfrak{so}(\mathbb{R}^d)$-valued ("area") process

$$(14) \qquad A_t^{i,j} = \frac{1}{2} \int_0^t X_{r-}^i dX_r^j - X_{r-}^j dX_r^i$$

as an Itô integral, and the "Marcus lift" (terminology from [19]) of X to a $G^2(\mathbb{R}^d)$-valued process given by

$$\mathbf{X}_t := \exp(X_t + A_t),$$

where exp is taken in the truncated tensor algebra $T^2(\mathbb{R}^d)$.

Note that, if $X = Y + K$ where Y is (another) semimartingale (w.r.t. the same filtration) and K is adapted, of bounded variation, then the respective Marcus lifts of X and Y are precisely related by the translation operator; that is, $\mathbf{X} = T_K\mathbf{Y}$. This can be seen by combining (12) with the fact that the second level of $\mathbf{X}$ is given precisely by the Marcus canonical integral $\int X \otimes \diamond dX$; see [19, Prop. 16] for details.

Recall that a function $F : [0, \infty) \to [0, \infty)$ is called *moderate* if it is continuous and increasing, $F(x) = 0$ if and only if $x = 0$, and there exists $c > 0$ such that $F(2x) \leq cF(x)$ for all $x > 0$.

LEMMA 4.1 (Uniform enhanced BDG inequality). *For any convex moderate function F there exist $c, C > 0$ such that for any $\mathbb{R}^d$-valued local martingale X*

$$c\mathbb{E}\left[F\left(|[X]_\infty|\right)\right] \leq \mathbb{E}\left[\sup_{s,t \geq 0} F\left(\|\mathbf{X}_{s,t}\|^2\right)\right] \leq C\mathbb{E}\left[F\left(|[X]_\infty|\right)\right].$$

PROOF. For a process M, denote $M_t^* := \sup_{0 \leq s \leq t}|M_s|$. Following the proof of [20, Thm. 14.8], it suffices to show that

$$\mathbb{E}\left[F(|A_\infty^*|)\right] \leq c_2\mathbb{E}\left[F(|[X]_\infty|)\right].$$

However we have

$$|[A]_\infty|^{1/2} \leq c_1|X_\infty^*|\,|[X]_\infty|^{1/2} \leq c_1(|X_\infty^*|^2 + |[X]_\infty|),$$

so one can apply the classical BDG inequality (e.g., [35, Thm. 2.1]) to A and F, and to X and $F(|\cdot|^2)$, which is also a convex moderate function, to obtain

$$\mathbb{E}\left[F(|A_\infty^*|)\right] \leq c_2\mathbb{E}\left[F\left(|[A]_\infty|^{1/2}\right)\right] \leq c_3\mathbb{E}\left[F(|X_\infty^*|^2) + F(|[X]_\infty|)\right] \leq c_4\mathbb{E}\left[F(|[X]_\infty|)\right].$$

$\square$

The following lemma is the crucial step in establishing finite p-variation of the lift of a local martingale.

LEMMA 4.2 (Interpolation). *For every $2 < q < p < r$ there exists $C = C(p, q, r)$ such that for every $\mathbb{R}^d$-valued local martingale X*

$$\mathbb{E}\left[\|\mathbf{X}\|_{p\text{-var}}^p\right] \leq C\mathbb{E}\left[|[X]_\infty|^{q/2} + |[X]_\infty|^{r/2}\right].$$

PROOF. For $\delta > 0$ define the increasing sequence of stopping times $(\tau_j^\delta)_{j=0}^\infty$ by $\tau_0^\delta = 0$ and for $j \geq 1$

$$\tau_j^\delta = \inf \left\{ t \geq \tau_{j-1}^\delta \mid \sup_{u,v \in [\tau_{j-1}^b, t]} d(\mathbf{X}_u, \mathbf{X}_v) > \delta \right\}$$

(where inf of the empty set is ∞). Define further $\nu(\delta) := \inf\{j \geq 0 \mid \tau_j^\delta = \infty\} - 1$. Observe that (cf. [9, p. 10])

$$(15) \qquad \qquad \|\mathbf{X}\|_{p\text{-var}}^p \leq \sum_{k=-\infty}^{\infty} 2^{p(k+1)} \nu(2^k).$$

Fix $\delta > 0$ and denote $\tau_j := \tau_j^\delta$. For $j = 0, 1, \ldots$ consider the sequence of local martingales $Y_t^j := X_{(\tau_j + t) \wedge \tau_{j+1}} - X_{\tau_j}$. Denote by $\mathbf{Y}_t^j$ the lift of Y_t^j, which coincides with $\mathbf{X}_{\tau_j, (\tau_j + t) \wedge \tau_{j+1}}$.

It holds that

$$\sum_{j=0}^{\infty} [Y^j]_\infty = \sum_{j=0}^{\nu(\delta)} [Y^j]_\infty = [X]_\infty,$$

and moreover $\sup_{s,t \geq 0} \|\mathbf{Y}_{s,t}^j\| \geq \delta$ for all $j = 0, \ldots, \nu(\delta) - 1$. Thus by the uniform enhanced BDG inequality (Lemma 4.1) with $F = |\cdot|^\alpha$, $\alpha \geq 1$,

$$\mathbb{E}\left[|[X]_\infty|^\alpha \right] \geq \mathbb{E}\left[\sum_{j=0}^{\infty} \left| [Y^j]_\infty \right|^\alpha \right] \geq c_\alpha \mathbb{E}\left[\sum_{j=0}^{\infty} \sup_{s,t \geq 0} \|\mathbf{Y}_{s,t}^j\|^{2\alpha} \right] \geq c_\alpha \delta^{2\alpha} \mathbb{E}\left[\nu(\delta) \right].$$

It follows that for $2 \leq 2\alpha < p$

$$\mathbb{E}\left[\sum_{k \leq 0} 2^{p(k+1)} \nu(2^k) \right] \leq c_\alpha^{-1} \mathbb{E}\left[|[X]_\infty|^\alpha \right] \sum_{k \leq 0} 2^{p(k+1) - 2\alpha k} \leq C(\alpha, p) \mathbb{E}\left[|[X]_\infty|^\alpha \right],$$

and likewise for $2\beta > p$

$$\mathbb{E}\left[\sum_{k > 0} 2^{p(k+1)} \nu(2^k) \right] \leq C(\beta, p) \mathbb{E}\left[|[X]_\infty|^\beta \right].$$

Taking $2\alpha = q$ and $2\beta = r$, the conclusion follows from the estimate (15). $\quad\square$

COROLLARY 4.3. *For every $\mathbb{R}^d$-valued semimartingale X, $p > 2$, and $T > 0$, it holds that a.s.*

$$\|\mathbf{X}\|_{p\text{-var};[0,T]} < \infty.$$

PROOF. Note that X can be decomposed into a process K of finite variation and a local martingale L with jump sizes bounded by a positive constant (see, e.g., [25, Prop. 4.17, p. 42]). Denoting by $\mathbf{L}$ the lift of L, it follows from a localization argument and Lemma 4.2, that $\|\mathbf{L}\|_{p\text{-var};[0,T]} < \infty$ a.s.. The conclusion follows by observing that $\mathbf{X} = T_K(\mathbf{L})$. $\qquad\square$

LEMMA 4.4 (Chebyshev inequality). *For all $p > 2$, there exists a constant $A > 0$ such that for every $\mathbb{R}^d$-valued local martingale X and $\lambda > 0$,*

$$\mathbb{P}\big[\|\mathbf{X}\|_{p\text{-var}} > \lambda\big] \leq \frac{A}{\lambda^2}\mathbb{E}\left[|[X]_\infty|\right].$$

PROOF. This crucially uses that $\mathbf{X}$ has finite p-variation for every local martingale X, and follows in exactly the same manner as [36, Lem. 1] or [20, Lem. 14.10]. $\qquad\square$

LEMMA 4.5 (Burkholder [6] Lemma 7.1). *Suppose X and Y are non-negative random variables, F is a moderate function, and $\beta > 1$ and $\delta, \varepsilon, \gamma, \eta > 0$ such that $\gamma\varepsilon < 1$,*

$$F(\beta\lambda) \leq \gamma F(\lambda), \quad F(\delta^{-1}\lambda) \leq \eta F(\lambda), \quad \forall \lambda > 0,$$

and

$$\mathbb{P}\left[X > \beta\lambda, Y < \delta\lambda\right] \leq \varepsilon\mathbb{P}\left[X > \lambda\right], \quad \forall \lambda > 0.$$

Then

$$\mathbb{E}\left[F(X)\right] \leq \frac{\gamma\eta}{1 - \gamma\varepsilon}\mathbb{E}\left[F(Y)\right].$$

LEMMA 4.6. *Let X be an $\mathbb{R}^d$-valued local martingale and D an adapted, non-decreasing process such that a.s., $|\Delta X_t| \leq D_{t-}$ for all $t \geq 0$. Then for every moderate function F (not necessarily convex), there exists $C = C(F) > 0$ such that*

$$\mathbb{E}\big[F(\|\mathbf{X}\|_{p\text{-var}})\big] \leq C\mathbb{E}\big[F(|[X]_\infty|^{1/2} + D_\infty)\big].$$

PROOF. We follow closely the proof of [36, Prop. 2] and [20, Thm. 14.12]. Since $\mathbf{X}$ is Marcus-like, i.e., $\log(\Delta\mathbf{X}_t) \in \mathbb{R}^d$, there exists a constant $c > 0$ such that for all $t \geq 0$

(16) $$\|\Delta\mathbf{X}_t\| = c|\Delta X_t| \leq c\,|[X]_t|^{1/2}.$$

Let $\delta > 0$, $\beta > c\delta + 1$, $\lambda > 0$, and define the stopping times

$$T = \inf\left\{t \geq 0 \mid \|\mathbf{X}\|_{p\text{-var};[0,t]} > \beta\lambda\right\},$$
$$S = \inf\left\{t \geq 0 \mid \|\mathbf{X}\|_{p\text{-var};[0,t]} > \lambda\right\},$$
$$R = \inf\left\{t \geq 0 \mid D_t \vee |[X]_t|^{1/2} > \delta\lambda\right\}.$$

Define the local martingale $N_t = X_{(t+S)\wedge R} - X_{S\wedge R}$ with lift $\mathbf{N}$ and note that

$$\|\mathbf{N}\|_{p\text{-var}} \geq \|\mathbf{X}\|_{p\text{-var};[0,R]} - \|\mathbf{X}\|_{p\text{-var};[0,R\wedge S]}.$$

On the event $\{T < \infty, R = \infty\}$, we have $\Delta X_S \leq D_{S-} \leq \delta\lambda$, and so from (16)

$$\|\mathbf{N}\|_{p\text{-var}} \geq \beta\lambda - \lambda - \|\Delta \mathbf{X}_S\| \geq (\beta - 1 - c\delta)\lambda.$$

By Lemma 4.4, it follows that

$$\mathbb{P}\left[T < \infty, R = \infty\right] \leq \mathbb{P}\left[\|\mathbf{N}\|_{p\text{-var}} > (\beta - c\delta - 1)\lambda\right] \leq \frac{A}{(\beta - c\delta - 1)^2\lambda^2}\mathbb{E}\left[|[N]_\infty|\right].$$

On the event $\{S = \infty\}$, it holds that $N \equiv 0$, whilst on $\{S < \infty\}$, we have $D_{R-} \leq \delta\lambda$ and thus

$$|[N]_\infty| = |[X]_R - [X]_{R\wedge S}| \leq |[X]_{R-}| + |\Delta X_R|^2 \leq 2\delta^2\lambda^2.$$

It follows that

$$\mathbb{E}\left[|[N]_\infty|\right] \leq 2\delta^2\lambda^2\mathbb{P}\left[S < \infty\right],$$

and thus we have for all $\lambda > 0$

$$\mathbb{P}\left[\|\mathbf{X}\|_{p\text{-var}} > \beta\lambda, D_\infty \vee [X]_\infty^{1/2} \leq \delta\lambda\right] \leq \frac{2A\delta^2}{(\beta - c\delta - 1)^2}\mathbb{P}\left[\|\mathbf{X}\|_{p\text{-var}} > \lambda\right].$$

The conclusion now follows by applying Lemma 4.5. $\square$

THEOREM 4.7 (p-variation rough path BDG).   *For every convex moderate function F and $p > 2$ there exists $c, C > 0$ such that for every $\mathbb{R}^d$-valued local martingale X*

$$c\mathbb{E}\left[F\left(|[X]_\infty|^{1/2}\right)\right] \leq \mathbb{E}\left[F\left(\|\mathbf{X}\|_{p\text{-var}}\right)\right] \leq C\mathbb{E}\left[F\left(|[X]_\infty|^{1/2}\right)\right].$$

PROOF. This again follows very closely the proof of [36, Prop. 2]. We may suppose $\mathbb{E}[|[X]_\infty|^{1/2}] < \infty$ (otherwise all concerned quantities are infinite). Let $D_t := \sup_{0 \leq s \leq t} |\Delta X_t|$ and $K_t^1 := \sum_{s \leq t} \Delta X_s \mathbf{1}_{\{|\Delta X_s| \geq 2D_{s-}\}}$. Note that K^1 is of integrable variation since

$$|K^1|_{1\text{-var}} \leq 4D_\infty \leq 4\,|[X]_\infty|^{1/2},$$

so there exists a unique previsible process K^2 such that $K^1 - K^2$ is a martingale (K^2 is the dual previsible projection of K^1 [43, Thm. 21.4] and is a special case of the Doob–Meyer decomposition). Finally, define the martingale $L := X - (K^1 - K^2)$; recall that $X = (K^1 - K^2) + L$ is called the Davis decomposition of X.

Observe that $|\Delta L_t| \le 4D_{t-}$. Indeed, $|\Delta(L-K^2)_t| \le 2D_{t-}$ by construction, and, if T is a stopping time, then either T is totally inaccessible, in which case $\Delta K_T^2 = 0$ since K^2 is previsible, or T is previsible, in which case $|\Delta K_T^2| = |\mathbb{E}\left[\Delta(K^2 - L)_T \mid \mathcal{F}_{T-}\right]| \le \mathbb{E}\left[2D_{T-} \mid \mathcal{F}_{T-}\right] = 2D_{T-}$, where $\mathcal{F}$ is the underlying filtration (cf. [42, p. 80-81]). Hence, by Lemma 4.6, we have

$$\mathbb{E}\left[F\left(\|\mathbf{L}\|_{p\text{-var}}\right)\right] \le C_1 \mathbb{E}\left[F\left(|[L]_\infty|^{1/2} + D_\infty\right)\right].$$

Since $D_\infty \le |[X]_\infty|^{1/2}$ and

$$|[L]_\infty|^{1/2} \le |[X]_\infty|^{1/2} + |[K]_\infty|^{1/2} \le |[X]_\infty|^{1/2} + |K|_{1\text{-var}},$$

we have
$$\mathbb{E}\left[F\left(\|\mathbf{L}\|_{p\text{-var}}\right)\right] \le C_2 \mathbb{E}\left[F\left(|[X]_\infty|^{1/2} + |K|_{1\text{-var}}\right)\right].$$

Furthermore, since F is convex, it follows from the Garsia-Neveu lemma (using the argument provided by [36, p. 306]) that

$$\mathbb{E}\left[F\left(|K^2|_{1\text{-var}}\right)\right] \le C_3 \mathbb{E}\left[F\left(|K^1|_{1\text{-var}}\right)\right],$$

and thus
$$\mathbb{E}\left[F\left(|K|_{1\text{-var}}\right)\right] \le C_4 \mathbb{E}\left[F\left(|[X]_\infty|^{1/2}\right)\right].$$

Finally, as $\mathbf{X} = T_K(\mathbf{L})$, we obtain

$$\mathbb{E}\left[F\left(\|\mathbf{X}\|_{p\text{-var}}\right)\right] \le C_5 \mathbb{E}\left[F\left(|K|_{1\text{-var}} + \|\mathbf{L}\|_{p\text{-var}}\right)\right] \le C_6 \mathbb{E}\left[F\left(|[X]_\infty|^{1/2}\right)\right].$$

$\square$

It was seen in [19, Thm. 20] that every (level M) càdlàg p-rough path $\mathbf{X}$ admits a unique minimal jump extension $\bar{\mathbf{X}}$ as a level-N rough path, $N \ge M$. (This generalizes Lyons' fundamental extension theorem to the jump setting.) Moreover, it is clear from the proof of [19, Thm. 20]

$$\|\bar{\mathbf{X}}\|_{p\text{-var}} \lesssim \|\mathbf{X}\|_{p\text{-var}}$$

(with a multiplicative constant that depends on N, M). Applied with $M = 2$ and Marcus lift $\mathbf{X} = \mathbf{X}(\omega)$ of an $\mathbb{R}^d$-valued local martingale, we obtain the following form of the BDG inequality.

COROLLARY 4.8 (p-variation level-N rough path BDG). *For every $N \geq 1$, convex moderate function F, and $p > 2$, there exists $c, C > 0$ such that for every $\mathbb{R}^d$-valued local martingale X*

$$c\mathbb{E}\big[F\big(\,|[X]_\infty|^{1/2}\,\big)\big] \leq \mathbb{E}\big[F\big(\|\bar{\mathbf{X}}\|_{p\text{-var}}\big)\big] \leq C\mathbb{E}\big[F\big(\,|[X]_\infty|^{1/2}\,\big)\big].$$

This is a useful result in the study of expected signatures, which is, loosely speaking, the study of $\mathbb{E}\big[\bar{\mathbf{X}}_{0,T}\big]$, with $\bar{\mathbf{X}}_{0,T} \in G^N(\mathbb{R}^d) \subset T^N(\mathbb{R}^d)$ and component-wise expectation in the linear space $T^N(\mathbb{R}^d)$. Since the norm of $\pi_m\big(\bar{\mathbf{X}}_{0,T}\big)$, the projection to $\big(\mathbb{R}^d\big)^{\otimes m}$, is bounded (up to a constant) by $\|\bar{\mathbf{X}}\|_{p\text{-var};[0,T]}^m$, we see that the very existence of the expected signature is guaranteed by the existence of all moments of $[X]_{0,T}$. In a Lévy setting with triplet (a, b, K), this clearly holds whenever $K\,(dy)\,1_{[|y|>1]}$ has moments of all orders. One can also apply this with a stopping time $T = T(\omega)$, e.g., the exit time of Brownian motion from a bounded domain. In either case, the expected signature is seen to exist (see also [19, Part III] and [37] for more on this).

A motivation for the study of expected signatures comes from one of the main results of [10, Sec. 6] which provides a solution to the moment problem for (random) signatures, i.e., determines conditions under which the sequence of expectations $(\mathbb{E}[\pi_m\bar{\mathbf{X}}_{0,T}])_{m\geq 0}$ uniquely determines the law of the full signature of $\bar{\mathbf{X}}$ (see [19, Thm. 54] where the moment problem was discussed for the Lévy case, and [7, 10] for other families of random geometric rough paths).

4.2. *Convergence of semimartingales and the UCV condition.* As an application of the BDG inequality, we obtain a convergence criterion for lifted local martingales in the rough path space $(\mathscr{D}_\phi^{p\text{-var}}([0,T], G^2(\mathbb{R}^d)), \alpha_{p\text{-var}})$ with a fixed path function ϕ, which is the main result of this subsection.

We first recall the uniformly controlled variation (UCV) condition for a sequence of semimartingales $(X^n)_{n\geq 1}$. For $X \in D([0,T], \mathbb{R}^d)$ and $\delta > 0$, we define

$$X_t^\delta = X_t - \sum_{s\leq t}(1 - \delta/|\Delta X_s|)^+ \Delta X_s.$$

Note that $X \mapsto X^\delta$ is a continuous function on the Skorokhod space and $\sup_{t\in[0,T]}|\Delta X_t^\delta| \leq 1$ with $\Delta X_t^\delta = \Delta X_t$ whenever $|\Delta X_t| \leq \delta$.

DEFINITION 4.9 (UCV, [33] Definition 7.5). *We say that a sequence of semimartingales $(X^n)_{n\geq 1}$ satisfies UCV if there exists $\delta > 0$ such that for all $\alpha > 0$ there exist decompositions $X^{n,\delta} = M^{n,\delta} + K^{n,\delta}$ and stopping times*

$\tau^{n,\alpha}$ such that for all $t \geq 0$

$$\sup_{n \geq 1} \mathbb{P}\left[\tau^{n,\alpha} \leq \alpha\right] \leq \frac{1}{\alpha} \text{ and } \sup_{n \geq 1} \mathbb{E}\left[M^{n,\delta}\right]_{t \wedge \tau^{n,\alpha}} + \left|K^{n,\delta}\right|_{1\text{-var};[0,t \wedge \tau^{n,\alpha}]}\right] < \infty.$$

Recall the following result of Kurtz–Protter [32, Thm. 2.2] (see also [33] Theorem 7.10 and p. 30).

THEOREM 4.10. *Let $X, (X^n)_{n \geq 1}, H, (H^n)_{n \geq 1}$ be càdlàg adapted processes (with respect to some filtrations $\mathcal{F}^n$). Suppose $(H^n, X^n)_{n \geq 1}$ converges in law (resp. in probability) to (H, X) in the Skorokhod topology as $n \to \infty$, and that $(X^n)_{n \geq 1}$ is a sequence of càdlàg semimartingales satisfying UCV. Then X is a semimartingale (with respect to some filtration $\mathcal{F}$) and $(H^n, X^n, \int_0^\cdot H^n_{t-} dX^n_t)$ converge in law (resp. in probability) to $(H, X, \int_0^\cdot H_{t-} dX_t)$ in the Skorokhod topology as $n \to \infty$.*

We can now state the main result which allows us to pass from convergence in the Skorokhod topology to convergence in rough path topology (see also Corollary 4.12).

THEOREM 4.11. *Let $(X^n)_{n \geq 1}$ be a sequence of semimartingales such that X^n converges in law (resp. in probability) to a semimartingale X in the Skorokhod topology. Suppose moreover that $(X^n)_{n \geq 1}$ satisfies the UCV condition. Then the lifted processes $(\mathbf{X}^n)_{n \geq 1}$ converge in law (resp. in probability) to the lifted process $\mathbf{X}$ in the Skorokhod space $D([0, T], G^2(\mathbb{R}^d))$. Moreover, for every $p > 2$, $(\|\mathbf{X}^n\|_{p\text{-var}})_{n \geq 1}$ is a tight collection of real random variables.*

PROOF. Since the stochastic area is given by the Itô integral (14), the convergence in law (resp. in probability) of $(\mathbf{X}^n)_{n \geq 1}$ to $\mathbf{X}$ is an immediate consequence of Theorem 4.10.

Let $\delta > 0$ for which we can apply the UCV condition to $(X^n)_{n \geq 1}$. We next claim that $(\|\mathbf{X}^{n,\delta}\|_{p\text{-var};[0,T]})_{n \geq 1}$ is tight. Indeed, for $\varepsilon > 0$ choose $\alpha > T$ so that $1/\alpha < \varepsilon/2$. Let $X^{n,\delta} = M^{n,\delta} + K^{n,\delta}$ be the decomposition from the UCV condition along with the stopping times $\tau^{n,\alpha}$. Then there exists $C > 0$ such that for all $n \geq 1$

$$\mathbb{P}\left[\|K^{n,\delta}\|_{1\text{-var};[0,T]} > C\right] \leq \mathbb{P}\left[\tau^{n,\alpha} \leq \alpha\right] + C^{-1}\mathbb{E}\left[\|K^{n,\delta}\|_{1\text{-var};[0,T \wedge \tau^{n,\delta}]}\right] < \varepsilon,$$

and

$$\sup_{n \geq 1} \mathbb{P}\left[\|\mathbf{M}^{n,\delta}\|_{p\text{-var};[0,T]} > C\right] \leq \sup_{n \geq 1} \mathbb{P}\left[\tau^{n,\alpha} \leq \alpha\right] + C^{-2}\mathbb{E}\left[\|\mathbf{M}^{n,\delta}\|^2_{p\text{-var};[0,T \wedge \tau^{n,\delta}]}\right]$$

$$\leq \sup_{n \geq 1} \mathbb{P}\left[\tau^{n,\alpha} \leq \alpha\right] + C^{-2}\mathbb{E}\left[|[M]_{T \wedge \tau^{n,\delta}}|\right] < \varepsilon,$$

where in the final line we used the enhanced BDG inequality Theorem 4.7. Using the fact that $\mathbf{X}^{n,\delta} = T_{K^{n,\delta}}(\mathbf{M}^{n,\delta})$ proves that $(\|\mathbf{X}^{n,\delta}\|_{p\text{-var};[0,T]})_{n\geq 1}$ is tight as claimed.

To conclude, observe that $\mathbf{X}^n = T_{L^n}(\mathbf{X}^{n,\delta})$ where $L^n := X^n - X^{n,\delta}$ is a process of bounded variation for which

$$|L^n|_{1\text{-var};[0,T]} \leq \sum_{|\Delta X_t^n|>\delta} |\Delta X_t^n|.$$

Since $(X^n)_{n\geq 1}$ is tight and $\sum_{|\Delta X_t^n|>\delta} |\Delta X_t^n|$ is a continuous function of X^n (for the Skorokhod topology), it follows that $(\|\mathbf{X}^n\|_{p\text{-var};[0,T]})_{n\geq 1}$ is tight as required. $\qquad\square$

COROLLARY 4.12. *Follow the notation of Theorem 4.11. Let $p > 2$ and ϕ an endpoint continuous, p-approximating path function defined on $J \subset G^2(\mathbb{R}^d) \times G^2(\mathbb{R}^d)$ such that $\mathbf{X}, \mathbf{X}^n \in \mathscr{D}_\phi([0,T], G^2(\mathbb{R}^d))$ a.s.. Then for every $p' > p$, $(\mathbf{X}^n, \phi) \to (\mathbf{X}, \phi)$ in law (resp. in probability) in the metric space $(\mathscr{D}_\phi^{p'\text{-var}}([0,T], G^2(\mathbb{R}^d)), \alpha_{p'\text{-var}})$.*

REMARK 4.13.   As we shall see in Proposition 4.15, a simple way to apply Corollary 4.12 is to assume that ϕ comes from the lift of a (left-invariant) path function $\phi : \mathbb{R}^d \to C^{q\text{-var}}([0,T], \mathbb{R}^d)$ (denoted by the same symbol) which is endpoint continuous and q-approximating for some $1 \leq q < 2$ (so that a canonical lift indeed exists), and does not create area, i.e.,

$$(17) \qquad \log S_2(\phi(x))_{0,1} = x \in \mathbb{R}^d \oplus \{0\} \subset \mathfrak{g}^2(\mathbb{R}^d).$$

Since $\mathbf{X}$ is Marcus-like, i.e., $\log(\Delta \mathbf{X}_t) \in \mathbb{R}^d \oplus \{0\}$, it indeed follows that $\mathbf{X} \in \mathscr{D}_\phi([0,T], G^2(\mathbb{R}^d))$ so that $\mathbf{X}^\phi$ is well defined (which corresponds to interpolating the jumps of $\mathbf{X}$ using ϕ); the same of courses applies to $\mathbf{X}^n$.

PROOF OF COROLLARY 4.12. Consider first the case of convergence in probability. By Theorem 4.11, $(\|\mathbf{X}^n\|_{p\text{-var}})_{n\geq 1}$ is tight, so for every $\varepsilon > 0$ we can find $R > 0$ such that

$$\sup_{n\geq 1} \mathbb{P}\big[\max\{\|\mathbf{X}^n\|, \|\mathbf{X}\|, \|\mathbf{X}^n\|_{p\text{-var}}, \|\mathbf{X}\|_{p\text{-var}}\} > R\big] \leq \varepsilon.$$

The conclusion now follows from Proposition 3.12. For the case of convergence in law, the proof follows in a similar way upon applying the Skorokhod representation theorem [27, Thm. 3.30] to the space $D([0,T], G^N(\mathbb{R}^d))$ and the sequence $(\mathbf{X}^n)_{n\geq 1}$. $\qquad\square$

As an application of Corollary 4.12, along with the fact that piecewise constant approximations satisfy UCV [32, Ex. 3.7], we obtain the following Wong–Zakai-type result (which shall be substantially generalized in Section 4.4 using different methods).

REMARK 4.14. The following result resembles the Wong–Zakai theorem of [31], where it was shown that ODEs driven by approximations of the form $X_t^h = h^{-1} \int_{t-h}^t X_s ds$ converge to an SDE of the Marcus type. Here we are able to complement [31] by showing this for the case of genuine piecewise linear (and a variety of other) approximations. Moreover the deterministic nature of our rough path approach is able to handle anticipating initial data (see Remark 4.19).

PROPOSITION 4.15 (Wong–Zakai with no area). *Let X be a semimartingale with Marcus lift $\mathbf{X}$, and let $\mathcal{D}_n \subset [0, T]$ be a sequence of deterministic partitions such that $|\mathcal{D}_n| \to 0$. Let $X^{[D_n]}$ be the piecewise constant approximations of X along the partition $\mathcal{D}_n$, and let $\mathbf{X}^{[\mathcal{D}_n]}$ be their (Marcus) lifts.*

Let $\phi : \mathbb{R}^d \to C^{q\text{-var}}([0, 1], \mathbb{R}^d)$ be an endpoint continuous, q-approximating path function for some $1 \le q < 2$ such that ϕ does not create area, i.e., (17) holds. By an abuse of notation, let $\phi : \exp(\mathbb{R}^d \oplus \{0\}) \to C^{q\text{-var}}([0, 1], G^2(\mathbb{R}^d))$ denote also the canonical lift of ϕ, treated as a path function defined on $\exp(\mathbb{R}^d \oplus \{0\}) \subset G^2(\mathbb{R}^d)$.

(1) *Consider the admissible pairs $(\mathbf{X}^{[\mathcal{D}_n]}, \phi)$ and $(\mathbf{X}, \phi)$ in $\mathscr{D}([0, T], G^2(\mathbb{R}^d))$. Then for every $p > 2$, $\alpha_{p\text{-var}}(\mathbf{X}^{[\mathcal{D}_n]}, \mathbf{X}) \to 0$ in probability as $n \to \infty$.*
(2) *Let $X^{\mathcal{D}_n, \phi}$ be the piecewise-ϕ interpolation of X along the partition $\mathcal{D}_n$. Let $U_{T \leftarrow 0}^{\mathbf{x}} \in \text{Diff}^m(\mathbb{R}^e)$ denote the flow associated to the RDE (6). Then $U_{T \leftarrow 0}^{X^{\mathcal{D}_n, \phi}}$ converges in probability to $U_{T \leftarrow 0}^{\mathbf{X}^\phi}$ as $n \to \infty$ (as $\text{Diff}^m(\mathbb{R}^e)$-valued random variables).*

PROOF. (1) follows immediately from Corollary 4.12 and the fact that $(X^{[\mathcal{D}_n]})_{n \ge 1}$ satisfies UCV [32, Ex. 3.7].

For (2), note that, since ϕ does not create area, $(\mathbf{X}^{[\mathcal{D}_n]})^\phi$ coincides (up to reparametrization) with the canonical lift of $X^{\mathcal{D}_n, \phi}$. The conclusion now follows from (1) and Theorem 3.13. $\square$

We now record a relation between canonical RDEs and Marcus SDEs.

PROPOSITION 4.16. *Let $X : [0, T] \to \mathbb{R}^d$ be a semimartingale and $\mathbf{X}$ its Marcus lift. Then for vector fields $V = (V_1, \ldots, V_d)$ in $\text{Lip}^\gamma(\mathbb{R}^e)$ for some $\gamma > 2$, it holds that the canonical RDE*

$$(18) \qquad dY_t = V(Y_t) \diamond d\mathbf{X}_t, \quad Y_0 \in \mathbb{R}^e$$

(i.e., the path function ϕ is the taken to be log-linear) coincides a.s. with the Marcus SDE

$$dY_t = V(Y_t) \diamond dX_t, \ \ Y_0 \in \mathbb{R}^e.$$

PROOF. Recall that, by definition, the Marcus SDE satisfies [31]

$$Y_t = Y_0 + \int_0^t V(Y_{s-})dX_s + \frac{1}{2}\int_0^t V'V(Y_s)d[X]_s^c + \sum_{0<s\leq t}\left\{e^{V\Delta X_s}(Y_{s-}) - Y_{s-} - V(Y_{s-})\Delta X_s\right\},$$

where $e^W(y)$ denotes the flow at time 1 along the vector field W from y, i.e., the solution at $t = 1$ to $z_0 = y$, $\dot{z}_t = W(z_t)$.

Recall likewise that, by definition, the rough canonical equation (in the sense of [19, Def. 37]) satisfies

$$Y_t = Y_0 + \int_0^t V(Y_{s-})d\mathbf{X}_s + \sum_{0<s\leq t}\left\{e^{V\Delta X_s}(Y_{s-}) - Y_{s-} - V(Y_{s-})\Delta X_s - V'V(Y_{s-})\frac{1}{2}(\Delta X_s)^{\otimes 2}\right\},$$

which agrees with the solution to the canonical RDE (18) from Definition 3.1 (see [19, Thm. 38]; we point out that [31] and [19] are not restricted to the case of finite activity). It remains to verify that a.s.

$$\int_0^t V(Y_{s-})d\mathbf{X}_s = \int_0^t V(Y_{s-})dX_s + \frac{1}{2}V'V(Y_s)d[X]_s^c + \sum_{0<s\leq t}V'V(Y_{s-})\frac{1}{2}(\Delta X_s)^{\otimes 2}.$$

To this end, observe that

$$\int_0^t V(Y_{s-})d\mathbf{X}_s = \lim_{|\mathcal{D}|\to 0}\sum_{t_i\in\mathcal{D}}V(Y_{t_i-})X_{t_i,t_{i+1}} + V'V(Y_{t_i-})\mathbf{X}_{t_i,t_{i+1}}^{(2)},$$

where $\mathbf{X}_{s,t}^{i,j} = \int_s^t X_{s,u-}^i dX_u^j + \frac{1}{2}[X^i, X^j]_{s,t}^c + \frac{1}{2}\sum_{r\in(s,t]}\Delta X_r^i \Delta X_r^j$. It follows from Lemma 4.35 that for a (deterministic) sequence of partitions with $|\mathcal{D}_n| \to 0$, we have a.s.

$$\int_0^t V(Y_{s-})d\mathbf{X}_s = \lim_{n\to\infty}\sum_{t_i\in\mathcal{D}_n}V(Y_{t_i-})dX_{t_i,t_{i+1}} + \frac{1}{2}V'V(Y_{t_i-})[X]_{t_i,t_{i+1}} + \sum_{r\in(t_i,t_{i+1}]}V'V(Y_{t_i-})\frac{1}{2}(\Delta X_r)^{\otimes 2}$$

from which the conclusion readily follows. $\square$

REMARK 4.17 (Marcus SDEs with piecewise constant driver). Observe that a simple special case of Proposition 4.16 (which does not require any

probabilistic considerations) is a piecewise constant path $X^{[\mathcal{D}]}$, which is constant between the points of a partition $\mathcal{D} \subset [0, T]$. In this case, the solution to

$$dY = V(Y) \diamond dX^{[\mathcal{D}]}, \ \ Y_0 \in \mathbb{R}^e,$$

agrees, for all $t \in \mathcal{D}$, with the ODE solution

$$d\tilde{Y} = V(\tilde{Y}) dX^{\mathcal{D}}, \ \ Y_0 \in \mathbb{R}^e,$$

where $X^{\mathcal{D}}$ is the piecewise linear path obtained from $X^{[\mathcal{D}]}$ by connecting with a straight line consecutive points $X_{t_n}^{[\mathcal{D}]} \rightsquigarrow X_{t_{n+1}}^{[\mathcal{D}]}$ for all $t_n \in \mathcal{D}$.

We are now ready to state the precise criterion for convergence in law (resp. in probability) of Marcus SDEs which was advertised in the introduction and which is analogous to the same criterion for Itô SDEs [32, Thm. 5.4].

THEOREM 4.18.   *Let $V = (V_1, \ldots, V_d)$ be a collection of Lip^γ vector fields on $\mathbb{R}^e$ for some $\gamma > 2$. Let $Y_0, (Y_0^n)_{n \geq 1}$ be a collection of (random) initial conditions in $\mathbb{R}^e$ and $X, (X^n)_{n \geq 1}$ be a collection of semimartingales such that $(X^n)_{n \geq 1}$ satisfies UCV and $(Y_0^n, X^n) \to (Y_0, X)$ in law (resp. in probability) as $n \to \infty$ (as $\mathbb{R}^e \times D([0, T], \mathbb{R}^d)$-valued random variables). Then the solutions to the Marcus SDEs*

$$dY_t^n = V(Y_t^n) \diamond dX_t^n, \ \ Y_0^n \in \mathbb{R}^e,$$

converge in law (resp. in probability) as $n \to \infty$ (in the Skorokhod topology) to the solution of the Marcus SDE

(19) $$dY_t = V(Y_t) \diamond dX_t, \ \ Y_0 \in \mathbb{R}^e.$$

PROOF. This is an immediate consequence of Corollary 4.12, Proposition 4.16, and the deterministic continuity of the solution map (part (ii) of Proposition 3.18).   □

REMARK 4.19.   We have not been explicit about filtrations, but of course, every semimartingale X^n above is adapted to some filtration $\{\mathcal{F}_t^n\}_{t \geq 0}$. In the same vain, as is standard in the context of SDEs, the initial datum Y_0^n is assumed to be $\mathcal{F}_0^n$-measurable, so that (19) makes sense as a bona fide integral equation (as recalled in the proof of Proposition 4.16).

Situations where Y_n^0 is independent of the driving noise X^n are then immediately handled. If, on the other hand, Y_n^0 depends in some anticipating fashion on the driving noise, then classical SDE theory (Marcus or Itô) breaks down and ideas from anticipating stochastic calculus are necessary

(such as composing the stochastic flow with anticipating initial data; in the Marcus context this would be possible thanks to [31, Thm. 3.4]). Our (essentially deterministic) rough path approach bypasses such problems entirely. We shall not pursue further application of rough paths to "anticipating Marcus SDEs" here, but note that this could be done analogously to [11].

4.3. *Examples.* We now give a list of examples to which Corollary 4.12, Proposition 4.15, and Theorem 4.18 apply. The main criterion of application is of course the UCV condition. For further examples of sequences of semimartingales satisfying UCV, see [32, Sec. 3]. We note that in the framework of Theorem 4.18, the UCV condition cannot in general be ommited (but see Theorem 4.29 below) as seen, e.g., in homogenization theory [34, 15, 28] and non-standard approximations to Brownian motion [41, 16].

EXAMPLE 4.20 (Piecewise constant approximations). Let X be a càdlàg semimartingale and $X^{[\mathcal{D}_k]}$ be its piecewise constant approximation (see Figure 4) along a sequence of deterministic partitions $\mathcal{D}_k \subset [0, T]$ such that $|\mathcal{D}_k| \to 0$. Then by Theorem 4.18, the solutions to

$$dY_t^{[\mathcal{D}_k]} = V(Y_t^{[\mathcal{D}_k]}) \diamond dX_t^{[\mathcal{D}_k]}, \quad Y_0^{[\mathcal{D}_k]} = y_0 \in \mathbb{R}^e$$

converge in probability (for the Skorokhod topology) to the solution of

$$dY_t = V(Y_t) \diamond dX_t, \quad Y_0 = y_0 \in \mathbb{R}^e.$$

Moreover, if X is continuous, then $Y^{[\mathcal{D}_k]}$ converges in probability for the uniform topology on $[0, T]$ to Y (which is now also the solution to the Stratonovich SDE).

EXAMPLE 4.21 (Piecewise linear approximations). Let X be a càdlàg semimartingale and now let $X^{\mathcal{D}_k}$ be its piecewise linear (i.e., Wong–Zakai) approximation (see Figure 5) along a sequence of deterministic partitions $\mathcal{D}_k \subset [0, T]$ such that $|\mathcal{D}_k| \to 0$. Consider the solutions to random ODEs

$$dY_t^{\mathcal{D}_k} = V(Y_t^{\mathcal{D}_k}) dX_t^{\mathcal{D}_k}, \quad Y_0^{\mathcal{D}_k} = y_0 \in \mathbb{R}^e,$$

and the Marcus SDE

$$dY_t = V(Y_t) \diamond dX_t, \quad Y_0 = y_0 \in \mathbb{R}^e.$$

Then, by Proposition 4.15, it holds that $Y_T^{\mathcal{D}_k} \to Y_T$ in probability.

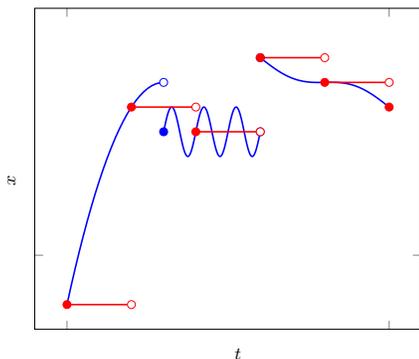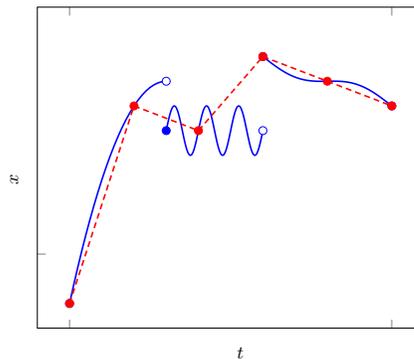Fig 4: Càdlàg path (blue) with piece-wise constant approximtation (red)

Fig 5: Càdlàg path (blue) with piecewise linear approximation (red, dashed)

Moreover, if X is continuous, then in light of the last part of Example 4.20 and Remark 4.17, $Y^{\mathcal{D}_k}$ converges in probability for the uniform topology on $[0, T]$ to Y (which, we emphasize again, is now the solution to the Stratonovich SDE), which agrees with the classical Wong–Zakai theorem for continuous semimartingales.

EXAMPLE 4.22 (Donsker approximations to Brownian motion). Consider an $\mathbb{R}^d$-valued random walk X^n with iid increments and finite second moments, rescaled so that $X^n \to B$ in law. Here we treat X^n as either piecewise constant or interpolated using any sufficiently nice path function ϕ which does not create area, i.e., satisfies (17) (e.g., piecewise linear). Then $(X^n)_{n \geq 1}$ satisfies UCV, so by Corollary 4.12 we again have convergence of the Marcus SDEs (or random ODEs in case of continuous interpolations)

$$dY_t^n = V(Y_t^n) \diamond dX_t^n \ \ Y_0 = y_0 \in \mathbb{R}^e$$

in law for the uniform topology on $[0, T]$ to the Stratonovich limit

$$dY_t = V(Y_t) \circ dB_t \ \ Y_0 = y_0 \in \mathbb{R}^e.$$

This is a special case of [9, Ex. 5.12] (see also Example 4.23) which improves the main result of Breuillard et al. [5] in the sense that no additional moment assumptions are required (highlighting a benefit of p-variation vs. Hölder topology).

EXAMPLE 4.23 (Null array approximations to Lévy processes). Generalizing Example 4.22, consider a null array of $\mathbb{R}^d$-valued random variables

$X_{n1}, \ldots, X_{nn}$, i.e., $\lim_{n \to \infty} \sup_k \mathbb{E}\left[|X_{nk}| \wedge 1\right] = 0$, and, for every $n \geq 1$, $X_{n1}, \ldots, X_{nn}$ are independent. Consider the associated random walk

$$X^n : [0,1] \to \mathbb{R}^d, \;\; X_t^n = \sum_{k=1}^{\lfloor tn \rfloor} X_{nk}.$$

Suppose $X^n \to X$ in law for a Lévy process X (see [27, Thm. 13.28] for necessary and sufficient conditions for this to occur), which in particular implies that for some $h > 0$

1. $\sum_{k=1}^n \mathbb{E}[X_{nk} \mathbf{1}_{\{|X_{nk}| < h\}}] \to b^h$,
2. $\sum_{k=1}^n \mathbb{E}[X_{nk}^i X_{nk}^j \mathbf{1}_{\{|X_{nj}| < h\}}] \to a_{i,j}^h$, and
3. $\sum_{k=1}^n \mathbb{E}[f(X_{nk})] \to \nu(f)$ for every $f \in C_b(\mathbb{R}^d)$ which is identically zero on a neighbourhood of zero,

where b^h, a^h, and ν are determined by the Lévy triplet of X (in particular ν is the Lévy measure of X). As a consequence, it is immediate to verify that $(X^n)_{n \geq 1}$ satisfies UCV. By Theorem 4.18, the solutions to

$$dY_t^n = V(Y_t^n) \diamond dX_t^n, \;\; Y_0^n = y_0 \in \mathbb{R}^e,$$

converge in law (for the Skorokhod topology) to the solution of the Marcus SDE

$$dY_t = V(Y_t) \diamond dX_t, \;\; Y_0 = y_0 \in \mathbb{R}^e.$$

If, once more, X^n are interpolated using any sufficiently nice path function ϕ (which in particular does not create area (17)), an application of Corollary 4.12 implies that Y_T^n (now solutions to random ODEs) converge in law to Y_T (now the solution to the random RDE driven by $\mathbf{X}^\phi$). Note that if ϕ is allowed to create area and $X_{n1}, \ldots, X_{nn}$ are further assumed iid for every $n \geq 1$, then this is precisely the case addressed in [9, Ex. 5.12] (though in this case one must consider a non-Marcus lift of X similar to the upcoming Theorem 4.29).

EXAMPLE 4.24 (Martingale CLT). Let $(X^n)_{n \geq 1}$ be a sequence of $\mathbb{R}^d$-valued càdlàg local martingales. Suppose that, as $n \to \infty$,

$$\mathbb{E}\big[\sup_{t \in (0,T]} |\Delta X_t^n|\big] \to 0 \;, \;\; [X^n, X^n]_t \to C(t) \;\; \forall t \in (0, T],$$

where $t \mapsto C(t) \in \mathbb{R}^{d \times d}$ is continuous and deterministic. Then $X^n \to X$, where X is a continuous Gaussian process with independent increments

and $\mathbb{E}\left[X(t)X(t)^T\right] = C(t)$ [14, Thm. 1.4, p. 339], and moreover the UCV condition is satisfied [33, p. 26]. Therefore solutions to

$$dY_t^n = V(Y_t^n) \diamond dX_t^n, \ \ Y_0^n = y_0 \in \mathbb{R}^e,$$

converge in law for the uniform topology on $[0, T]$ to the solution of the Stratonovich SDE

$$dY_t = V(Y_t) \circ dX_t, \ \ Y_0 = y_0 \in \mathbb{R}^e.$$

4.4. *Wong–Zakai revisited.* In this subsection we significantly expand Proposition 4.15 by showing convergence in probability of very general (area-creating) interpolations of càdlàg semimartingales. If the interpolation creates area, we in general no longer expect to converge to the Marcus lift of X (which is the reason one cannot apply Proposition 4.15), and therefore we first modify the lift $\mathbf{X}$ appropriately.

Throughout the section, we fix a (left-invariant) q-approximating path function $\phi : \mathbb{R}^d \to C^{q\text{-var}}([0, 1], \mathbb{R}^d)$ for some $1 \leq q < 2$ (which we take to be defined on the entire space $\mathbb{R}^d$ only for simplicity). Consider the two maps

$$\psi : \mathbb{R}^d \to G^2(\mathbb{R}^d)$$
$$a : \mathbb{R}^d \to \mathfrak{g}^{(2)}(\mathbb{R}^d) \cong \mathfrak{so}(\mathbb{R}^d)$$

defined uniquely by

$$S_2(\phi(x))_{0,1} = \psi(x) = \exp(x + a(x))$$

(so that $a(x)$ is the area generated by the path $\phi(x) : [0, 1] \to \mathbb{R}^d$).

Let us also fix a càdlàg semimartingale $X : [0, T] \to \mathbb{R}^d$ and a sequence of deterministic partitions $\mathcal{D}_k \subset [0, T]$ such that $\lim_{k \to \infty} |\mathcal{D}_k| = 0$. Consider the following assumption.

ASSUMPTION 4.25. *There exists a càdlàg bounded variation process* $B : [0, T] \to \mathfrak{g}^{(2)}(\mathbb{R}^d)$ *such that*

$$\sup_{t \in [0,T]} \left| B_t - \sum_{\mathcal{D}_k \ni t_j \leq t} a(X_{t_j} - X_{t_{j-1}}) \right| \to 0 \ \text{in probability as } k \to \infty.$$

Before stating the Wong–Zakai theorem, we give several examples of ϕ for which Assumption 4.25 is satisfied.

EXAMPLE 4.26 (No area). If $\phi(x)$ does not create area for all $x \in \mathbb{R}^d$, so that $a \equiv 0$ (e.g., when ϕ is the linear interpolation on $\mathbb{R}^d$), then evidently Assumption 4.25 is satisfied with $B_t \equiv 0$.

EXAMPLE 4.27 (Hoff-type process). Suppose that $\mathbb{R}^d = \mathbb{R}^2$, so that $\mathfrak{g}^{(2)}(\mathbb{R}^2) \cong \mathbb{R}$. Let ϕ travel to (x, y) first linearly along the x-coordinate and then linearly along the y-coordinate:

$$\phi(x, y)_t = \mathbf{1}_{\{t \in [0, 1/2]\}} 2tx + \mathbf{1}_{\{t \in (1/2, 1]\}}(x + (2t - 1)y).$$

Then $a(x, y) = \frac{1}{2}xy$, so that Assumption 4.25 is satisfied with $B_t = \frac{1}{2}[X^1, X^2]_t$.

EXAMPLE 4.28 (Regular a). Suppose more generally that a is twice differentiable at 0, so that by Lemma 4.31

$$a(X_{s,t}) = \frac{1}{2}D^2 a(0)(X_{s,t}^{\otimes 2}) + o(|X_{s,t}|^2).$$

We see in this case that Assumption 4.25 is satisfied with

$$B_t = \frac{1}{2} \sum_{i,j=1}^{d} D^2 a(0)^{i,j}[X^i, X^j]_t^c + \sum_{s \leq t} a(\Delta X_s).$$

For a partition $\mathcal{D} \subset [0, T]$, let $X^{\mathcal{D}, \phi}$ be the piecewise-ϕ interpolation of X along $\mathcal{D}$, and $\mathbf{X}^{\mathcal{D}, \phi}$ its canonical lift. The following is the main result of this subsection.

THEOREM 4.29 (Wong–Zakai). *Suppose that Assumption 4.25 is satisfied. Let $\bar{\mathbf{X}} : [0, T] \to G^2(\mathbb{R}^d)$ be the modified level-2 lift of X defined by*

$$\bar{\mathbf{X}} := \exp(X + \bar{A}), \quad \bar{A}_t := A_t + B_t,$$

where $\mathbf{X} = \exp(X + A)$ is the Marcus lift of X. Suppose further that ϕ is endpoint continuous.

(1) *Consider the admissible pair $(\bar{\mathbf{X}}, \phi) \in \mathscr{D}([0, T], G^2(\mathbb{R}^d))$. Then for every $p > 2$, it holds that*

$$\alpha_{p\text{-var}}(\mathbf{X}^{\mathcal{D}_k, \phi}, \bar{\mathbf{X}}) \to 0 \text{ in probability as } k \to \infty.$$

(2) *Let $U_{T \leftarrow 0}^{\mathbf{x}} \in \mathrm{Diff}^m(\mathbb{R}^e)$ denote the flow associated to the RDE (6). Then $U_{T \leftarrow 0}^{X^{\mathcal{D}_k, \phi}}$ converges in probability to $U_{T \leftarrow 0}^{\bar{\mathbf{X}}^\phi}$ (as a $\mathrm{Diff}^m(\mathbb{R}^e)$-valued random variable).*

REMARK 4.30. Note that the jumps of B_t must necessarily be of the form $a(\Delta X_t)$. Hence $\Delta \bar{\mathbf{X}}_t \in \psi(\mathbb{R}^d)$, so that indeed $\bar{\mathbf{X}} \in \mathscr{D}_\phi([0, T], G^2(\mathbb{R}^d))$ and $\bar{\mathbf{X}}^\phi$ is well-defined (this is an abuse of notation since ϕ is a path function on $\mathbb{R}^d$, but because $\phi(x)$ of finite q-variation, it can be canonically lifted to a path function $\phi : \psi(\mathbb{R}^d) \to G^2(\mathbb{R}^d)$).

For the proof of the theorem, we require several lemmas.

LEMMA 4.31.   *Let $\eta_{q\text{-var}}$ be a q-variation modulus of ϕ (see Definition 2.14). Then*

$$|a(x)| \leq \eta_{q\text{-var}}(r)|x|^2, \ \ \forall r > 0, \ \ \forall x \in \mathbb{R}^d \text{ s.t. } |x| \leq r.$$

PROOF. This is immediate from the property

$$|x| + |a(x)|^{1/2} \asymp \|\psi(x)\| \leq \|\phi(x)\|_{q\text{-var};[0,1]} \leq \eta_{q\text{-var}}(r)|x|.$$

$\square$

The following lemma essentially involves no probability.

LEMMA 4.32.   *Suppose ϕ is endpoint continuous. Then for a.e. sample path $\bar{\mathbf{X}} \in D([0,T], G^2(\mathbb{R}^d))$, it holds that for all $\varepsilon > 0$ there exists $r > 0$, such that for all partitions $\mathcal{D} \subset [0,T]$ with $|\mathcal{D}| < r$, there exists $\delta_0 > 0$, such that for all $\delta < \delta_0$, there exists $\lambda \in \Lambda$ such that*

$$|\lambda| < 2|\mathcal{D}|,$$
$$\bar{\mathbf{X}}_{t_n} = \bar{\mathbf{X}}^{\phi,\delta}_{\lambda(t_n)}, \ \ \forall t_n \in \mathcal{D}, \ \ \text{and}$$
$$\max_{t_n \in \mathcal{D}} \sup_{t \in [t_n, t_{n+1}]} d(\mathbf{X}^{\mathcal{D},\phi}_{t_n,t}, \bar{\mathbf{X}}^{\phi,\delta}_{\lambda(t_n),\lambda(t)}) < \varepsilon.$$

PROOF. Since $\bar{\mathbf{X}}$ is càdlàg, for every $\varepsilon > 0$ we can find $r > 0$ sufficiently small and a partition $\mathcal{P} = (s_0, \ldots, s_m)$ so that $|s_{i+1} - s_i| > 2r$ and

$$\sup_{u,v \in [s_i, s_{i+1})} \|\bar{\mathbf{X}}_{u,v}\| < \varepsilon.$$

Then whenever $|\mathcal{D}| < r$, for every $t_n \in \mathcal{D}$, there exists at most one $s_i \in \mathcal{P}$ such that $s_i \in [t_n, t_{n+1}]$. Now using the fact that ϕ is q-approximating and endpoint continuous, the claim readily follows. $\square$

LEMMA 4.33.   *The family of real random variables $(\|\mathbf{X}^{\mathcal{D},\phi}\|_{p\text{-var};[0,T]})_{\mathcal{D} \subset [0,T]}$ is tight.*

REMARK 4.34.   As the proof of Lemma 4.33 will reveal, the biggest difficulty is overcome by the enhanced BDG inequality for càdlàg local martingales (Theorem 4.7). We wish to emphasize that the lemma is even helpful in the context of a rough paths proof of the Wong–Zakai theorem for *continuous* semimartingales with piecewise linear interpolations (so that $a \equiv 0$), since

an analogous tightness result is still needed in this case and is non-trivial (cf. [20, Thm. 14.15]).

We also mention that part (1) of Proposition 4.15 in particular shows that $(\|\mathbf{X}^{[\mathcal{D}_k]}\|_{p\text{-var}})_{k\geq 1}$ is tight, which can significantly simply the proof of Lemma 4.33 (at least if one restricts attention to the family $(\|\mathbf{X}^{\mathcal{D}_k,\phi}\|_{p\text{-var}})_{k\geq 1}$). However we give a direct proof of the general result here.

PROOF OF LEMMA 4.33. We can decompose $X = L + K$ (non-uniquely) where K is of bounded variation and L is a local martingale with jumps bounded by some $M > 0$ (e.g., [25, Prop. 4.17, p. 42]). Using the localizing sequence $\tau_m = \inf\{t \in [0,T] \mid |L_t| + |K_t| > m\}$, we may further suppose that L is bounded.

For a partition $\mathcal{D} \subset [0,T]$, consider the piecewise constant path $\tilde{\mathbf{X}}^{[\mathcal{D}]} : [0,T] \to G^2(\mathbb{R}^d)$ which is constant on $[t_n, t_{n+1})$ and $\tilde{\mathbf{X}}^{[\mathcal{D}]}_{t_n} = \mathbf{X}^{\mathcal{D},\phi}_{t_n}$ for every $t_n \in \mathcal{D}$. Consider also the piecewise constant semimartingale $X^{[\mathcal{D}]}$ which is constant on $[t_n, t_{n+1})$ and $X^{[\mathcal{D}]}_{t_n} = X_{t_n}$ for every $t_n \in \mathcal{D}$. Let $\mathbf{X}^{[\mathcal{D}]}$ be the (Marcus) lift of $X^{[\mathcal{D}]}$. By definition of $a : \mathbb{R}^d \to \mathfrak{g}^{(2)}(\mathbb{R}^d)$, we have for all $t_n \in \mathcal{D}$

$$\tilde{\mathbf{X}}^{[\mathcal{D}]}_{t_n} = \mathbf{X}^{[\mathcal{D}]}_{t_n} \otimes \exp\Big(\sum_{\mathcal{D}\ni t_k \leq t_n} a(X_{t_k} - X_{t_{k-1}})\Big).$$

It follows that

$$\|\tilde{\mathbf{X}}^{[\mathcal{D}]}\|^p_{p\text{-var};[0,T]} \leq C\big(\|\mathbf{X}^{[\mathcal{D}]}\|^p_{p\text{-var};[0,T]} + \|Y^{[\mathcal{D}]}\|^{p/2}_{p/2\text{-var};[0,T]}\big),$$

where $Y^{[\mathcal{D}]}_t = \sum_{\mathcal{D}\ni t_n \leq t} a(X_{t_n} - X_{t_{n-1}}) \in \mathfrak{g}^{(2)}(\mathbb{R}^d)$.

Observe that $\mathbf{X}^{\mathcal{D},\phi}$ is a reparametrization of $(\tilde{\mathbf{X}}^{[\mathcal{D}]})^\phi$ (where we use the same abuse of notation as in Remark 4.30), so in particular

$$\|\mathbf{X}^{\mathcal{D},\phi}\|_{p\text{-var};[0,T]} = \|(\tilde{\mathbf{X}}^{[\mathcal{D}]})^\phi\|_{p\text{-var};[0,T]}.$$

Moreover, defining the piecewise constant local martingale $L^{[\mathcal{D}]}$, its lift $\mathbf{L}^{[\mathcal{D}]}$, and the piecewise constant path of bounded variation $K^{[\mathcal{D}]}$ in the same way as $X^{[\mathcal{D}]}$, we note that $\mathbf{X}^{[\mathcal{D}]} = T_{K^{[\mathcal{D}]}}(\mathbf{L}^{[\mathcal{D}]})$. Following Lemma 2.16, it suffices to show that the families

$$(\|Y^{[\mathcal{D}]}\|_{p/2\text{-var}})_{\mathcal{D}\subset[0,T]}, \quad (\|\mathbf{L}^{[\mathcal{D}]}\|_{p\text{-var}})_{\mathcal{D}\subset[0,T]}, \quad (\|K^{[\mathcal{D}]}\|_{1\text{-var}})_{\mathcal{D}\subset[0,T]}$$

are tight. This in turn follows respectively from Lemma 4.31, the enhanced BDG inequality Theorem 4.7, and the fact that $\|K^{[\mathcal{D}]}\|_{1\text{-var}} \leq \|K\|_{1\text{-var}}$. $\square$

PROOF OF THEOREM 4.29. Note that (2) follows directly from (1) and Theorem 3.13. To show (1), note that $\|\bar{\mathbf{X}}\|_{p\text{-var}} < \infty$ a.s. (by Corollary 4.3), so following Lemma 4.33 and interpolation (Lemma 3.11), it suffices to show that to show that

$$\alpha_\infty(\mathbf{X}^{\mathcal{D}_k,\phi}, \bar{\mathbf{X}}) \to 0 \text{ in probability as } k \to \infty.$$

Let $\bar{A}, \bar{A}^{\phi,\delta}$ and $A^{\mathcal{D}_k,\phi}$ denote the stochastic area of $\bar{\mathbf{X}}, \bar{\mathbf{X}}^{\phi,\delta}$ and $\mathbf{X}^{\mathcal{D}_k,\phi}$ respectively. We have for $t \in [t_n, t_{n+1}] \subset \mathcal{D}_k$

$$\mathbf{X}^{\mathcal{D}_k,\phi}_t = \psi(X_{0,t_1}) \dots \psi(X_{t_{n-1},t_n}) S_2(\phi(X_{t_n,t_{n+1}}))_{\frac{t-t_n}{t_{n+1}-t_n}},$$

and so by the Campbell–Baker–Hausdorff formula

$$A^{\mathcal{D}_k,\phi}_t = A^{\mathcal{D}_k,\phi}_{t_n,t} + \sum_{j=0}^{n-1} a(X_{t_j,t_{j+1}}) + \frac{1}{2}\big[X_{0,t_n}, X^{\mathcal{D}_k,\phi}_{t_n,t}\big] + \frac{1}{2}\sum_{j=0}^{n-1}[X_{0,t_j}, X_{t_j,t_{j+1}}].$$

Likewise

$$\bar{A}_t = \bar{A}_{t_n,t} + \sum_{j=0}^{n-1} \bar{A}_{t_j,t_{j+1}} + \frac{1}{2}[X_{0,t_n}, X_{t_n,t}] + \frac{1}{2}\sum_{j=0}^{n-1}[X_{0,t_j}, X_{t_j,t_{j+1}}].$$

Recalling further that $\bar{A}_{t_j,t_{j+1}} = A_{t_j,t_{j+1}} + B_{t_j,t_{j+1}}$ and $X^{\mathcal{D}_k,\phi}_{t_n} = X_{t_n}$, it follows that for all $t_n \in \mathcal{D}_k$

$$d(\mathbf{X}^{\mathcal{D}_k,\phi}_{t_n}, \bar{\mathbf{X}}_{t_n}) \asymp \Big| \sum_{j=0}^{n-1} A_{t_j,t_{j+1}} + B_{t_j,t_{j+1}} - a(X_{t_{j+1}} - X_{t_j}) \Big|^{1/2}.$$

Combining Lemma 4.32 with Assumption 4.25, we see that the proof is complete once we show

$$\max_{t_n \in \mathcal{D}_k} \Big| \sum_{j=0}^{n-1} A_{t_j,t_{j+1}} \Big| \to 0 \text{ in probability as } k \to \infty,$$

which in turn follows from Lemma 4.35. $\square$

4.5. Appendix: Vanishing areas.

LEMMA 4.35.   *Let $X : [0,T] \to \mathbb{R}^d$ be a càdlàg semimartingale, $Y : [0,T] \to \mathbf{L}((\mathbb{R}^d)^{\otimes 2}, \mathbb{R})$ a locally bounded previsible process, and $(\mathcal{D}_k)_{k\geq 1}$ a sequence of deterministic partitions of $[0,T]$ such that $\lim_{k\to\infty} |\mathcal{D}_k| = 0$. Define $\mathbb{X}_{s,t} := \int_s^t (X_{r-} - X_s) \otimes dX_r$ as Itô integrals. Then*

$$\max_{t_n \in \mathcal{D}_k} \Big| \sum_{j=0}^{n-1} Y_{t_j} \mathbb{X}_{t_j,t_{j+1}} \Big| \to 0 \text{ in probability as } k \to \infty.$$

REMARK 4.36.   In the case $Y \equiv 1$, observe that Lemma 4.35 is an immediate consequence of the convergence in part (1) of Proposition 4.15 (where ϕ is taken as the piecewise linear interpolation).

PROOF. As in the proof of Lemma 4.33, we can decompose $X = L + K$, where K is of bounded variation and L is a local martingale with bounded jumps (e.g., [25, Prop. 4.17, p. 42]). Using the localizing sequence

$$\tau_m = \inf\{t \in [0, T] \mid |Y_t| + |L_t| + |K_t| > m\},$$

we may further suppose that Y and L are uniformly bounded and that X is bounded on $[0, \tau_m)$ by $m > 0$ and is constant on $[\tau_m, T]$. We now write

$$\mathbb{X}_{s,t} = \int_s^t (X_{r-} - X_s)(dL_r + dK_r).$$

For a fixed càdlàg sample path X, for every $\varepsilon > 0$ we can find $r > 0$ sufficiently small and a partition $\mathcal{P} = (s_0, \ldots, s_m)$ so that $|s_{i+1} - s_i| > r$ and

$$\sup_{u,v \in [s_i, s_{i+1})} |X_{u,v}| < \varepsilon.$$

Then whenever $|\mathcal{D}| < r/2$, for every $t_n \in D$, there exists at most one $s_i \in \mathcal{P}$ such that $s_i \in [t_n, t_{n+1}]$.

Since K is a process of finite variation, if $s_i \in [t_n, t_{n+1}]$ then

$$\left| \int_{t_n}^{t_{n+1}} (X_{s-} - X_{t_n}) dK_s \right| \le \varepsilon |K|_{1\text{-var};[t_n,s_i]} + 2m|K|_{1\text{-var};[s_i,t_{n+1}]},$$

and if no $s_i \in \mathcal{P}$ is in $[t_n, t_{n+1}]$, then the upper bound is $\varepsilon |K|_{1\text{-var};[t_n,t_{n+1}]}$. Denoting by $[t_n, t_{n+1}]$ the interval in $\mathcal{D}_k$ containing $s_i \in \mathcal{P}$, we then have

$$\sum_{t_n \in \mathcal{D}_k} \left| Y_{t_n} \int_{t_n}^{t_{n+1}} (X_{s-} - X_{t_n}) dK_s \right| \le C|Y|_\infty \left(\varepsilon |K|_{1\text{-var}} + 2m \sum_{s_i \in \mathcal{P}} |K|_{1\text{-var};[s_i,t_{n+1}]} \right),$$

from which it follows that the LHS converges to zero a.s. as $k \to \infty$. It remains to show that

$$(20) \qquad \max_{t_n \in \mathcal{D}_k} \left| \sum_{j=0}^{n-1} Y_{t_j} \int_{t_j}^{t_{j+1}} (X_{s-} - X_{t_n}) dL_s \right| \to 0$$

in probability as $k \to \infty$. By Itô isometry

$$\mathbb{E}\left[\left(Y_{t_n} \int_{t_n}^t (X_{s-} - X_{t_n}) dL_s \right)^2 \right] \le C\mathbb{E}\left[|Y_{t_n}|^2 \int_{t_n}^t |X_{s-} - X_{t_n}|^2 d[L]_s \right].$$

As before, if $s_i \in \mathcal{P}$ is in $[t_n, t_{n+1}]$, then

$$\int_{t_n}^{t} |X_{s-} - X_{t_n}|^2 d|[L]_s| \leq \varepsilon^2 |[L]_{t_n, s_i}| + (2m)^2 |[L]_{s_i, t_{n+1}}|,$$

and if no $s_i \in \mathcal{P}$ is in $[t_n, t_{n+1}]$, then the upper bound is $\varepsilon^2 [L]_{t_n, t_{n+1}}$. Hence

$$\sum_{t_n \in \mathcal{D}_k} |Y_{t_n}|^2 \int_{t_n}^{t_{n+1}} |X_{s-} - X_{t_n}|^2 d|[L]_s| \leq C|Y|_\infty^2 \Big(\varepsilon^2 |[L]_\infty| + (2m)^2 \sum_{s_i \in \mathcal{P}} |[L]_{s_i, t_{n+1}}| \Big),$$

from which it follows that the LHS converges to zero a.s. as $k \to \infty$. As L is bounded (so in particular bounded in L^2), we obtain by dominated convergence

$$\mathbb{E}\Big[\sum_{t_n \in \mathcal{D}_k} \Big(Y_{t_n} \int_{t_n}^{t_{n+1}} (X_{s-} - X_{t_n}) dL_s \Big)^2 \Big] \leq C\mathbb{E}\Big[\sum_{t_n \in \mathcal{D}_k} |Y_{t_n}|^2 \int_{t_n}^{t_{n+1}} |X_{s-} - X_{t_n}|^2 d|[L]_s| \Big] \to 0.$$

Finally, applying the classical BDG inequality to the discrete-time martingale

$$\sum_{j=0}^{n} Y_{t_j} \int_{t_j}^{t_{j+1}} (X_{s-} - X_{t_n}) dL_s,$$

we see that (20) holds in L^2, and thus in probability as desired. $\qquad \square$

5. Beyond semimartingales. We have seen in the previous section that (general) multi-dimensional semimartingales give rise to (càdlàg) geometric p-rough paths. Marcus lifts of general semimartingales provide us with concrete and important examples of driving signals for canonical RDEs, providing a decisive and long-awaited [47] rough path view on classical stochastic differential equations with jumps. We now discuss several examples to which the theory of Section 3 can be applied which fall outside the scope of classical semimartingale theory.

5.1. *Semimartingales perturbed by paths of finite q-variation.* Keeping focus on $\mathbb{R}^d$-valued processes and their (canonical) lifts, we remark that any process with a decomposition $X = Y + B$, where Y is a semimartingale and B is a process with finite q-variation for some $q < 2$, admits a canonical Marcus lift given by $\mathbf{X} = T_B(\mathbf{Y})$. Note that due to the deterministic nature of the (Young) integrals used to construct $T_B(\mathbf{Y})$, we require no adaptedness assumptions on B. We summarize the existence of the lift in the following proposition, which is an immediate consequence of Proposition 3.23.

PROPOSITION 5.1. *Let $1 \leq q < 2$ and consider an $D^{q\text{-var}}([0,T],\mathbb{R}^d)$-valued random variable B, and an $\mathbb{R}^d$-valued semimartingale Y. Write A_Y for the area of Y and define its Marcus lift $\mathbf{Y} = \exp{(Y + A_Y)}$. Then the process $X := Y + B$ admits a canonical lift, given by $\mathbf{X} = T_B\mathbf{Y}$, which is a Marcus-like geometric p-rough path for any $p > 2$.*

We mention that the class of paths with such a decomposition contains some well-studied processes.

EXAMPLE 5.2 (PII).   The important class of processes with independent increments (PII) goes beyond semimartingale theory. In fact, every such process X can be decomposed (non-uniquely) as $X = Y + B$, where Y is a PII and a semimartingale and B is a deterministic càdlàg path. Moreover, X is a semimartingale if and only if B has finite variation on compacts. Provided that the process B has finite q-variation for some $q < 2$, we immediately see that X admits a lift to a (càdlàg) geometric p-rough path for any $p > 2$.

We note that there is a natural interest in differential equations driven by PIIs (under the assumptions of Proposition 5.1 this is meaningful!) since the resulting (pathwise) solutions Z to canonical RDEs

$$dZ = V(Z) \diamond d\mathbf{X}$$

will be (time-inhomogeneous) Markov processes. A further study and characterization of such processes seems desirable. (For instance, they may not be nicely characterized by their generator. Consider the case $V \equiv 1, X = B \in C^{q-var}([0,T],\mathbb{R}) \setminus C^1([0,T],\mathbb{R})$ with $q \in (1,2)$.)

5.2. *Markovian and Gaussian càdlàg rough paths.*   One can use Dirichlet forms to construct Markovian rough paths which are not lifts of semimartingales. In the continuous setting this has been developed in detail in [17]. Including a non-local term in the Dirichlet form will allow to extend this construction to the jump case, but we will not investigate this here. We also note that Gaussian càdlàg rough paths can be constructed; as in the continuous theory, the key condition is finite ρ-variation of the covariance (cf. [18, Sec. 10.2]) but without assuming its continuity.

5.3. *Group-valued processes.*   This point was already made in the context of Lévy rough paths [19, 9], which substantially generalizes the notion of the Marcus lift of an $\mathbb{R}^d$-valued Lévy process (such processes, for example, arise naturally as limits of stochastic flows, see [9, Sec. 5.3.1]). In the same spirit, one can defined "genuine semimartingale rough paths" as $G^N(\mathbb{R}^d)$-valued process with local characteristics modelled after Lévy (rough path)

triplets. (Remark that the Lie group $G^N(\mathbb{R}^d)$ is, in particular, a differentiable manifold so that the theory of manifold-valued semimartingales applies. The issue is to identify those which constitute (geometric) rough paths, which can be done analyzing the local characteristics, as was done in the Lévy case in the afore-mentioned papers. In the same spirit, the afore-mentioned Dirichlet-form construction also extends to the group setting.)

References.

[1] APPLEBAUM, D. (2009). *Lévy processes and stochastic calculus*, second ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge University Press, Cambridge. MR2512800

[2] APPLEBAUM, D. and KUNITA, H. (1993). Lévy flows on manifolds and Lévy processes on Lie groups. *J. Math. Kyoto Univ.* **33** 1103–1123. MR1251218

[3] APPLEBAUM, D. and TANG, F. (2001). Stochastic flows of diffeomorphisms on manifolds driven by infinite-dimensional semimartingales with jumps. *Stochastic Process. Appl.* **92** 219–236. MR1817587

[4] BILLINGSLEY, P. (1999). *Convergence of probability measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons, Inc., New York A Wiley-Interscience Publication. MR1700749

[5] BREUILLARD, E., FRIZ, P. and HUESMANN, M. (2009). From random walks to rough paths. *Proc. Amer. Math. Soc.* **137** 3487–3496. MR2515418

[6] BURKHOLDER, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probability* **1** 19–42. MR0365692

[7] CASS, T. and OGRODNIK, M. (2017). Tail estimates for Markovian rough paths. *Ann. Probab.* **45** 2477–2504. MR3693967

[8] CHECHKIN, A. and PAVLYUKEVICH, I. (2014). Marcus versus Stratonovich for systems with jump noise. *J. Phys. A* **47** 342001, 15. MR3251981

[9] CHEVYREV, I. (2015). Random walks and Lévy processes as rough paths. *ArXiv e-prints*. To appear in Probab. Theory Related Fields.

[10] CHEVYREV, I. and LYONS, T. (2016). Characteristic functions of measures on geometric rough paths. *Ann. Probab.* **44** 4049–4082. MR3572331

[11] COUTIN, L., FRIZ, P. and VICTOIR, N. (2007). Good rough path sequences and applications to anticipating stochastic calculus. *Ann. Probab.* **35** 1172–1193. MR2319719

[12] COUTIN, L. and LEJAY, A. (2005). Semi-martingales and rough paths theory. *Electron. J. Probab.* **10** no. 23, 761–785. MR2164030

[13] DAVIS, G. J. and HU, T. Y. (1995). On the structure of the intersection of two middle third Cantor sets. *Publ. Mat.* **39** 43–60. MR1336355

[14] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov processes. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York Characterization and convergence. MR838085

[15] FRIZ, P., GASSIAT, P. and LYONS, T. (2015). Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.* **367** 7939–7955. MR3391905

[16] FRIZ, P. and OBERHAUSER, H. (2009). Rough path limits of the Wong-Zakai type with a modified drift term. *J. Funct. Anal.* **256** 3236–3256. MR2504524

[17] FRIZ, P. and VICTOIR, N. (2008). On uniformly subelliptic operators and stochastic area. *Probab. Theory Related Fields* **142** 475–523. MR2438699

[18] FRIZ, P. K. and HAIRER, M. (2014). *A course on rough paths. Universitext*. Springer, Cham With an introduction to regularity structures. MR3289027

[19] FRIZ, P. K. and SHEKHAR, A. (2017). General rough integration, Lévy rough paths and a Lévy-Kintchine-type formula. *Ann. Probab.* **45** 2707–2765. MR3693973

[20] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional stochastic processes as rough paths. Cambridge Studies in Advanced Mathematics* **120**. Cambridge University Press, Cambridge. MR2604669

[21] FRIZ, P. K. and ZHANG, H. (2017). Differential equations driven by rough paths with jumps. *ArXiv e-prints*.

[22] FUJIWARA, T. (1991). Stochastic differential equations of jump type on manifolds and Lévy flows. *J. Math. Kyoto Univ.* **31** 99–119. MR1093330

[23] FUJIWARA, T. and KUNITA, H. (1985). Stochastic differential equations of jump type and Lévy processes in diffeomorphisms group. *J. Math. Kyoto Univ.* **25** 71–106. MR777247

[24] FUJIWARA, T. and KUNITA, H. (1999). Canonical SDE's based on semimartingales with spatial parameters. II. Inverse flows and backward SDE's. *Kyushu J. Math.* **53** 301–331. MR1713105

[25] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit theorems for stochastic processes*, second ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer-Verlag, Berlin. MR1943877

[26] JAKUBOWSKI, A., MÉMIN, J. and PAGÈS, G. (1989). Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbf{D}^1$ de Skorokhod. *Probab. Theory Related Fields* **81** 111–137. MR981569

[27] KALLENBERG, O. (1997). *Foundations of modern probability. Probability and its Applications (New York)*. Springer-Verlag, New York. MR1464694

[28] KELLY, D. and MELBOURNE, I. (2016). Smooth approximation of stochastic differential equations. *Ann. Probab.* **44** 479–520. MR3456344

[29] KUNITA, H. (1990). *Stochastic flows and stochastic differential equations. Cambridge Studies in Advanced Mathematics* **24**. Cambridge University Press, Cambridge. MR1070361

[30] KUNITA, H. (1996). Stochastic differential equations with jumps and stochastic flows of diffeomorphisms. In *Itô's stochastic calculus and probability theory* 197–211. Springer, Tokyo. MR1439526

[31] KURTZ, T. G., PARDOUX, É. and PROTTER, P. (1995). Stratonovich stochastic differential equations driven by general semimartingales. *Ann. Inst. H. Poincaré Probab. Statist.* **31** 351–377. MR1324812

[32] KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070. MR1112406

[33] KURTZ, T. G. and PROTTER, P. E. (1996). Weak convergence of stochastic integrals and differential equations. In *Probabilistic models for nonlinear partial differential equations (Montecatini Terme, 1995). Lecture Notes in Math.* **1627** 1–41. Springer, Berlin. MR1431298

[34] LEJAY, A. and LYONS, T. (2005). On the importance of the Lévy area for studying the limits of functions of converging stochastic processes. Application to homogenization. In *Current trends in potential theory. Theta Ser. Adv. Math.* **4** 63–84. Theta, Bucharest. MR2243956

[35] LENGLART, E., LÉPINGLE, D. and PRATELLI, M. (1980). Présentation unifiée de certaines inégalités de la théorie des martingales. In *Seminar on Probability, XIV (Paris, 1978/1979) (French). Lecture Notes in Math.* **784** 26–52. Springer, Berlin With an appendix by Lenglart. MR580107

[36] LÉPINGLE, D. (1976). La variation d'ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **36** 295–316. MR0420837

[37] Lyons, T. and Ni, H. (2015). Expected signature of Brownian motion up to the first exit time from a bounded domain. *Ann. Probab.* **43** 2729–2762. MR3395473

[38] Lyons, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoamericana* **14** 215–310. MR1654527

[39] Marcus, S. (1978). Modeling and analysis of stochastic differential equations driven by point processes. *IEEE Transactions on Information Theory* **24** 164-172.

[40] Marcus, S. I. (1981). Modeling and approximation of stochastic differential equations driven by semimartingales. *Stochastics* **4** 223-245.

[41] McShane, E. J. (1972). Stochastic differential equations and models of random processes. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory* 263–294. Univ. California Press, Berkeley, Calif. MR0402921 (53 ##6735)

[42] Meyer, P.-A. (1972). *Martingales and stochastic integrals. I. Lecture Notes in Mathematics. Vol. 284.* Springer-Verlag, Berlin-New York. MR0426145

[43] Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2. Cambridge Mathematical Library.* Cambridge University Press, Cambridge Itô calculus, Reprint of the second (1994) edition. MR1780932

[44] Simon, T. (2003). Small deviations in p-variation for multidimensional Lévy processes. *J. Math. Kyoto Univ.* **43** 523–565. MR2028666

[45] Whitt, W. (2002). *Stochastic-process limits. Springer Series in Operations Research.* Springer-Verlag, New York An introduction to stochastic-process limits and their application to queues. MR1876437

[46] Williams, D. R. E. (1998). Differential equations driven by discontinuous paths, PhD thesis, Imperial College London.

[47] Williams, D. R. E. (2001). Path-wise solutions of stochastic differential equations driven by Lévy processes. *Rev. Mat. Iberoamericana* **17** 295–329. MR1891200

I. Chevyrev
Mathematical Institute
University of Oxford
Andrew Wiles Building
Radcliffe Observatory Quarter
Woodstock Road
Oxford OX2 6GG
United Kingdom
E-mail: chevyrev@maths.ox.ac.uk

P. K. Friz
Institut für Mathematik
Technische Universität Berlin
Strasse des 17. Juni 136
10623 Berlin
Germany
and
Weierstrass–Institut für Angewandte
Analysis und Stochastik
Mohrenstrasse 39
10117 Berlin
Germany
E-mail: friz@math.tu-berlin.de