

# Random walks and Lévy processes as rough paths

Ilya Chevyrev<sup>1</sup> 

Received: 9 November 2015 / Revised: 22 February 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** We consider random walks and Lévy processes in a homogeneous group  $G$ . For all  $p > 0$ , we completely characterise (almost) all  $G$ -valued Lévy processes whose sample paths have finite  $p$ -variation, and give sufficient conditions under which a sequence of  $G$ -valued random walks converges in law to a Lévy process in  $p$ -variation topology. In the case that  $G$  is the free nilpotent Lie group over  $\mathbb{R}^d$ , so that processes of finite  $p$ -variation are identified with rough paths, we demonstrate applications of our results to weak convergence of stochastic flows and provide a Lévy–Khintchine formula for the characteristic function of the signature of a Lévy process. At the heart of our analysis is a criterion for tightness of  $p$ -variation for a collection of càdlàg strong Markov processes.

**Keywords** Homogeneous groups · Rough paths · Lévy processes · Random walks · Tightness of  $p$ -variation · Stochastic flows · Characteristic functions of signatures

**Mathematics Subject Classification** Primary 60G51; Secondary 60H10

## 1 Introduction

This paper focuses on several questions regarding Lévy processes and random walks in homogeneous groups, with a particular focus on applications to rough paths theory.

---

The author is supported by a Junior Research Fellowship of St John's College, Oxford.

---

✉ Ilya Chevyrev  
chevyrev@maths.ox.ac.uk

<sup>1</sup> Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK

Let  $G$  be a homogeneous group (in the sense of [15]) equipped with a sub-additive homogeneous norm and the corresponding left-invariant metric. We can summarise the three main results of the paper as follows.

- (Theorem 5.1) Given a Lévy process  $\mathbf{X}$  in  $G$ , we determine (almost) all values of  $p > 0$  for which the sample paths of  $\mathbf{X}$  have almost surely finite  $p$ -variation.
- (Theorem 5.5) We give sufficient conditions for a sequence of (interpolated and reparametrised) random walks in  $G$  to converge weakly to a (interpolated and reparametrised) Lévy process in  $G$  in  $p$ -variation topology.
- (Theorem 5.17) In the case that  $G = G^N(\mathbb{R}^d)$ , the step- $N$  free nilpotent Lie group over  $\mathbb{R}^d$ , we determine a Lévy–Khinchine formula for the characteristic function (in the sense of [11]) of the signature of the random rough path constructed from a Lévy process in  $G$ .

We apply the second of these results in the context of rough paths to show weak convergence of stochastic flows in several examples. Notably, we provide a significant generalisation of a result of Kunita [29] and of a related result of Breuillard, Friz and Huesmann [7].

We take a moment to discuss how our work relates to the appearance of càdlàg rough paths in the current literature. Friz and Shekhar [17] recently introduced a broad extension of rough paths theory to the càdlàg setting. Their work in particular generalises the notion of rough integration and RDEs and significantly extends earlier work of Williams [36] who gave pathwise solutions to differential equations driven by Lévy processes in  $\mathbb{R}^d$ .

As a family of càdlàg rough paths of particular interest, Lévy process in  $G^N(\mathbb{R}^d)$  of finite  $p$ -variation for some  $1 \leq p < N + 1$ , were studied in [17]. Such Lévy  $p$ -rough paths bear a resemblance to Markovian rough paths constructed from subelliptic Dirichlet forms on  $L^2(G^N(\mathbb{R}^d))$ , first studied in [19] and recently in [10–12], in the sense that both processes may be viewed as stochastic rough paths whose evolution depends entirely on its first  $N$  iterated integrals.

The method we employ here to give meaning to càdlàg rough paths is to connect left- and right-limits with continuous paths and treat the resulting object as a classical rough path. We therefore do not address directly the concept of a càdlàg RDE in this paper, but emphasise that our methods relate closely to Marcus SDEs and that Theorems 5.1 and 5.17 can be seen as generalisations of two related results in [17] (discussed further in Sect. 5). We mention however that the method of proof used for our main results, which is based on approximating a Lévy process by a sequence of random walks, is different to the methods used in [17].

We also point out that our methods treat general interpolations, which depend arbitrarily on the endpoints of jumps, on the same footing as the simpler linear interpolation used in Marcus SDEs. Examples of interest of such non-linear interpolations date back to the works of McShane [33] and Sussman [35] on approximations of Brownian motion (discussed further in Examples 5.12 and 5.14), and recently in the work of Flint, Hambly and Lyons [14].

A crucial result for our analysis, which we believe to be of independent interest, is a criterion for tightness of  $p$ -variation of strong Markov processes taking values in a Polish space (Theorem 4.8). This result is a generalisation of the main result of

Manstavičius [32], which provides a criterion for a strong Markov process to have sample paths of a.s. finite  $p$ -variation. Our proof of Theorem 4.8 is a simplification of the stopping times technique adopted in [32].

Finally, we mention that while most applications presented in this paper concern geometric rough paths, and thus only require consideration of the free nilpotent Lie group, we have attempted to make statements in their natural level of generality. In particular, we believe that our results may prove to be of interest for studying random walks and Lévy processes in the Butcher group, which correspond to branched rough paths in the sense of [21, 22] (see also Remark 3.2 below).

## Outline of the paper

In Sect. 2 we discuss iid arrays and Lévy processes taking values in a general Lie group. Our only contribution in this section is the construction of a sequence of random walks  $(\mathbf{X}^n)_{n \geq 1}$  associated with a Lévy process  $\mathbf{X}$  such that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  in the Skorokhod topology, and for which tightness of  $p$ -variation is simple to verify. In Sect. 3 we recall several preliminary facts about homogeneous groups and spaces of paths of finite  $p$ -variation.

Section 4 is devoted to the proof of Theorem 4.1, which shows tightness of  $p$ -variation for a collection of random walks in a homogeneous group. This is a central result used in the proofs of the three main aforementioned theorems, which we state and prove in Sect. 5. In Sect. 5.3.1 we also provide several applications of Theorem 5.5 to weak convergence of stochastic flows.

In “Appendix A” we introduce the concept of path functions, which serve to connect the left- and right-limits of càdlàg paths, and collect several technical results used throughout Sect. 5. In “Appendix B” we describe conditions under which sample paths of a Lévy process possess infinite  $p$ -variation (used to complete the proof of Theorem 5.1).

## 2 Iid arrays and Lévy processes in Lie groups

### 2.1 Notation

Throughout this section, we fix a Lie group  $G$  with Lie algebra  $\mathfrak{g}$ , and identify  $\mathfrak{g}$  with the space of left-invariant vector fields on  $G$ . Let  $u_1, \dots, u_m$  be a basis for  $\mathfrak{g}$ . We equip  $\mathfrak{g}$  with the inner product for which  $u_1, \dots, u_m$  is an orthonormal basis. For an element  $y \in \mathfrak{g}$  we write  $y = \sum_{i=1}^m y^i u_i$ . When  $x$  is an element of a normed space, we denote its norm by  $|x|$ .

We further fix an open neighbourhood  $U \subset G$  of the identity  $1_G \in G$ , such that  $U$  has compact closure and  $\exp : \mathfrak{g} \mapsto G$  is a diffeomorphism from a neighbourhood of zero in  $\mathfrak{g}$  onto  $U$ . Let  $\xi_i \in C_c^\infty(G, \mathbb{R})$  be smooth functions of compact support such that  $\log(x) = \sum_{i=1}^m \xi_i(x) u_i$  for all  $x \in U$  (that is,  $\xi_i$  provide exponential coordinates of the first kind on  $U$ ). We denote  $\xi : G \mapsto \mathfrak{g}$ ,  $\xi(x) = \sum_{i=1}^m \xi_i(x) u_i$ .

For a metric space  $E$ , denote by  $D([0, T], E)$  the space of càdlàg functions  $\mathbf{x} : [0, T] \mapsto E$  equipped with the Skorokhod topology (see, e.g., [3, Section 12]). We shall use the symbol  $\mathcal{o}$  to denote spaces of paths whose starting point is the identity element  $1_G$ . For example  $D_{\mathcal{o}}([0, T], G)$  denotes the set of all  $\mathbf{x} \in D([0, T], G)$  such that  $\mathbf{x}_0 = 1_G$ .

## 2.2 Preliminaries on iid arrays and Lévy processes

An *array* in  $G$  is a sequence of a finite collection of  $G$ -valued random variables  $(X_{n1}, \dots, X_{nn})_{n \geq 1}$ . We call the array *iid* if, for every  $n \geq 1$ ,  $X_{n1}, \dots, X_{nn}$  are iid. We will always suppose that an iid array  $X_{nj}$  is infinitesimal, i.e.,  $\lim_{n \rightarrow \infty} \mathbb{P}[X_{n1} \notin V] = 0$  for every neighbourhood  $V$  of  $1_G$ . Furthermore, for all  $n \geq 1$  we let

$$B_n := \mathbb{E}[\xi(X_{n1})] \in \mathfrak{g},$$

and for all  $i, j \in \{1, \dots, m\}$

$$A_n^{i,j} := \mathbb{E}[\xi_i(X_{n1})\xi_j(X_{n1})].$$

For a collection of elements  $x_1, \dots, x_n$  in  $G$ , we define the *associated walk*  $\mathbf{x} \in D_{\mathcal{o}}([0, 1], G)$  by

$$\mathbf{x}_t = \begin{cases} 1_G & \text{if } t \in [0, n^{-1}) \\ x_1 \dots x_{\lfloor tn \rfloor} & \text{if } t \in [n^{-1}, 1], \end{cases}$$

and for an array  $X_{nj}$ , we refer to the associated random walk  $\mathbf{X}^n$  to mean the sequence of associated walks built from the collections  $(X_{n1}, \dots, X_{nn})$ .

Recall that a (left) Lévy process in  $G$  is a  $D_{\mathcal{o}}([0, T], G)$ -valued random variable  $\mathbf{X}$  with independent and stationary (right) increments. We refer to Liao [30] for further details.

We call a *Lévy triplet* (or simply *triplet*) a collection  $(A, B, \Pi)$  of an  $m \times m$  covariance matrix  $(A^{i,j})_{i,j=1}^m$ , an element  $B = \sum_{i=1}^m B^i u_i \in \mathfrak{g}$ , and a Lévy measure  $\Pi$  on  $G$  (see [30, p. 12]).

A classical theorem of Hunt [26] asserts that for every Lévy process  $\mathbf{X}$  in  $G$ , there exists a unique triplet  $(A, B, \Pi)$  such that the generator of  $\mathbf{X}$  is given for all  $f \in C_0^2(G)$  and  $x \in G$  by

$$\begin{aligned} \lim_{t \rightarrow 0} t^{-1} \mathbb{E}[f(x\mathbf{X}_t) - f(x)] &= \sum_{i=1}^m B^i (u_i f)(x) + \frac{1}{2} \sum_{i,j=1}^m A^{i,j} (u_i u_j f)(x) \\ &\quad + \int_G \left[ f(xy) - f(x) - \sum_{i=1}^m \xi_i(y) (u_i f)(x) \right] \Pi(dy). \end{aligned}$$

Conversely, every Lévy triplet gives rise to a unique Lévy process.

We will heavily use a characterisation due to Feinsilver [13] of when a  $G$ -valued random walk converges in law to a Markov process as a  $D_o([0, 1], G)$ -valued random variable. The following is a special case of the main results of [13].

**Theorem 2.1** (Feinsilver [13]). *Let  $X_{nj}$  be an iid array of  $G$ -valued random variables and  $\mathbf{X}^n$  the associated random walk. Denote by  $F_n$  the probability measure on  $G$  associated with  $X_{n1}$ . Let  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ .*

*Then  $\mathbf{X}^n \xrightarrow{D} \mathbf{X}$  as  $D_o([0, 1], G)$ -valued random variables if and only if*

- (1)  $\lim_{n \rightarrow \infty} nF_n(f) = \Pi(f)$  for every  $f \in C_b(G)$  which is identically zero on a neighbourhood of  $1_G$ ,
- (2)  $\lim_{n \rightarrow \infty} nB_n = B$ , and
- (3) for all  $i, j \in \{1, \dots, m\}$ ,

$$\lim_{n \rightarrow \infty} nA_n^{i,j} = A^{i,j} + \int_G \xi_i(x) \xi_j(x) \Pi(dx).$$

The following notion of a scaling function will be used throughout the paper.

**Definition 2.2** (Scaling function). A continuous bounded function  $\theta : G \rightarrow \mathbb{R}$  is called a *scaling function* if

- (i)  $\theta(1_G) = 0$ ,
- (ii)  $\theta(x) > 0$  for all  $x \neq 1_G$ ,
- (iii) there exists  $C > 0$  such that  $|\xi|^2 \leq C\theta$ , and
- (iv) there exists  $c > 0$  such that  $\theta(x) > c$  for all  $x \in G \setminus U$ .

Let  $X_{nj}$  be an iid array in  $G$ . We say that  $\theta$  *scales* the array  $X_{nj}$  if

$$\sup_{n \geq 1} n\mathbb{E}[\theta(X_{n1})] < \infty.$$

The importance behind the above definition is that given a scaling function  $\theta$  which scales  $X_{nj}$ , the rate with which  $\theta$  decays at  $1_G$  will determine the values of  $p > 0$  for which the  $p$ -variation of the associated random walk is tight (Theorem 4.1).

**Example 2.3** In the case  $G = \mathbb{R}^d$ , the prototypical example of a scaling function is  $1 \wedge |\cdot|^2$ . For a general Lie group  $G$ , the example extends as follows: let  $c > 0$  be sufficiently small such that  $W := \{\exp(y) \mid y \in \mathfrak{g}, |y| \leq c\}$  is contained in  $U$ . Then

$$\theta(x) := \mathbf{1}\{x \in U\}(c^2 \wedge |\xi(x)|^2) + c^2 \mathbf{1}\{x \notin U\} \quad (2.1)$$

is a scaling function.

**Remark 2.4** Suppose that  $\theta$  is defined by (2.1) and that  $X_{nj}$  is an iid array in  $G$  such that the associated random walk converges in law to a Lévy process. Then a simple consequence of Theorem 2.1 is that  $\theta$  scales the array  $X_{nj}$ .

### 2.3 Approximating walk

In this subsection, given a Lévy process  $\mathbf{X}$  in  $G$ , we construct an iid array  $X_{nj}$  for which the associated random walk  $\mathbf{X}^n$  converges in law to  $\mathbf{X}$ . The array  $X_{nj}$  has the advantage that it takes values in either the support of the Lévy measure of  $\mathbf{X}$ , or in a set which shrinks to the identity as  $n \rightarrow \infty$ . This makes the walk  $\mathbf{X}^n$  significantly easier to analyse than the increments of  $\mathbf{X}$  itself and will be used in the proofs of Theorems 5.1 and 5.17.

Throughout this subsection, let  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ . For  $i \in \{1, \dots, m\}$  define

$$\Gamma_i := \left\{ 0 \leq \gamma < 2 \mid \int_G |\xi_i(x)|^\gamma \Pi(dx) = \infty \right\}.$$

Define also the sets of indexes

$$J = \left\{ j \in \{1, \dots, m\} \mid A^{j,j} > 0 \right\},$$

$$\tilde{K} = \{k \in \{1, \dots, m\} \mid 1 \notin \Gamma_k\}.$$

For  $k \in \tilde{K}$  define

$$\tilde{B}^k = B^k - \int_G \xi_k(x) \Pi(dx),$$

and let  $K = \{k \in \tilde{K} \mid \tilde{B}^k \neq 0\}$ .

For  $n$  sufficiently large so that  $\Pi(U^c) < n/2$ , let

$$h_n = \inf \{h \geq 0 \mid \Pi(\{|\xi(x)| > h\} \cup U^c) \leq n/2\}.$$

Define  $U_n = \{x \in U \mid |\xi(x)| \leq h_n\}$  and note that  $w_n := \Pi\{U_n^c\} \leq n/2$ . Remark that  $\lim_{n \rightarrow \infty} h_n = 0$  which implies that  $U_n$  shrinks to  $1_G$  as  $n \rightarrow \infty$ .

Define on  $G$  the probability measure  $\mu_n(dx) := w_n^{-1} \mathbf{1}_{\{x \in U_n^c\}} \Pi(dx)$ . Observe that by Hölder's inequality, for all  $q \geq 1$

$$\int_{U_n^c} |\xi_i(x)| \Pi(dx) \leq (n/2)^{1-1/q} \left( \int_G |\xi_i(x)|^q \Pi(dx) \right)^{1/q}. \quad (2.2)$$

For every  $n \geq 1$ , let  $Y_n = Y_n^1 u_1 + \dots + Y_n^m u_m$  be a  $\mathfrak{g}$ -valued random variable such that for all  $k \in \tilde{K}$

$$b_n^k := \mathbb{E}[Y_n^k] = (1 - w_n/n)^{-1} n^{-1} \tilde{B}^k,$$

and for all  $k \notin \tilde{K}$

$$b_n^k := \mathbb{E} \left[ Y_n^k \right] = (1 - w_n/n)^{-1} n^{-1} \left( B^k - \int_{U_n^c} \xi_k(x) \Pi(dx) \right),$$

and with covariances for all  $i, j \in \{1, \dots, m\}$

$$\mathbb{E} \left[ (Y_n^i - b_n^i)(Y_n^j - b_n^j) \right] = (1 - w_n/n)^{-1} n^{-1} A^{i,j}.$$

In particular, note that  $Y_n^i = b_n^i$  a.s. for all  $i \notin J$ . Remark that setting  $q = 2$  in (2.2) implies

$$\lim_{n \rightarrow \infty} n^{-1} \int_{U_n^c} |\xi_i(x)| \Pi(dx) = 0,$$

from which it follows that  $\sup_{n \geq 1} n|b_n^i| < \infty$ . Moreover, it holds that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[ (Y_n^i - b_n^i)(Y_n^j - b_n^j) \right] = 0.$$

It follows that we can choose  $Y_n$  such that  $\exp(Y_n)$  has support in a neighbourhood  $V_n$  of  $1_G$ , such that  $V_n$  shrinks to  $1_G$  as  $n \rightarrow \infty$ . Denote by  $\nu_n$  the probability measure of the  $G$ -valued random variable  $\exp(Y_n)$ .

Finally, let  $X_{n1}$  be the  $G$ -valued random variable associated to the probability measure  $(w_n/n)\mu_n + (1 - w_n/n)\nu_n$ , and let  $X_{n2}, \dots, X_{nn}$  be independent copies of  $X_{n1}$ .

Consider the random walk  $\mathbf{X}^n$  associated with  $X_{nj}$ . Then a straightforward application of Theorem 2.1 implies that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D_o([0, 1], G)$ -valued random variables. We also record the following two simple lemmas whose proofs we omit.

**Lemma 2.5** *Let  $0 < q_1, \dots, q_m \leq 2$  be real numbers such that  $q_i \notin \Gamma_i$  for all  $i \in \{1, \dots, m\}$ ,  $q_i = 2$  for all  $i \in J$ , and  $q_i \geq 1$  for all  $i \in K$ . Let  $\theta$  be a scaling function such that  $\theta(x) = \sum_{i=1}^m |\xi_i(x)|^{q_i}$  for  $x$  in a neighbourhood of  $1_G$ . Then  $\theta$  scales the array  $X_{n1}, \dots, X_{nn}$ .*

**Lemma 2.6** *Let  $\theta$  be a scaling function on  $G$  which scales  $X_{nj}$ . Let  $V$  be a neighbourhood of  $1_G$ , and let  $f : \text{supp}(\Pi) \cup V \mapsto \mathbb{R}$  be a bounded measurable function such that  $f$  is continuous on  $\text{supp}(\Pi)$ . Furthermore, suppose that*

$$\lim_{x \rightarrow 1_G} \frac{1}{\theta(x)} \left| f - f(1_G) - \sum_{i=1}^m b_i \xi_i(x) - \frac{1}{2} \sum_{i,j=1}^m a_{i,j} \xi_i(x) \xi_j(x) \right| = 0.$$

*Then for all  $n$  sufficiently large,  $X_{n1} \in \text{supp}(\Pi) \cup V$  a.s., and*

$$\lim_{n \rightarrow \infty} n \mathbb{E} [f(X_{n1}) - f(1_G)] = Q,$$

where

$$Q := \sum_{i=1}^m B^i b_i + \frac{1}{2} \sum_{i,j=1}^m A^{i,j} a_{i,j} + \int_G \left[ f(x) - f(1_G) - \sum_{i=1}^m b_i \xi_i(x) \right] \Pi(dx).$$

### 3 Homogeneous groups

In this section we collect several preliminary facts about homogeneous groups. For details, we refer to [15] and [25].

Throughout this section, we fix a homogeneous group  $G$ . That is,  $G$  is a nilpotent, connected, and simply connected Lie group endowed with a one-parameter family of dilations (group automorphisms)  $(\delta_\lambda)_{\lambda>0}$ , which, upon identifying  $G$  with its Lie algebra  $\mathfrak{g}$  by the exp map, is given by

$$\delta_\lambda(u_i) = \lambda^{d_i} u_i$$

for a basis  $u_1, \dots, u_m$  of  $\mathfrak{g}$  and real numbers  $d_m \geq \dots \geq d_1 \geq 1$ . We equip  $G$  with a sub-additive homogeneous norm  $\|\cdot\|$  which induces a left-invariant metric  $d(x, y) = \|x^{-1}y\|$  (see [25]).

For the remainder of the section, we identify  $G$  with  $\mathfrak{g}$  by the diffeomorphism  $\exp : \mathfrak{g} \mapsto G$ , and write  $x = \sum x^i u_i$  for  $x \in G$ . Note that  $\|x\| = \sum_{i=1}^m |x^i|^{1/d_i}$  is also a homogeneous norm on  $G$  and thus equivalent to  $\|\cdot\|$ .

For a multi-index  $\alpha = (\alpha^1, \dots, \alpha^m)$ ,  $\alpha^i \geq 0$ , we define  $\deg(\alpha) = \sum_{i=1}^m \alpha^i d_i$ , and for  $x \in G$ , write  $x^\alpha = (x^1)^{\alpha^1} \dots (x^m)^{\alpha^m}$ . By the Campbell–Baker–Hausdorff (CBH) formula, for all  $i \in \{1, \dots, m\}$  there exist constants  $C_{\alpha,\beta}^i$  such that

$$(xy)^i = x^i + y^i + \sum_{\alpha,\beta} C_{\alpha,\beta}^i x^\alpha y^\beta, \quad (3.1)$$

where the (finite) sum runs over all non-zero multi-indexes  $\alpha, \beta$  such that  $\deg(\alpha) + \deg(\beta) = d_i$ .

**Example 3.1** Recall that a Lie group  $G$  is called graded if its Lie algebra is endowed with a decomposition

$$\mathfrak{g} = \mathfrak{g}^1 \oplus \dots \oplus \mathfrak{g}^N \quad (3.2)$$

such that  $[\mathfrak{g}^i, \mathfrak{g}^j] \subseteq \mathfrak{g}^{i+j}$ , where  $\mathfrak{g}^k = 0$  for  $k > N$  (and where we allow the possibility that  $\mathfrak{g}^k = 0$  for some  $k \leq N$ ). Every graded Lie group can be equipped with a natural family of dilations  $(\delta_\lambda)_{\lambda>0}$ , and thus a homogeneous structure, for which  $d_1, \dots, d_m$  are rational numbers with  $d_1 = 1$ , given by  $\delta_\lambda(u) = \lambda^{k/\alpha} u$  for all  $u \in \mathfrak{g}^k$ , where  $\alpha = \min\{k \geq 1 \mid \mathfrak{g}^k \neq 0\}$  (and conversely, if  $d_1, \dots, d_m$  are rational for a homogeneous group  $G$ , then  $G$  can be given a graded structure [15, p. 5]).

Recall also that a graded Lie group  $G$  is called a step- $N$  Carnot group (or stratified group in the terminology of [15]) if the decomposition (3.2) further satisfies  $[\mathfrak{g}_i, \mathfrak{g}_j] = \mathfrak{g}_{i+j}$ , where  $\mathfrak{g}_k = 0$  for  $k > N$ . Every Carnot group is a homogeneous group with

a natural family of dilations given by  $\delta_\lambda(u) = \lambda^k u$  for all  $u \in \mathfrak{g}_k$  (so that  $d_i \in \{1, \dots, N\}$ ), and for which the metric  $d$  can be taken as the Carnot–Carathéodory distance [2, p. 38].

The Carnot group which will be particularly relevant in Sect. 5.3 for applications in rough paths theory is the step- $N$  free nilpotent Lie group  $G^N(\mathbb{R}^d)$  over  $\mathbb{R}^d$ , which we recall is, by definition, the space where geometric  $p$ -rough paths (for  $\lfloor p \rfloor = N$ ) take value. For further details concerning the theory of geometric rough paths, we refer to [18].

**Remark 3.2** Another homogeneous group which plays an important role in the theory of rough paths is the step- $N$  Butcher group  $\mathcal{G}^N(\mathbb{R}^d)$  over  $\mathbb{R}^d$  (see [21, 22]). Recall that  $G^N(\mathbb{R}^d)$  is canonically embedded in  $\mathcal{G}^N(\mathbb{R}^d)$ , and that  $\mathcal{G}^N(\mathbb{R}^d)$  admits a natural grading under which  $\mathcal{G}^N(\mathbb{R}^d)$  is not a Carnot group (see [22, Remark 2.15]).

The group  $\mathcal{G}^N(\mathbb{R}^d)$  is, by definition, the space where branched rough paths take value (which form a genuine extension of the notion of geometric rough paths). We mention that branched rough paths were recently studied in [8] to give a rough path perspective on renormalisation of stochastic PDEs in the theory of regularity structures [9, 23]. Lévy processes in  $\mathcal{G}^N(\mathbb{R}^d)$  in particular form a family of stationary stochastic processes closed under appropriate renormalisation maps (see [8, Section 4]).

### 3.1 Paths of finite $p$ -variation

For  $p > 0$  and functions  $\mathbf{x}, \mathbf{y} : [s, t] \mapsto G$ , define the  $p$ -variation distance

$$d_{p\text{-var};[s,t]}(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}_s, \mathbf{y}_s) + \sup_{\mathcal{D} \subset [s,t]} \left( \sum_{t_j \in \mathcal{D}} d(\mathbf{x}_{t_j, t_{j+1}}, \mathbf{y}_{t_j, t_{j+1}})^p \right)^{1/p},$$

where the supremum runs over all partitions  $\mathcal{D}$  of  $[s, t]$  and where we have used the shorthand notation  $\mathbf{x}_{u,v} = \mathbf{x}_u^{-1} \mathbf{x}_v$ . Define also the  $p$ -variation of  $\mathbf{x}$  by  $\|\mathbf{x}\|_{p\text{-var};[s,t]} = d_{p\text{-var};[s,t]}(\mathbf{x}, 1_G)$ , and

$$d_{0;[s,t]}(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}_s, \mathbf{y}_s) + \sup_{u,v \in [s,t]} d(\mathbf{x}_{u,v}, \mathbf{y}_{u,v})$$

and

$$d_{\infty;[s,t]}(\mathbf{x}, \mathbf{y}) = \sup_{u \in [s,t]} d(\mathbf{x}_u, \mathbf{y}_u), \quad \|\mathbf{x}\|_{\infty;[s,t]} = d_{\infty;[s,t]}(\mathbf{x}, 1_G).$$

We will drop the reference to the interval  $[s, t]$  when it is clear from the context. For convenience, we record the following standard interpolation estimates.

**Lemma 3.3** (1) For all  $p' > p > 0$  and  $\mathbf{x}, \mathbf{y} : [s, t] \mapsto G$

$$d_{p'\text{-var}}(\mathbf{x}, \mathbf{y}) \leq 2^{\max\{0, (1-p)/p'\}} (\|\mathbf{x}\|_{p\text{-var}} + \|\mathbf{y}\|_{p\text{-var}})^{p/p'} d_0(\mathbf{x}, \mathbf{y})^{1-p/p'}.$$

(2) There exists  $C > 0$  such that for all  $\mathbf{x}, \mathbf{y} : [s, t] \mapsto G$  with  $\mathbf{x}_s = \mathbf{y}_s$

$$d_\infty(\mathbf{x}, \mathbf{y}) \leq d_0(\mathbf{x}, \mathbf{y}) \leq C \max\{d_\infty(\mathbf{x}, \mathbf{y}), d_\infty(\mathbf{x}, \mathbf{y})^{1/d_m}(\|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty)^{1-1/d_m}\}.$$

*Proof* (1) is obvious. To show (2), it follows from an application of the CBH formula (3.1) and the equivalence of  $\|\cdot\|$  and  $\|\cdot\|_m$ , that for all  $g, h \in G$

$$\left\| g^{-1}hg \right\| \leq C_1 \max\{\|h\|, \|h\|^{1/d_m} \|g\|^{1-1/d_m}\}.$$

The conclusion now follows by the identical argument used to prove [18, Proposition 8.15].  $\square$

For  $p \geq 1$ , let  $C^{p\text{-var}}([0, T], G)$  denote the space of continuous paths of finite  $p$ -variation equipped with the metric  $d_{p\text{-var};[0,T]}$ . Note that  $C^{p\text{-var}}([0, T], G)$  is a complete metric space due to the lower semi-continuity of  $\mathbf{x} \mapsto \|\mathbf{x}\|_{p\text{-var};[0,T]}$  (under pointwise convergence).

Note that, except in trivial cases,  $C^{p\text{-var}}([0, T], G)$  is non-separable. However, it is not difficult to show that  $C^{p'\text{-var}}([0, T], G)$  contains a separable subset  $C^{0,p'\text{-var}}([0, T], G)$  which contains  $C^{p\text{-var}}([0, T], G)$  for all  $1 \leq p < p'$ . Indeed, let  $C^g([0, T], G)$  denote the space of curves which are concatenations of one-parameter subgroups of  $G$ , i.e., all curves  $\gamma : [0, T] \mapsto G$  of the form

$$\gamma(t) = \gamma(t_{k-1}) \exp\left(\frac{t - t_{k-1}}{t_k - t_{k-1}} \log x_k\right), \quad t \in [t_{k-1}, t_k], \quad k \in \{1, \dots, n\}, \quad (3.3)$$

where  $\mathcal{D} = (t_0 = 0 < t_1 < \dots < t_n = T)$  is a partition of  $[0, T]$  and  $x_1, \dots, x_n \in G$  (and where for clarity we have broken the convention of identifying  $G$  with  $\mathfrak{g}$ ). Then for  $p \geq 1$ , define  $C^{0,p\text{-var}}([0, T], G)$  as the closure of  $C^g([0, T], G) \cap C^{p\text{-var}}([0, T], G)$  in  $C^{p\text{-var}}([0, T], G)$ .

**Remark 3.4** In the case that  $G$  is a Carnot group with decomposition (3.2),  $C^{0,p\text{-var}}([0, T], G)$  is precisely the closure of the horizontal lifts of smooth paths  $\gamma \in C^\infty([0, T], \mathfrak{g}^1)$ .

To show the claimed properties of  $C^{0,p\text{-var}}([0, T], G)$ , note that for  $x \in G$ , the path  $\gamma : t \mapsto \exp(t \log x)$  has finite  $p$ -variation if and only if  $x^i = 0$  for all  $i \in \{1, \dots, m\}$  such that  $d_i > p$ , in which case there exists  $C_1 = C_1(p, G) > 0$  such that  $\|\gamma\|_{p\text{-var};[0,1]} \leq C_1 \|x\|$ . For  $\mathbf{x} : [0, T] \mapsto G$ , and a partition  $\mathcal{D} \subset [0, T]$ , let  $\mathbf{x}^\mathcal{D} \in C^g([0, T], G)$  be the interpolation of  $\mathbf{x}$  along  $\mathcal{D}$  defined as  $\gamma$  in (3.3) with  $x_k = \mathbf{x}_{t_{k-1}}^{-1} \mathbf{x}_{t_k}$  and  $\mathbf{x}_0^\mathcal{D} = \mathbf{x}_0$ . One can then readily show (e.g., by Lemma A.5) that  $\sup_{\mathcal{D} \subset [0,T]} \|\mathbf{x}^\mathcal{D}\|_{p\text{-var}} \leq C_2 \|\mathbf{x}\|_{p\text{-var}}$ . Hence for all  $\mathbf{x} \in C^{p\text{-var}}([0, T], G)$  and  $p' > p \geq 1$ , by Lemma 3.3,  $d_{p'\text{-var};[0,T]}(\mathbf{x}^\mathcal{D}, \mathbf{x}) \rightarrow 0$  as  $|\mathcal{D}| \rightarrow 0$ , which shows that  $C^{p\text{-var}}([0, T], G) \subseteq C^{0,p'\text{-var}}([0, T], G)$  as claimed. The fact that  $C^{0,p\text{-var}}([0, T], G)$  is separable (and thus Polish) is also easy to show (e.g., by considering  $\gamma \in C^g([0, T], G)$  with rational coordinates and using a similar argument as the proof of Lemma A.5).

The following result will be important in our classification of  $G$ -valued Lévy processes of finite  $p$ -variation.

**Proposition 3.5** *Let  $p > 0$  and  $(\mathbf{X}_n)_{n \geq 1}$  be a sequence of  $D([0, T], G)$ -valued random variables such that  $(\|\mathbf{X}_n\|_{p\text{-var};[0,T]})_{n \geq 1}$  is a tight collection of real random variables. Suppose that  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D([0, T], G)$ -valued random variables.*

- (1) *It holds that  $\|\mathbf{X}\|_{p\text{-var};[0,T]} < \infty$  a.s..*
- (2) *Suppose further that  $p \geq 1$  and  $\mathbf{X}^n, \mathbf{X}$  are  $C([0, T], G)$ -valued random variables. Then for all  $p' > p$ ,  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $C^{0,p'\text{-var}}([0, T], G)$ -valued random variables.*

*Proof* (1) Note that  $\mathbf{x} \mapsto \|\mathbf{x}\|_{p\text{-var}}$  is a lower semi-continuous function on  $D$ . Since  $D([0, T], G)$  is Polish, we may apply the Skorokhod representation theorem [27, Theorem 3.30], from which the conclusion easily follows.

- (2) It follows from Lemma 3.3 that every set of the form  $A \cap \{\mathbf{x} \in C([0, T], G) \mid \|\mathbf{x}\|_{p\text{-var};[0,T]} < R\}$ , where  $R > 0$  and  $A$  is a compact subset of  $C([0, T], G)$  (for uniform topology), is a compact subset of  $C^{0,p'\text{-var}}([0, T], G)$ . Hence  $(\mathbf{X}_n)_{n \geq 1}$  is a tight collection of  $C^{0,p'\text{-var}}([0, T], G)$ -valued r.v.'s, and so converges in law along a subsequence to some  $C^{0,p'\text{-var}}([0, T], G)$ -valued r.v.  $\tilde{\mathbf{X}}$ . Since  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $C([0, T], G)$ -valued r.v.'s, it necessarily follows that  $\tilde{\mathbf{X}} \stackrel{\mathcal{D}}{=} \mathbf{X}$ , which concludes the proof.  $\square$

**Remark 3.6** A version of Helly's selection principle (see [34, Theorem 2.4]) states that any uniformly bounded sequence of functions  $\mathbf{x}^n : [0, T] \mapsto G$  for which  $\sup_{n \geq 1} \|\mathbf{x}^n\|_{p\text{-var};[0,T]} < \infty$  for some  $p \geq 1$ , has a subsequence such that  $\mathbf{x}^{n_k} \rightarrow \mathbf{x}$  pointwise.

## 4 $p$ -variation tightness of random walks

We continue to use the notation of the previous section. Consider an iid array  $X_{nk}$  in the homogeneous group  $G$ , and let  $\mathbf{X}^n$  be the associated random walk. The main result of this section is Theorem 4.1, which provides sufficient conditions under which  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight. In its simplest form, Theorem 4.1 implies that whenever  $\mathbf{X}^n$  converges in law to a Lévy process in  $G$ , and the array  $X_{nk}$  is scaled by a scaling function  $\theta$ , then  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight for all  $p > \kappa > 0$ , where  $\kappa$  depends only on the scaling function  $\theta$ .

Let  $\xi_1, \dots, \xi_m \in C_c^\infty(G)$  and  $\xi : G \mapsto \mathfrak{g}$  be smooth functions and  $U$  a neighbourhood of  $1_G$  for which the conditions at the start of Sect. 2 are satisfied with respect to the basis  $u_1, \dots, u_m$ .

**Theorem 4.1** *Let  $X_{n1}, \dots, X_{nm}$  be an iid array of  $G$ -valued random variables and  $\mathbf{X}^n$  the associated random walk. For every  $i \in \{1, \dots, m\}$ , let  $0 < q_i \leq 2$  be a real number, and define*

$$\kappa = \max\{q_1 d_1, \dots, q_m d_m\}.$$

Consider the following conditions:

- (A) for every fixed  $h \in [0, 1]$ ,  $(\mathbf{X}_h^n)_{n \geq 1}$  is a tight collection of  $G$ -valued random variables;
- (B) for all  $i \in \{1, \dots, m\}$ ,  $\sup_{n \geq 1} n \|\mathbb{E}[\xi_i(X_{n1})]\| < \infty$ ;
- (C) the array  $X_{nk}$  is scaled by a scaling function  $\theta$ , where  $\theta \equiv \sum_{i=1}^m |\xi_i|^{q_i}$  on a neighbourhood of  $1_G$ .

Then, provided (A), (B) and (C) hold,  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D_o([0, 1], G)$ -valued random variables and, for every  $p > \kappa$ ,  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is a tight collection of real random variables.

**Remark 4.2** Suppose that for a Lévy process  $\mathbf{X}$  in  $G$ ,  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D_o([0, T], G)$ -valued random variables. Then conditions (A) and (B) are automatically satisfied by Theorem 2.1 (and (C) is satisfied upon choosing  $q_i = 2$  for all  $i \in \{1, \dots, m\}$  by Remark 2.4).

The remainder of the section is devoted to the proof of Theorem 4.1, which can be split into three parts. The first part is collected in Sect. 4.1 and comprises a general  $p$ -variation tightness criterion for strong Markov processes. The second part, which is the most technical part of the proof, is collected in Sect. 4.2 and establishes the bounds required to apply the results of Sect. 4.1 for the case  $p > d_m$ . The third part is collected in Sect. 4.3 and treats the case  $p \leq d_m$ . Roughly speaking, in the third part we decompose  $\mathbf{X}^n$  into the lift of a walk in a lower level group, for which the previous two parts apply, and a perturbation on the higher levels, for which the  $p$ -variation can be controlled directly.

#### 4.1 $p$ -Variation tightness of strong Markov processes

In this section we give a criterion for  $p$ -variation tightness of strong Markov processes in a Polish space (Theorem 4.8), which is inspired by the work of Manstavičius [32].

Let  $(E, d)$  be a metric space and  $\mathbf{x} : [0, T] \mapsto E$  a function. Define

$$M(\mathbf{x}) := \sup_{s, t \in [0, T]} d(\mathbf{x}_t, \mathbf{x}_s),$$

and, for  $\delta > 0$ ,

$$\nu_\delta(\mathbf{x}) := \sup \left\{ k \geq 0 \mid \exists (t_i)_{i=1}^{2k}, t_1 < t_2 \leq t_3 < \dots < t_{2k}, d(\mathbf{x}_{t_{2i}}, \mathbf{x}_{t_{2i-1}}) > \delta, i \leq k \right\}.$$

Note that quantity  $\nu_\delta(\mathbf{x})$  measures the maximum number of oscillations of  $\mathbf{x}$  of magnitude greater than  $\delta$  over non-overlapping intervals. Observe the following basic inequality which serves to control  $\|\mathbf{x}\|_{p\text{-var};[0,T]}$ :

$$\|\mathbf{x}\|_{p\text{-var};[0,T]}^p \leq \sum_{r=1}^{\infty} 2^{-r p + p} \nu_{2^{-r}}(\mathbf{x}) + M(\mathbf{x})^p \nu_1(\mathbf{x}). \quad (4.1)$$

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

$$\tau_j^\delta(\mathbf{x}) = \begin{cases} \inf \left\{ t \in [\tau_{j-1}^\delta(\mathbf{x}), T] \mid \sup_{u,v \in [\tau_{j-1}^\delta(\mathbf{x}), t]} d(\mathbf{x}_u, \mathbf{x}_v) > \delta \right\} \\ \infty & \text{if the above set is empty.} \end{cases}$$

**Lemma 4.3** *Let  $\mathbf{X}$  be a  $D([0, T], E)$ -valued random variable for a Polish space  $(E, d)$ . Let  $\delta, h > 0$  such that there exists  $q \in (0, 1)$  for which a.s. for all  $i \geq 0$*

$$\mathbb{P}[\tau_{i+1}^\delta(\mathbf{X}) - \tau_i^\delta(\mathbf{X}) \leq h \mid \tau_i^\delta(\mathbf{X}), \dots, \tau_0^\delta(\mathbf{X})] \leq q \quad (4.2)$$

(where we use the convention  $\infty - \infty = \infty$ ). Then

$$\mathbb{E}[v_\delta(\mathbf{X})] \leq \lceil T/h \rceil \frac{1}{1-q}.$$

*Proof* Note that for any function  $\mathbf{x} : [0, T] \mapsto E$ , it holds that  $v_\delta(\mathbf{x})$  is the largest integer  $j$  for which  $\tau_j^\delta(\mathbf{x}) \leq T$ , and thus

$$\mathbb{P}[v_\delta(\mathbf{X}) \geq j] = \mathbb{P}[\tau_j^\delta(\mathbf{X}) \leq T].$$

For  $i \geq 0$ , consider the event  $A_i = \{\tau_{i+1}^\delta(\mathbf{X}) - \tau_i^\delta(\mathbf{X}) > h\}$ , and note that

$$\mathbb{P}[\tau_j^\delta(\mathbf{X}) \leq T] \leq \mathbb{P}[\text{at most } \lceil T/h \rceil \text{ of } (A_i)_{i=0}^{j-1} \text{ occur}].$$

Consider a real random variable  $Z$  distributed by the negative binomial distribution with parameters  $(\lceil T/h \rceil, q)$ , i.e.,  $Z$  counts the total number of iid Bernoulli trials with success probability  $q$  until exactly  $\lceil T/h \rceil$  failures occur. It follows from the uniform bound (4.2) that

$$\mathbb{P}[\text{at most } \lceil T/h \rceil \text{ of } (A_i)_{i=0}^{j-1} \text{ occur}] \leq \mathbb{P}[Z \geq j]$$

(where one considers  $A_i$  as a failure with probability at least  $1 - q$ ), so that

$$\mathbb{E}[v_\delta(\mathbf{X})] \leq \mathbb{E}[Z] = \lceil T/h \rceil \frac{1}{1-q}.$$

□

We now show how one can verify the condition of Lemma 4.3 for a strong Markov process. We first restrict attention to the set of times on which a process is allowed to move.

**Definition 4.4** For a metric space  $(E, d)$  and a  $D([0, T], E)$ -valued random variable  $\mathbf{X}$ , call a (deterministic) open interval  $(s, t) \subset [0, T]$  *stationary* if

$$\mathbb{P}[\forall u \in (s, t), \mathbf{X}_u = \mathbf{X}_s] = 1.$$

Let  $Z_{\mathbf{X}} \subseteq [0, T]$  denote the union of all stationary intervals, and let  $R_{\mathbf{X}} = [0, T] \setminus Z_{\mathbf{X}}$  be its complement.

*Example 4.5* For the random walk  $\mathbf{X}^n \in D([0, 1], G)$  associated with an iid array  $X_{nj}$  in a Lie group  $G$ , we have  $R_{\mathbf{X}^n} = \{0, 1/n, \dots, (n-1)/n, 1\}$ .

We emphasise that the role of  $R_{\mathbf{X}}$  is only technical in that it allows us to easily formulate bounds uniform in  $s \in R_{\mathbf{X}}$  (such as those in Theorem 4.8 and Corollary 4.9) which hold for random walks and for which the same bounds would not hold when taken uniformly over all  $s \in [0, T]$  (though for completely harmless reasons). The following lemma is a variant of Gihman–Skorokhod [20, Lemma 2, p. 420] (in which the notion of  $R_{\mathbf{X}}$  does not appear).

**Lemma 4.6** (Maximum inequality). *Let  $\mathbf{X}$  be a càdlàg (not necessarily strong) Markov process taking values in a Polish space  $(E, d)$ .*

*Let  $h, \delta > 0$  and suppose there exists  $c \in [0, 1)$  such that*

$$\sup_{s \in R_{\mathbf{X}}} \sup_{x \in E} \sup_{t \in [s, s+h]} \mathbb{P}^{s,x} [d(\mathbf{X}_s, \mathbf{X}_t) > \delta] \leq c. \quad (4.3)$$

*Then for all  $s \in R_{\mathbf{X}}$  and  $x \in E$ , it holds that*

$$\mathbb{P}^{s,x} \left[ \sup_{t \in [s, s+h]} d(\mathbf{X}_s, \mathbf{X}_t) > 2\delta \right] \leq \frac{\mathbb{P}^{s,x} [d(\mathbf{X}_s, \mathbf{X}_{s+h}) > \delta]}{1 - c}.$$

*Proof* Let  $s \in R_{\mathbf{X}}$  and observe that a.s.

$$\sup_{t \in [s, s+h]} d(\mathbf{X}_s, \mathbf{X}_t) = \sup_{t \in [s, s+h] \cap R_{\mathbf{X}}} d(\mathbf{X}_s, \mathbf{X}_t).$$

Consider a nested sequence of partitions  $\mathcal{D}_n \subset [s, s+h] \cap R_{\mathbf{X}}$  such that

$$\lim_{n \rightarrow \infty} \sup_{t \in [s, s+h] \cap R_{\mathbf{X}}} d_+(t, \mathcal{D}_n) = 0,$$

where  $d_+(t, \mathcal{D}_n) := t - \sup \{u \leq t \mid u \in \mathcal{D}_n\}$ . Since  $\mathbf{X}$  is càdlàg, it holds that

$$\sup_{t \in [s, s+h] \cap R_{\mathbf{X}}} d(\mathbf{X}_s, \mathbf{X}_t) = \lim_{n \rightarrow \infty} \sup_{t_i \in \mathcal{D}_n} d(\mathbf{X}_s, \mathbf{X}_{t_i}),$$

where the right side is non-decreasing in  $n$  since  $\mathcal{D}_n$  are nested.

It thus suffices to show that for any partition  $\mathcal{D} = (t_0 = s, \dots, t_n) \subset [s, s+h] \cap R_{\mathbf{X}}$ , we have

$$\mathbb{P}^{s,x} \left[ \sup_{t_i \in \mathcal{D}} d(\mathbf{X}_s, \mathbf{X}_{t_i}) > 2\delta \right] \leq \frac{\mathbb{P}^{s,x} [d(\mathbf{X}_s, \mathbf{X}_{s+h}) > \delta]}{1-c}. \quad (4.4)$$

To this end, for  $i \in \{0, \dots, n\}$ , consider the events

$$C_i := \{d(\mathbf{X}_{t_i}, \mathbf{X}_{s+h}) > \delta\}$$

and

$$B_i := \{d(\mathbf{X}_s, \mathbf{X}_{t_i}) > 2\delta\}.$$

Define the  $\sigma$ -algebras  $\mathcal{F}_{s,t} := \sigma(\mathbf{X}_u)_{s \leq u \leq t}$ . Observe that (4.3) implies that a.s.

$$\mathbb{E}^{s,x} [\mathbf{1}\{C_i\} \mid \mathcal{F}_{s,t_i}] \leq c.$$

Moreover, consider the disjoint events  $F_i := B_1^c \cap \dots \cap B_{i-1}^c \cap B_i$ . Then for all  $i \in \{0, \dots, n\}$

$$F_i \cap C_i^c \subseteq C_0.$$

Since each  $F_i \cap C_i^c$  is disjoint and  $F_i$  is  $\mathcal{F}_{s,t_i}$ -measurable, we have

$$\begin{aligned} \mathbb{P}^{s,x} [C_0] &\geq \sum_{i=1}^n \mathbb{P}^{s,x} [F_i \cap C_i^c] \\ &\geq \sum_{i=1}^n \mathbb{E}^{s,x} [\mathbf{1}\{F_i\} \mathbb{E}^{s,x} [\mathbf{1}\{C_i^c\} \mid \mathcal{F}_{s,t_i}]] \\ &\geq (1-c) \sum_{i=1}^n \mathbb{P}^{s,x} [F_i]. \end{aligned}$$

Finally, (4.4) now follows from the fact that

$$\sum_{i=1}^n \mathbb{P}^{s,x} [F_i] = \mathbb{P}^{s,x} \left[ \sup_{t_i \in \mathcal{D}} d(\mathbf{X}_s, \mathbf{X}_{t_i}) > 2\delta \right].$$

□

**Corollary 4.7** *Let  $\mathbf{X}$  be a càdlàg strong Markov process taking values in a Polish space  $(E, d)$ . Let  $h, \delta > 0$  and  $c \in [0, 1)$  satisfy (4.3). Then for all  $i \geq 0$ , a.s.*

$$\mathbb{P} \left[ \tau_{i+1}^{4\delta}(\mathbf{X}) - \tau_i^{4\delta}(\mathbf{X}) \leq h \mid \tau_i^{4\delta}(\mathbf{X}), \dots, \tau_0^{4\delta}(\mathbf{X}) \right] \leq \frac{c}{1-c}.$$

*Proof* Observe that  $\tau_i^{4\delta}(\mathbf{X})$  takes values a.s. in  $R_{\mathbf{X}}$ , and that the event  $\{\tau_{i+1}^{4\delta}(\mathbf{X}) - \tau_i^{4\delta}(\mathbf{X}) \leq h\}$  is contained inside  $\{\sup_{t \in [\tau_i, \tau_i + h]} d(\mathbf{X}_{\tau_i}, \mathbf{X}_t) > 2\delta\}$ . Conditioning on the stopping times  $\{\tau_i^{4\delta}(\mathbf{X}), \dots, \tau_0^{4\delta}(\mathbf{X})\}$  and using the assumption that  $\mathbf{X}$  is a strong Markov process, the desired result now follows from Lemma 4.6.  $\square$

We now obtain the following  $p$ -variation tightness criterion for strong Markov processes. Recall the quantity  $M(\mathbf{X}) = \sup_{s, t \in [0, T]} d(\mathbf{X}_s, \mathbf{X}_t)$ .

**Theorem 4.8** *Let  $\mathcal{M}$  be collection of càdlàg strong Markov processes on  $[0, T]$  taking values in a Polish space  $(E, d)$ . Suppose that*

- (a)  $(M(\mathbf{X}))_{\mathbf{X} \in \mathcal{M}}$  is tight, and
- (b) *there exist constants  $a, \kappa, b > 0$  and  $c \in [0, 1/2)$  such that for all  $\delta \in (0, b]$*

$$\sup_{\mathbf{X} \in \mathcal{M}} \sup_{s \in R_{\mathbf{X}}} \sup_{x \in E} \sup_{t \in [s, s+h(\delta)]} \mathbb{P}^{s, x} [d(\mathbf{X}_s, \mathbf{X}_t) > \delta] \leq c,$$

where  $h(\delta) := a\delta^\kappa$ .

Then for any  $p > \kappa$ ,  $(\|\mathbf{X}\|_{p\text{-var}; [0, T]})_{\mathbf{X} \in \mathcal{M}}$  is a tight collection of real random variables.

*Proof* Let  $p > \kappa$ . We claim that it suffices to show

$$\sup_{\mathbf{X} \in \mathcal{M}} \sum_{r=0}^{\infty} 2^{-rp} \mathbb{E} [v_{2^{-r}}(\mathbf{X})] < \infty. \quad (4.5)$$

Indeed, observe that (4.5) implies that  $(v_1(\mathbf{X}))_{\mathbf{X} \in \mathcal{M}}$  is tight. It then follows, by (a) and the estimate (4.1), that (4.5) implies  $(\|\mathbf{X}\|_{p\text{-var}; [0, T]})_{\mathbf{X} \in \mathcal{M}}$  is tight as claimed.

It thus remains to show (4.5). By (b) and Corollary 4.7, it holds that for all  $\delta \in (0, b]$

$$\sup_{\mathbf{X} \in \mathcal{M}} \mathbb{P} \left[ \tau_{i+1}^{4\delta}(\mathbf{X}) - \tau_i^{4\delta}(\mathbf{X}) \leq h(\delta) \mid \tau_i^{4\delta}(\mathbf{X}), \dots, \tau_0^{4\delta}(\mathbf{X}) \right] \leq \frac{c}{1-c}.$$

Hence, by Lemma 4.3, for all  $\delta \in (0, b]$

$$\sup_{\mathbf{X} \in \mathcal{M}} \mathbb{E} [v_{4\delta}(\mathbf{X})] \leq \lceil T/h(\delta) \rceil \frac{1}{1 - \frac{c}{1-c}} \leq (1 + T\delta^{-\kappa}/a) \frac{1-c}{1-2c},$$

from which (4.5) readily follows.  $\square$

**Corollary 4.9** *Let  $(\mathbf{X}^n)_{n \geq 1}$  be a sequence of càdlàg strong Markov processes on  $[0, T]$  taking values in a Polish space  $(E, d)$ . Suppose that*

- (i) *for every fixed rational  $h \in [0, T]$ ,  $(\mathbf{X}_h^n)_{n \geq 1}$  is a tight collection of  $E$ -valued random variables, and*
- (ii) *there exist constants  $K, \beta, \gamma, b > 0$  such that for all  $\delta \in (0, b]$  and  $h > 0$*

$$\sup_{n \geq 1} \sup_{s \in R_{\mathbf{X}^n}} \sup_{x \in E} \sup_{t \in [s, s+h]} \mathbb{P}^{s, x} [d(\mathbf{X}_s^n, \mathbf{X}_t^n) > \delta] \leq K \frac{h^\beta}{\delta^\gamma}.$$

Then  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D([0, T], E)$ -valued random variables, and for any  $p > \gamma/\beta$ ,  $(\|\mathbf{X}^n\|_{p\text{-var};[0,T]})_{n \geq 1}$  is a tight collection of real random variables.

*Proof* First, note that (ii) applied to small  $h$  allows us to verify the Aldous condition for the sequence  $(\mathbf{X}^n)_{n \geq 1}$  (see, e.g., [28, p. 188], though note one should restrict attention to sequences of stopping times  $\tau_n$  taking values in  $R_{\mathbf{X}^n}$  a.s., which is a trivial modification to the usual Aldous condition). Together with (i), it follows that  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D([0, T], E)$ -valued random variables ([28, Theorems 4.8.1, 4.8.2]).

Observe that  $M$  is a continuous function on  $D([0, T], E)$ , from which it follows that  $(M(\mathbf{X}^n))_{n \geq 1}$  is tight. Moreover, observe that (ii) implies that there exists  $a > 0$  such that for all  $\delta \in (0, b]$

$$\sup_{n \geq 1} \sup_{s \in R_{\mathbf{X}^n}} \sup_{x \in E} \sup_{t \in [s, s+h]} \mathbb{P}^{s,x} [d(\mathbf{X}_s^n, \mathbf{X}_t^n) > \delta] \leq \frac{1}{3},$$

where  $h = a\delta^{\gamma/\beta}$ . It follows that the conditions of Theorem 4.8 are satisfied with  $\kappa = \gamma/\beta$ , so that indeed  $(\|\mathbf{X}^n\|_{p\text{-var};[0,T]})_{n \geq 1}$  is tight for all  $p > \gamma/\beta$ .  $\square$

## 4.2 Proof of Theorem 4.1 in the case $p > d_m$

We continue using the notation of Sect. 3. In particular, we identify  $G$  with  $\mathfrak{g}$  via the exp map.

*Remark 4.10* We note here that Corollary 4.9 and the bound (4.6) in the upcoming Lemma 4.12 are sufficient to establish that conditions (A), (B) and (C) imply that  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D_o([0, 1], G)$ -valued random variables and that  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight for all  $p > \kappa \vee d_m$ , which proves the statement of Theorem 4.1 subject to the restriction  $p > d_m$ .

Observe that an inductive application of the CBH formula (3.1), along with the multinomial identity  $(z_1 + \dots + z_n)^j = \sum_{k_1 + \dots + k_n = j} \binom{j}{k_1, \dots, k_n} z_1^{k_1} \dots z_n^{k_n}$ , yields the following lemma.

**Lemma 4.11** *For all  $x_1, \dots, x_k \in G$  and every index  $i \in \{1, \dots, m\}$ , it holds that*

$$(x_1 \dots x_k)^i = \sum_{1 \leq a_1 \leq k} x_{a_1}^i + \sum_{r=2}^{\lfloor d_i \rfloor} \sum_{\alpha_1, \dots, \alpha_r} \sum_{1 \leq a_1 < \dots < a_r \leq k} c_{\alpha_1, \dots, \alpha_r}^i x_{a_1}^{\alpha_1} \dots x_{a_r}^{\alpha_r},$$

where  $\sum_{\alpha_1, \dots, \alpha_r}$  indicates the (finite) sum over all non-zero multi-indexes  $\alpha_1, \dots, \alpha_r$  such that  $\deg(\alpha_1) + \dots + \deg(\alpha_r) = d_i$  and  $c_{\alpha_1, \dots, \alpha_r}^i$  are constants.

Recall that  $\mathbf{X}^n$  denotes the random walk associated to the iid array  $X_{n1}, \dots, X_{nn}$ .

**Lemma 4.12** *Use the notation from Theorem 4.1 and suppose that (B) and (C) hold. Let  $\gamma := d_m \vee \kappa$ , and for  $i \in \{1, \dots, m\}$  denote by  $\mathbf{Y}^{n,i} \in D_o([0, 1], \mathbb{R})$  the random walk associated with the  $\mathbb{R}$ -valued iid array  $X_{nk}^i$ .*

Then there exists  $K > 0$  such that for all  $n \geq 1$ ,  $k \in \{1, \dots, n\}$  and  $\delta \in (0, 1]$

$$\mathbb{P} \left[ \left\| \mathbf{X}_{k/n}^n \right\| > \delta \right] \leq K \frac{k/n}{\delta^\gamma}, \quad (4.6)$$

and, for all  $i \in \{1, \dots, m\}$  such that  $q_i \leq 1$ ,

$$\mathbb{P} \left[ \left\| \mathbf{Y}_{k/n}^{n,i} \right\| > \delta \right] \leq K \frac{k/n}{\delta^{q_i}}. \quad (4.7)$$

*Proof* We first claim that it suffices to consider the case  $\|X_{n1}\| \leq \varepsilon$  a.s. for all  $n \geq 1$ , where  $\varepsilon > 0$  may be taken arbitrarily small. Indeed, let  $\varepsilon > 0$  and note that there exists  $c > 0$  such that  $\theta(x) > c$  for all  $x \in G$  with  $\|x\| > \varepsilon$ . Since  $\theta$  scales  $X_{nk}$ , it follows that there exists  $C_1 > 0$  such that for all  $n \geq 1$

$$\mathbb{P} [\|X_{n1}\| \geq \varepsilon] \leq c^{-1} \mathbb{E} [\theta(X_{n1})] \leq C_1/n,$$

and hence

$$\mathbb{P} \left[ \max_{1 \leq a \leq k} \|X_{na}\| \geq \varepsilon \right] = 1 - (1 - \mathbb{P} [\|X_{n1}\| \geq \varepsilon])^k \leq C_1 k/n.$$

It follows that for all  $n \geq 1$  and  $k \in \{1, \dots, n\}$

$$\mathbb{P} \left[ \left\| \mathbf{X}_{k/n}^n \right\| > \delta \right] \leq \mathbb{P} \left[ \left\| \mathbf{X}_{k/n}^n \right\| > \delta, \max_{1 \leq a \leq k} \|X_{na}\| < \varepsilon \right] + C_1 k/n,$$

and similarly for  $\mathbb{P} \left[ \left\| \mathbf{Y}_{k/n}^{n,i} \right\| > \delta \right]$ . Replacing  $X_{nk}$  by

$$X'_{nk} = \begin{cases} X_{nk} & \text{if } \|X_{nk}\| < \varepsilon \\ 1_G & \text{otherwise,} \end{cases}$$

we note that (B) and (C) imply that the same conditions hold for the iid array  $X'_{nk}$ . It thus suffices to prove the statement of the lemma for the iid array  $X'_{nk}$  instead as claimed.

We henceforth assume that  $\|X_{n1}\| < \varepsilon$  a.s., where  $\varepsilon > 0$  is sufficiently small so that  $x \in U$  whenever  $\|x\| < \varepsilon$ . We first show (4.7). Let  $i \in \{1, \dots, m\}$  such that  $q_i \leq 1$ . Then there exists  $C_2 > 0$  such that

$$\mathbb{E} \left[ \left\| \mathbf{Y}_{k/n}^{n,i} \right\|^{q_i} \right] = \mathbb{E} \left[ \left\| \sum_{a=1}^k X'_{na} \right\|^{q_i} \right] \leq \mathbb{E} \left[ \sum_{a=1}^k \|X'_{na}\|^{q_i} \right] \leq C_2 k/n.$$

where the second inequality is due to (C). It follows by Markov's inequality that there exists  $K > 0$  such that (4.7) holds for all  $n \geq 1$ ,  $k \in \{1, \dots, n\}$ , and  $\delta \in (0, 1]$ .

We now show (4.6). By Lemma 4.11, it suffices to show that for all  $i \in \{1, \dots, m\}$ ,  $r \in \{1, \dots, \lfloor d_i \rfloor\}$ , multi-indexes  $\alpha_1, \dots, \alpha_r$  such that  $\deg(\alpha_1) + \dots + \deg(\alpha_r) = d_i$  (with  $\alpha_1^i = 1$  in the case that  $r = 1$ ), there exists  $K > 0$  such that for all  $n \geq 1$ ,  $k \in \{1, \dots, n\}$  and  $\delta \in (0, 1]$

$$\mathbb{P} \left[ \left| \sum_{1 \leq a_1 < \dots < a_r \leq k} X_{na_1}^{\alpha_1} \dots X_{na_r}^{\alpha_r} \right|^{1/d_i} > \delta \right] \leq K \frac{k/n}{\delta^\gamma}. \quad (4.8)$$

To this end, let us fix  $i \in \{1, \dots, m\}$ ,  $r \in \{1, \dots, \lfloor d_i \rfloor\}$ , and multi-indexes  $\alpha_1, \dots, \alpha_r$  such that  $\deg(\alpha_1) + \dots + \deg(\alpha_r) = d_i$ . Consider first the case  $r \geq 2$ . Define

$$\gamma_i := d_i \vee \max \{q_j d_j \mid j \in \{1, \dots, m\}, d_j \leq d_i\}.$$

By Markov's and Jensen's inequalities (observing that  $\gamma_i \leq 2d_i$ )

$$\begin{aligned} & \mathbb{P} \left[ \left| \sum_{1 \leq a_1 < \dots < a_r \leq k} X_{na_1}^{\alpha_1} \dots X_{na_r}^{\alpha_r} \right|^{1/d_i} > \delta \right] \\ & \leq \delta^{-\gamma_i} \mathbb{E} \left[ \left( \sum_{1 \leq a_1 < \dots < a_r \leq k} X_{na_1}^{\alpha_1} \dots X_{na_r}^{\alpha_r} \right)^2 \right]^{\gamma_i/2d_i}. \end{aligned} \quad (4.9)$$

To bound the last expression, for a multi-index  $\alpha = (\alpha^1, \dots, \alpha^m)$ , denote  $|\alpha| = \alpha^1 + \dots + \alpha^m$ . Note that due to the assumption  $\|X_{n1}\| < \varepsilon$  a.s., (B) is equivalent to

$$\sup_{|\alpha|=1} \sup_{n \geq 1} |\mathbb{E}[X_{n1}^\alpha]| < \infty. \quad (4.10)$$

Furthermore, by (C) and the Cauchy–Schwartz inequality,

$$\sup_{|\alpha| \geq 2} \sup_{n \geq 1} n \mathbb{E}[|X_{n1}^\alpha|] < \infty. \quad (4.11)$$

Consider now the expression

$$\mathbb{E} \left[ \left( \sum_{1 \leq a_1 < \dots < a_r \leq k} X_{na_1}^{\alpha_1} \dots X_{na_r}^{\alpha_r} \right)^2 \right]. \quad (4.12)$$

Since  $X_{n1}, \dots, X_{nn}$  are independent, (4.12) splits into a sum of terms of the form  $\mathbb{E}[X_{n1}^{\beta_1}] \dots \mathbb{E}[X_{nk}^{\beta_k}]$  with  $\beta_i \geq 0$ . Call the *simple degree* of such a term the number of  $\beta_i > 0$ . The minimum simple degree of any term is evidently  $r$  and the maximum

is  $2r$ , and one readily sees that there exists  $C_3 > 0$  such that for all  $n \geq 1$  and  $k \in \{1, \dots, n\}$ , the number of terms of simple degree  $s \in \{r, \dots, 2r\}$  is bounded above by  $C_3 k^s$ . Furthermore, since  $X_{n1}, \dots, X_{nn}$  are identically distributed, it follows from (4.10) and (4.11) that there exists  $C_4 > 0$  such that the absolute value of every term of simple degree  $s$  is bounded above by  $C_4 n^{-s}$ . Since  $2 \leq r \leq s$  and  $k \leq n$ , it follows that

$$\mathbb{E} \left[ \left( \sum_{1 \leq a_1 < \dots < a_r \leq k} X_{na_1}^{\alpha_1} \dots X_{na_r}^{\alpha_r} \right)^2 \right] \leq C_5 (k/n)^2.$$

Therefore, from (4.9) and the fact that  $d_i \leq \gamma_i \leq \gamma$ , we obtain (4.8). This completes the case  $r \geq 2$ .

It remains to consider the case  $r = 1$ . Define now  $\gamma_i := d_i(q_i \vee 1)$ . It holds that

$$\mathbb{P} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i \right|^{1/d_i} > \delta \right] \leq \delta^{-\gamma_j} \mathbb{E} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i \right|^{q_i \vee 1} \right].$$

Denote  $\mu_n = \mathbb{E} [X_{n1}^i]$ . Then there exist  $C_6, C_7 > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i \right|^{q_i \vee 1} \right] &= \mathbb{E} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i - \mu_n + \mu_n \right|^{q_i \vee 1} \right] \\ &\leq C_6 \mathbb{E} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i - \mu_n \right|^{q_i \vee 1} + |k/n|^{q_i \vee 1} \right] \\ &\leq C_7 \left( \mathbb{E} \left[ \sum_{1 \leq a \leq k} |X_{na}^i - \mu_n|^{q_i \vee 1} \right] + (k/n)^{q_i \vee 1} \right), \end{aligned}$$

where the first inequality is due to (4.10), and the second inequality is due to the (discrete) Burkholder–Davis–Gundy inequality and the fact that  $q_i \leq 2$ . It now follows from (C) and (4.10) that

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{1 \leq a \leq k} X_{na}^i \right|^{q_i \vee 1} \right] &\leq C_8 \left( k \mathbb{E} [ |X_{n1}^i|^{q_i \vee 1} ] + k |\mu_n|^{q_i \vee 1} + (k/n)^{q_i \vee 1} \right) \\ &\leq C_9 \left( k/n + kn^{-(q_i \vee 1)} + (k/n)^{q_i \vee 1} \right) \\ &\leq C_{10} (k/n). \end{aligned}$$

Since  $\gamma_j \leq \gamma$ , this completes the case  $r = 1$  and the proof of the lemma.  $\square$

As mentioned in Remark 4.10, Corollary 4.9 and the bound (4.6) are now sufficient to prove Theorem 4.1 for the case that  $p > d_m$ .

### 4.3 Proof of Theorem 4.1 in the case $p \leq d_m$

**Lemma 4.13** *Use the notation from Lemma 4.12 and suppose that (A), (B) and (C) hold. For all  $i \in \{1, \dots, m\}$ , it holds that  $(\mathbf{Y}_h^{n,i})_{n \geq 1, h \in [0,1]}$  is a tight collection of real random variables.*

*Proof* By Remark 4.10,  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D_o([0, 1], G)$ -valued random variables, from which it follows that  $(\max_{1 \leq k \leq n} |X_{nk}^i|)_{n \geq 1}$  is tight for all  $i \in \{1, \dots, m\}$ . We may thus suppose that  $\|X_{n1}\| \leq R$  a.s. for some large  $R > 0$  and all  $n \geq 1$ .

Consider the decomposition  $X_{nk}^i = A_{nk} + B_{nk}$  where

$$A_{nk} = X_{nk}^i \mathbf{1}\{\|X_{nk}\| < \varepsilon\}$$

and

$$B_{nk} = X_{nk}^i \mathbf{1}\{\varepsilon \leq \|X_{nk}\| \leq R\}.$$

We take here  $\varepsilon > 0$  sufficiently small so that  $\|x\| < \varepsilon$  implies  $x \in U$ . It suffices to prove that  $(\sum_{a=1}^k B_{na})_{n \geq 1, k \in \{1, \dots, n\}}$  and  $(\sum_{a=1}^k A_{na})_{n \geq 1, k \in \{1, \dots, n\}}$  are tight collections of real random variables.

Let  $C_1 = C_1(\varepsilon) > 0$  be such that  $C_1 \theta(x) > |x^i| \mathbf{1}\{\varepsilon \leq \|x\| \leq R\}$  for all  $x \in G$ . Since  $\theta$  scales  $X_{nk}$ , it holds that

$$\sup_{n \geq 1, k \in \{1, \dots, n\}} \mathbb{E} \left[ \left| \sum_{a=1}^k B_{na} \right| \right] \leq \sup_{n \geq 1} C_1 n \mathbb{E} [\theta(X_{n1})] < \infty,$$

and thus  $(\sum_{a=1}^k B_{na})_{n \geq 1, k \in \{1, \dots, n\}}$  is tight.

Now observe that (B) and (C) imply that  $\sup_{n \geq 1} n \|\mathbb{E}[A_{n1}]\| < \infty$ . Moreover (C) implies that there exists  $C_2 > 0$  such that  $|\xi_i(x)|^2 \leq C_2 \theta(x)$  for all  $x \in G$  and  $i \in \{1, \dots, m\}$ . Since  $A_{na} = \mathbf{1}\{\|X_{na}\| < \varepsilon\} \xi_i(X_{na})$ , and  $A_{n1}, \dots, A_{nn}$  are iid, it follows that

$$\sup_{n \geq 1, k \in \{1, \dots, n\}} \mathbb{E} \left[ \left( \sum_{a=1}^k A_{na} \right)^2 \right] \leq \sup_{n \geq 1} \sum_{a,b=1}^n |\mathbb{E}[A_{na} A_{nb}]| < \infty,$$

and thus  $(\sum_{a=1}^k A_{na})_{n \geq 1, k \in \{1, \dots, n\}}$  is also tight.  $\square$

For  $i \in \{1, \dots, m\}$ , let  $\mathfrak{g}^{>i}$  be the subspace of  $\mathfrak{g}$  spanned by  $\{u_j \mid j > i\}$ . Note that  $\mathfrak{g}^{>i}$  is an ideal of  $\mathfrak{g}$ , and so we can define the Lie algebra  $\mathfrak{g}^i = \mathfrak{g}/\mathfrak{g}^{>i}$  and the projection map  $\pi^i : \mathfrak{g} \mapsto \mathfrak{g}^i$ . The dilations  $\delta_\lambda$  on  $\mathfrak{g}$  give rise to a natural family of

dilations on  $\mathfrak{g}^i$ , and thus to a homogeneous group  $G^i$  associated with  $\mathfrak{g}^i$ . Equivalently,  $G^i = \mathfrak{g}/\mathfrak{g}^{>i}$ , where we have identified  $\mathfrak{g}$  with  $G$  and  $\mathfrak{g}^{>i}$  with a normal subgroup of  $G$ . We implicitly equip  $G^i$  with an arbitrary sub-additive homogeneous norm  $\|\cdot\|$ . For notational convenience, we also let  $G^0 = \{1\}$  be the trivial group and  $\pi^0 : G \mapsto G^0$  the trivial map.

**Corollary 4.14** *Use the notation from Lemma 4.12 and suppose that (A), (B) and (C) hold.*

- (i) *Let  $i \in \{1, \dots, m\}$  and  $p > d_i \vee \kappa$ . Then  $(\|\pi^i \mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight.*
- (ii) *For every  $i \in \{1, \dots, m\}$  such that  $q_i \leq 1$ ,  $(\|\mathbf{Y}^{n,i}\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight for all  $p > q_i$ .*

*Proof* (i) Observe that  $\pi^i \mathbf{X}^n$  is the random walk associated with the  $G^i$ -valued iid array  $\pi^i X_{nk}$ , from which the conclusion follows by Corollary 4.9 and the bound (4.6) of Lemma 4.12 (cf. Remark 4.10).

- (ii) From Corollary 4.9 and the bound (4.7) of Lemma 4.12, it suffices to check that condition (i) of Corollary 4.9 holds for the processes  $(\mathbf{Y}^{n,i})_{n \geq 1}$ . However this follows from Lemma 4.13. □

Recall that we identify  $G$  with  $\mathfrak{g}$  via the exp map. For functions  $\mathbf{z} : [0, T] \mapsto G$  and  $\mathbf{y} : [0, T] \mapsto \mathbb{R}$ , define the function

$$\mathbf{x} = \mathbf{z} + \mathbf{y} : [0, T] \mapsto G, \quad \mathbf{x}_t = \mathbf{z}_t + \mathbf{y}_t u_m.$$

(where addition is taken in  $\mathfrak{g}$ ). The following lemma is a simple consequence of the fact that  $(\mathbf{x}_s^{-1} \mathbf{x}_t)^i = (\mathbf{z}_s^{-1} \mathbf{z}_t)^i$  for all  $i \in \{1, \dots, m-1\}$ , and

$$(\mathbf{x}_s^{-1} \mathbf{x}_t)^m = (\mathbf{z}_s^{-1} \mathbf{z}_t)^m + \mathbf{y}_t - \mathbf{y}_s.$$

**Lemma 4.15** *Let  $\mathbf{z} : [0, T] \mapsto G$  and  $\mathbf{y} : [0, T] \mapsto \mathbb{R}$  be functions, and let  $\mathbf{x} = \mathbf{z} + \mathbf{y}$ . Then for any  $p > 0$  there exists  $C = C(p, G) > 0$  such that*

$$\|\mathbf{x}\|_{p\text{-var};[0,T]} \leq C \left( \|\mathbf{z}\|_{p\text{-var};[0,T]} + \|\mathbf{y}\|_{p/d_m\text{-var};[0,T]}^{1/d_m} \right).$$

**Lemma 4.16** *Let  $p > 0$  and  $i \in \{1, \dots, m\}$  be the largest index such that  $d_i \leq p$  (with  $i = 0$  if no such index exists). Consider elements  $x_1, \dots, x_n \in G$  and let  $\mathbf{x} \in D_o([0, 1], G)$  be the associated walk. For  $j \in \{i+1, \dots, m\}$ , let  $\mathbf{y}^j \in D_o([0, 1], \mathbb{R})$  be the walk associated with the real numbers  $x_1^j, \dots, x_n^j$ .*

*Then there exists  $C = C(p, G) > 0$ , such that*

$$\|\mathbf{x}\|_{p\text{-var};[0,1]} \leq C \left( \|\pi^i \mathbf{x}\|_{p\text{-var};[0,1]} + \sum_{j=i+1}^m \|\mathbf{y}^j\|_{p/d_j\text{-var};[0,1]}^{1/d_j} \right).$$

*Proof* By induction on  $m$  and Lemma 4.15, it suffices to show that if  $p < d_m$  and  $x_k^m = 0$  for all  $k \in \{1, \dots, n\}$ , then  $\|\mathbf{x}\|_{p\text{-var};[0,1]} \leq C_1 \|\pi^{m-1}\mathbf{x}\|_{p\text{-var};[0,1]}$ . This in turn follows from the CBH formula (3.1) and an application of Young's partition coarsening argument (see, e.g., [31, p. 50]).  $\square$

We now have all the ingredients for the proof of Theorem 4.1.

*Proof of Theorem 4.1* The fact that  $(\mathbf{X}^n)_{n \geq 1}$  is a tight collection of  $D_o([0, 1], G)$ -valued random variables follows directly from Corollary 4.9 and the bound (4.6) of Lemma 4.12 (cf. Remark 4.10).

Let  $p > \kappa$ . Decreasing  $p$  if necessary, we may suppose  $p \neq d_i$  for all  $i \in \{1, \dots, m\}$ . Let  $i \in \{1, \dots, m\}$  be the largest index such that  $d_i < p$  (with  $i = 0$  if no such index exists). Define  $\mathbf{Y}^{n,j}$  as in Lemma 4.12, and note that  $q_j < p/d_j < 1$  for all  $j \in \{i+1, \dots, m\}$ . It follows by Corollary 4.14 that  $(\|\pi^i \mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  and  $(\|\mathbf{Y}^{n,j}\|_{p/d_j\text{-var};[0,1]})_{n \geq 1}$  are tight for all  $j \in \{i+1, \dots, m\}$ . We conclude by Lemma 4.16 that  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is also tight.  $\square$

## 5 Lévy processes in homogeneous groups

### 5.1 Finite $p$ -variation of Lévy processes

Consider a homogeneous group  $G$  and recall the notation of Sect. 3. Recall also the definitions of  $\Gamma_i$ ,  $J$ , and  $K$  from Sect. 2.3. The following is the main result of this subsection.

**Theorem 5.1** *Let  $p > 0$  and  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ .*

- (1) *Then  $\|\mathbf{X}\|_{p\text{-var};[0,1]} < \infty$  a.s. provided that all of the following hold:*
  - (i)  $p > 2d_j$  for all  $j \in J$ ;
  - (ii)  $p > d_k$  for all  $k \in K$ ;
  - (iii)  $p/d_i > \sup\{\Gamma_i\}$  for all  $i \in \{1, \dots, m\}$ .
- (2) *Then  $\|\mathbf{X}\|_{p\text{-var};[0,1]} = \infty$  a.s. provided that one of the following holds:*
  - (iv)  $p \leq 2d_j$  for some  $j \in J$ ;
  - (v)  $p < d_k$  for some  $k \in K$ ;
  - (vi)  $p/d_i \in \Gamma_i$  for some  $i \in \{1, \dots, m\}$ .

*Remark 5.2* Note that Theorem 5.1 does not completely determine all values of  $p$  for which  $\|\mathbf{X}\|_{p\text{-var};[0,1]} < \infty$  a.s. (e.g., when  $p/d_i = \sup\{\Gamma_i\} \notin \Gamma_i$  for some  $i \in \{1, \dots, m\}$ ). Comparing Theorem 5.1 with known results for  $\mathbb{R}$ -valued Lévy processes [6], we suspect that (ii) and (iii) can be replaced by  $p \geq d_k, \forall k \in K$ , and  $p/d_i \notin \Gamma_i, \forall i \in \{1, \dots, m\}$ , respectively, which would complete the characterisation.

*Remark 5.3* In [17], the authors determined sufficient conditions under which a Lévy process in the step-2 free nilpotent Lie group  $G^2(\mathbb{R}^d)$  possesses finite  $p$ -variation for  $p \in (2, 3)$ , along with a partial converse that their conditions cannot in general be weakened ([17, Theorem 50]). In this context, Theorem 5.1 generalises this result to all  $N \geq 1$  and  $p > 0$  and provides a sharp converse. In particular, the Carnot–Carathéodory Blumenthal–Gettoor index  $\beta$  introduced in [17] for a Lévy measure on

$G^N(\mathbb{R}^d)$  relates to our definition of  $\Gamma_i$  by  $\beta = \max\{d_1 \sup\{\Gamma_1\}, \dots, d_m \sup\{\Gamma_m\}\}$ , in which case (iii) reads  $p > \beta$ .

For the proof of Theorem 5.1, we require the following lemma.

**Lemma 5.4** *Let  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ . Assume  $p > 0$  satisfies (i), (ii), and (iii) of Theorem 5.1.*

*Let  $X_{nj}$  be the associated iid array constructed in Sect. 2.3 and  $\mathbf{X}^n$  the associated random walk. Then  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight.*

*Proof* Let  $0 < p' < p$  such that  $p'$  also satisfies (i), (ii), and (iii) of Theorem 5.1. For all  $i \in \{1, \dots, m\}$ , define  $q_i := 2 \wedge (p'/d_i)$ , and let  $\theta$  be a scaling function on  $G$  such that  $\theta \equiv \sum_{i=1}^m |\xi_i|^{q_i}$  in a neighbourhood of  $1_G$ .

Observe that  $q_i \notin \Gamma_i$  for all  $i \in \{1, \dots, m\}$ ,  $q_j = 2$  for all  $j \in J$ , and  $q_k > 1$  for all  $k \in K$ . Thus, by Lemma 2.5,  $\theta$  scales the array  $X_{nj}$ . Moreover, since  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D_o([0, 1], G)$ -valued random variables, it follows that the array  $X_{nj}$  satisfies the conditions of Theorem 4.1 with the above  $\theta$  and  $q_1, \dots, q_m$  (see Remark 4.2). Since  $p > \max\{q_1 d_1, \dots, q_m d_m\}$ , it follows that  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight.  $\square$

*Proof of Theorem 5.1* (1) follows from Lemma 5.4 and part (1) of Proposition 3.5, while (2) follows directly from Corollary B.3 and Proposition B.4.  $\square$

## 5.2 Convergence in $p$ -variation

In this subsection we consider continuous random paths  $(\mathbf{X}^{n,\phi})_{n \geq 1}$ ,  $\mathbf{X}^\phi$ , constructed from a random walk  $\mathbf{X}^n$  and a Lévy process  $\mathbf{X}$  by connecting their left- and right-limits with a path function  $\phi$ , and give conditions under which  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C^{p\text{-var}}([0, 1], G)$ -valued random variables. All relevant material on path functions is collected in “Appendix A”.

**Theorem 5.5** *Let  $X_{nj}$  be an iid array in  $G$  and  $\mathbf{X}^n$  the associated random walk.*

*Let  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ . Suppose that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D_o([0, 1], G)$ -valued random variables and that  $\theta$  scales  $X_{nj}$ , where  $\theta \equiv \sum_{i=1}^m |\xi_i|^{q_i}$  in a neighbourhood of  $1_G$  for some  $0 < q_i \leq 2$ .*

*Let  $W \subseteq G$  be a closed subset such that  $\text{supp}(\Pi) \subseteq W$  and  $X_{n1} \in W$  a.s. for all  $n \geq 1$ . Let  $p > \max\{1, q_1 d_1, \dots, q_m d_m\}$  and  $\phi : W \mapsto C_o^{p\text{-var}}([0, 1], G)$  a  $p$ -approximating, endpoint continuous path function.*

*Then  $\|\mathbf{X}^\phi\|_{p\text{-var};[0,1]} < \infty$  a.s., and for every  $p' > p$ ,  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C_o^{0,p'\text{-var}}([0, 1], G)$ -valued random variables.*

**Remark 5.6** In the statement of Theorem 5.5, note that, a.s.,  $\mathbf{X}_{t-}^{-1} \mathbf{X}_t \in \text{supp}(\Pi)$  for every jump time  $t$  of  $\mathbf{X}$  (e.g., [30, Proposition 1.4]). Hence, for any (measurable) path function  $\phi$  defined on  $\text{supp}(\Pi)$ ,  $\mathbf{X}^\phi$  is indeed a well-defined  $C_o([0, 1], G)$ -valued random variable.

*Proof* By Theorem 4.1, it holds that  $(\|\mathbf{X}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$  is tight, and thus, by Proposition A.7,  $(\|\mathbf{X}^{n,\phi}\|_{p\text{-var};[0,1]})_{n \geq 1}$  is also tight. Since  $\phi$  is endpoint continuous on  $W$ , it follows by Proposition A.4 that  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C_o([0, 1], G)$ -valued random variables. The conclusion now follows from Proposition 3.5.  $\square$

### 5.3 Applications to rough paths theory

We apply the results so far developed in the paper to the theory of rough paths and stochastic flows. Following Example 3.1, denote by  $G^N(\mathbb{R}^d)$  the step- $N$  free nilpotent Lie group over  $\mathbb{R}^d$  and let  $\mathfrak{g}^N(\mathbb{R}^d)$  be its Lie algebra. For the remainder of the paper, unless otherwise stated, we shall always let  $G = G^N(\mathbb{R}^d)$  and  $\mathfrak{g} = \mathfrak{g}^N(\mathbb{R}^d)$ . Being a Carnot group,  $G$  comes equipped with a natural homogeneous structure and we note that  $u_1, \dots, u_d$  can be identified with a basis of  $\mathbb{R}^d$ .

For  $1 \leq p < N + 1$ , we let  $WG\Omega_p(\mathbb{R}^d) := C^{p\text{-var}}([0, T], G)$ , equipped with the metric  $d_{p\text{-var};[0,T]}$ , denote the space of weakly geometric  $p$ -rough paths. Given an element  $\mathbf{x} \in WG\Omega_p(\mathbb{R}^d)$ , and a collection  $(f_i)_{i=1}^d$  of vector fields in  $\text{Lip}^{\gamma+k-1}(\mathbb{R}^e)$  for  $\gamma > p \geq 1$  and an integer  $k \geq 1$ , there is a unique solution to the rough differential equation (RDE)

$$d\mathbf{y}_t = f(\mathbf{y}_t)d\mathbf{x}_t, \quad \mathbf{y}_0 \in \mathbb{R}^e. \quad (5.1)$$

We refer to [18] for further details on (geometric) rough paths theory.

#### 5.3.1 Stochastic flows

Let  $U_{T \leftarrow 0}^{\mathbf{x}} : \mathbf{y}_0 \mapsto \mathbf{y}_T$  denote the flow associated to (5.1), which we recall is an element of  $\text{Diff}^k(\mathbb{R}^e)$ , the group of  $C^k$ -diffeomorphisms of  $\mathbb{R}^e$ . Recall that the map  $U_{T \leftarrow 0}^{\mathbf{x}} : WG\Omega_p(\mathbb{R}^d) \mapsto \text{Diff}^k(\mathbb{R}^e)$  is a continuous function on  $WG\Omega_p(\mathbb{R}^d)$  when  $\text{Diff}^k(\mathbb{R}^e)$  is equipped with the  $C^k$ -topology ([18, Theorem 11.12]). The following result is now an immediate corollary of Theorem 5.5.

**Corollary 5.7** *Suppose the assumptions of Theorem 5.5 are verified for some  $1 \leq p < N + 1$ . Let  $\gamma > p$ ,  $k \geq 1$  an integer, and  $(f_i)_{i=1}^d$  a collection of vector fields in  $\text{Lip}^{\gamma+k-1}(\mathbb{R}^e)$ . Let  $U_{1 \leftarrow 0}^{\mathbf{x}} : WG\Omega_p(\mathbb{R}^d) \mapsto \text{Diff}^k(\mathbb{R}^e)$  be the associated flow map.*

*Then  $U_{1 \leftarrow 0}^{\mathbf{X}^{n,\phi}} \xrightarrow{\mathcal{D}} U_{1 \leftarrow 0}^{\mathbf{X}^\phi}$  as  $\text{Diff}^k(\mathbb{R}^e)$ -valued random variables.*

We demonstrate how one can apply Corollary 5.7 to show weak convergence of stochastic flows in the following three examples, the first of which extends a result of Kunita [29].

**Example 5.8** (Linear interpolation, Kunita [29]). Let  $Y_{n1}, \dots, Y_{nn}$  be an iid array in  $\mathbb{R}^d$  such that the associated random walk  $\mathbf{Y}^n$  converges in law as a  $D_o([0, 1], \mathbb{R}^d)$ -valued random variable to a Lévy process  $\mathbf{Y}$  in  $\mathbb{R}^d$ .

We claim that ODE flows driven by the piecewise linear interpolation of the random walk  $\mathbf{Y}^n$  along  $\text{Lip}^{\gamma+k-1}$  vector fields, for any  $\gamma > 2$ ,  $k \geq 1$ , converge in law as  $\text{Diff}^k(\mathbb{R}^e)$ -valued random variables.

Indeed, setting  $G := G^2(\mathbb{R}^d)$ , consider the  $G$ -valued iid array  $X_{nj} := e^{Y_{nj}}$ . It follows that  $X_{nj}$  is scaled by any scaling function  $\theta$  on  $G$  for which  $\theta \geq \sum_{i=1}^d |\xi_i|^2$ . Moreover, using the fact that  $\xi_i \circ \exp \in C_c^\infty(\mathbb{R}^d)$ , one can readily see by Theorem 2.1 that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D_o([0, 1], G)$ -valued random variables, where  $\mathbf{X}$  is a  $G$ -valued Lévy process. Finally, consider the 1-approximating, endpoint continuous path function

$$\phi : \exp(\mathbb{R}^d) \mapsto C_o([0, 1], G), \quad \phi(e^x)_t = e^{tx}.$$

Then  $\mathbf{X}^{n,\phi}$  is (a reparametrisation of) the lift of the piecewise linear interpolation of  $\mathbf{Y}^n$ . Furthermore, the conditions of Theorem 5.5 are satisfied for all  $p > 2$ , so that  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C_o^{0,p\text{-var}}([0, 1], G)$ -valued random variables, from which the desired claim follows (Corollary 5.7).

*Remark 5.9* In the previous example, it is easy to see that RDEs driven by  $\mathbf{X}^\phi$  coincide (up to reparametrisation) with general (Marcus) RDEs driven by  $\mathbf{X}$  (in the sense of [17, Section 6]) and thus with Marcus SDEs driven by  $\mathbf{Y}$ .

*Remark 5.10* The previous example extends the main result of Kunita [29, Theorem 4, Corollary p. 329]. The main restriction of Kunita's result is the assumption that the vector fields  $f_1, \dots, f_d$ , along which  $\mathbf{Y}^n$  drives an ODE, generate a finite dimensional Lie algebra, which essentially allows one to reduce the problem to a random walk on a Lie group (see [29, p. 340]). Our approach, based on convergence under rough path topologies, bypasses this restriction and provides a natural interpretation of the limiting stochastic flow as the solution of an RDE.

*Remark 5.11* Breuillard, Friz and Huesmann [7] showed a result analogous to the above example in a special case where the limiting Lévy process  $\mathbf{Y}$  is Brownian motion. The main analytic tool used in [7] is the Kolmogorov–Lamperti criterion to show tightness of  $(\|\mathbf{Y}^n\|_{1/p\text{-Hölder};[0,1]})_{n \geq 1}$ . This is of course stronger than tightness of  $(\|\mathbf{Y}^n\|_{p\text{-var};[0,1]})_{n \geq 1}$ , and cannot hold whenever the limiting Lévy process has jumps, which demonstrates an example where the tightness criterion Theorem 4.8 can be used as an effective alternative to the classical Kolmogorov–Lamperti criterion.

In the following example we demonstrate how Example 5.8 generalises to non-linear interpolations with essentially no extra effort.

*Example 5.12* (Non-linear interpolation). As in Example 5.8, let  $Y_{nj}$  be an iid array in  $\mathbb{R}^d$  such that  $\mathbf{Y}^n \xrightarrow{\mathcal{D}} \mathbf{Y}$  for a Lévy process  $\mathbf{Y}$  in  $\mathbb{R}^d$ .

Instead of piecewise interpolations, consider now any  $q$ -approximating endpoint continuous path function  $\psi : \mathbb{R}^d \mapsto C_o^{q\text{-var}}([0, 1], \mathbb{R}^d)$  for some  $1 \leq q < 2$ . Set again  $G = G^2(\mathbb{R}^d)$  and define the injective map  $f : \mathbb{R}^d \mapsto G$  by

$$f(x) = S_2(\psi(x))_1,$$

where  $S_2 : C_o^{q\text{-var}}([0, 1], \mathbb{R}^d) \mapsto C_o^{q\text{-var}}([0, 1], G)$  denotes the level-2 lifting map.

Consider the iid array  $X_{nj} := f(Y_{nj})$ . It follows readily from the assumption that  $\psi$  is  $q$ -approximating that  $X_{nj}$  is again scaled by any scaling function  $\theta$  on  $G$  for which

$\theta \geq \sum_{i=1}^d |\xi_i|^2$ . We now make the assumption on  $\psi$  and  $Y_{n1}$  that for all  $i, j \in \{1, \dots, m\}$  the following limits exist:

$$D^i := \lim_{n \rightarrow \infty} n \mathbb{E} [\xi_i(f(Y_{n1}))],$$

$$C^{i,j} := \lim_{n \rightarrow \infty} n \mathbb{E} [\xi_i(f(Y_{n1})) \xi_j(f(Y_{n1}))] - \int_{\mathbb{R}^d} \xi_i(f(x)) \xi_j(f(x)) \Pi(dx).$$

This occurs, for example, whenever every  $\xi_i \circ f$  is twice differentiable at zero, but in general will depend on the array  $Y_{nj}$  and the path function  $\psi$ .

Under this assumption, it follows from Theorem 2.1 that the random walk  $\mathbf{X}^n$  associated with the array  $X_{nj}$  converges in law to the Lévy process  $\mathbf{X}$  with triplet  $(C, D, \Xi)$ , where  $\Xi$  is the pushforward of  $\Pi$  by  $f$ .

Define now the  $q$ -approximating, endpoint continuous path function  $\phi : f(\mathbb{R}^d) \mapsto C_o^{q\text{-var}}([0, 1], G)$  by

$$\phi(f(x)) = S_2(\psi(x)).$$

Observe that the conditions of Theorem 5.5 are again satisfied for all  $p > 2$ , so that  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C_o^{0,p\text{-var}}([0, 1], G)$ -valued random variables.

Note that  $\mathbf{X}^{n,\phi}$  is, up to reparametrisation, the lift of  $\mathbf{Y}^{n,\psi}$  (which is itself, up to reparametrisation, the random walk  $\mathbf{Y}^n$  interpolated by the path function  $\psi$ ). It follows that ODE flows driven by  $\mathbf{Y}^{n,\psi}$  along  $\text{Lip}^{\gamma+k-1}$  vector fields, for any  $\gamma > 2, k \geq 1$ , converge in law as  $\text{Diff}^k$ -valued r.v.'s to the corresponding RDE flow driven by  $\mathbf{X}^\phi$  (Corollary 5.7).

*Remark 5.13* McShane [33] considered non-linear interpolations of the increments of Brownian motion and showed strong convergence of the corresponding ODEs to the associated Stratonovich SDE with an adjusted drift. We note that the family of path functions  $\psi$  to which the above example applies includes the non-linear interpolations considered by McShane ([33, p. 285]) (provided that  $Y_{nk}$  are also sufficiently well behaved, e.g., the increments of Brownian motion, to ensure that the limits  $C^{i,j}$  and  $D^i$  exist). The above example can thus be seen as a weak convergence analogue for general Lévy processes of the results in [33]. In a similar way, the following example is analogous to the results of Sussman [35] on non-linear approximations of Brownian motion.

*Example 5.14* (Perturbed walk). As in Examples 5.8 and 5.12, let  $Y_{nj}$  be an iid array in  $\mathbb{R}^d$  such that  $\mathbf{Y}^n \xrightarrow{\mathcal{D}} \mathbf{Y}$  for a Lévy process  $\mathbf{Y}$  in  $\mathbb{R}^d$ .

Let  $N \geq 2$  and as before denote  $G = G^N(\mathbb{R}^d)$  and  $\mathfrak{g} = \mathfrak{g}^N(\mathbb{R}^d)$ . Fix a path  $\gamma \in C_o^{1\text{-var}}([0, 1], \mathbb{R}^d)$  such that  $v := \log(S_N(\gamma)_{0,1})$  is in the center of  $\mathfrak{g}$  (that is,  $v^i = 0$  for all  $i \in \{1, \dots, m\}$  such that  $d_i < N$ ).

In this example we wish to consider the random path  $\mathbf{Z}^n \in C^{1\text{-var}}([0, 1], \mathbb{R}^d)$  defined by linearly joining the points of  $\mathbf{Y}^n$ , and, between each linear chord, running along the path  $n^{-1/N}\gamma$ .

Define the closed subset

$$W := \left\{ \exp(y) \exp(\lambda v) \mid y \in \mathbb{R}^d, \lambda \geq 0 \right\} \subseteq G.$$

Note that for every  $x \in W$  decomposes uniquely as  $x = \exp(y) \exp(\lambda v)$  for some  $y \in \mathbb{R}^d$  and  $\lambda \geq 0$ . Define then the 1-approximating, endpoint continuous path function  $\phi : W \mapsto C_o^{1\text{-var}}([0, 1], G)$  by

$$\phi(\exp(y) \exp(\lambda v))_t = \begin{cases} \exp(2ty) & \text{if } t \in [0, 1/2] \\ \exp(y) S_N(\lambda^{1/N} \gamma)_{2t-1} & \text{if } t \in (1/2, 1]. \end{cases}$$

Consider the  $G$ -valued iid array  $X_{nj} := \exp(Y_{nj}) \exp(n^{-1}v)$  and the associated random walk  $\mathbf{X}^n$ . Observe that  $\mathbf{X}^{n,\phi}$  is (a reparametrisation of) the level- $N$  lift of the path  $\mathbf{Z}^n$  described above.

We now claim that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  for a Lévy process  $\mathbf{X}$  in  $G$ . A straightforward way to show this is to take local coordinates  $\sigma_1, \dots, \sigma_d \in C_c^\infty(\mathbb{R}^d)$  so that  $\sigma := \sum_{i=1}^d \sigma_i u_i$  is the identity in a neighbourhood of zero, and write the triplet of  $\mathbf{Y}$  as  $(A, B, \Pi)$  with respect to  $\sigma_1, \dots, \sigma_d$ . Define the functions

$$f_n : \mathbb{R}^d \mapsto \mathfrak{g}, \quad f_n(y) = \xi(e^y e^{v/n}),$$

so that  $\xi(X_{nk}) = f_n(Y_{nk})$ . Note that, since  $v$  is in the centre of  $\mathfrak{g}$ , there exists a neighbourhood of zero  $V \subset \mathbb{R}^d$  and  $n_0 > 0$  such that for all  $n \geq n_0$

$$f_n(y) = \sum_{i=1}^d \sigma_i(y) u_i + \xi(e^{v/n}) + h_n(y),$$

where  $h_n \equiv 0$  on  $V$ . It readily follows that

$$\lim_{k \rightarrow \infty} \sup_{n \geq 1} k \left| \mathbb{E} [f_n(Y_{k1}) - \xi(e^{v/n})] - Q_n \right| = 0,$$

where

$$Q_n := \sum_{i=1}^d B^i u_i + \int_{\mathbb{R}^d} h_n(y) \Pi(dy) \in \mathfrak{g}.$$

Observe now that for all  $y \in \mathbb{R}^d$ ,  $\lim_{n \rightarrow \infty} h_n(y) = \xi(e^y) - \sigma(y)$ , so that by dominated convergence,

$$\lim_{n \rightarrow \infty} Q_n = \sum_{i=1}^d B^i u_i + \int_{\mathbb{R}^d} (\xi(e^y) - \sigma(y)) \Pi(dy) =: Q \in \mathfrak{g},$$

from which it follows that

$$\lim_{n \rightarrow \infty} n \mathbb{E} [f_n(Y_{n1}) - f_n(0)] = Q.$$

Since  $nf_n(0) = v$  for all  $n$  sufficiently large, we obtain that the following limit exists:

$$D := \lim_{n \rightarrow \infty} n \mathbb{E} [\xi(X_{n1})] \in \mathfrak{g}.$$

Furthermore, letting  $\Xi$  denote the pushforward of  $\Pi$  by  $\exp$ , one can show in exactly the same way that

$$C^{i,j} := \lim_{n \rightarrow \infty} n \mathbb{E} [\xi_i(X_{n1}) \xi_j(X_{n1})] - \int_G \xi_i(x) \xi_j(x) \Xi(dx)$$

exists for all  $i, j \in \{1, \dots, m\}$ , and that

$$\lim_{n \rightarrow \infty} n \mathbb{E} [f(X_{n1})] = \int_G f(x) \Xi(dx)$$

for every  $f \in C_b(G)$  which is identically zero on a neighbourhood of  $1_G$ . It follows by Theorem 2.1 that  $\mathbf{X}^n \xrightarrow{\mathcal{D}} \mathbf{X}$  as claimed, where  $\mathbf{X}$  is the Lévy process with triplet  $(C, D, \Xi)$ .

Finally, one readily sees that  $X_{nj}$  is scaled by any scaling function  $\theta$  on  $G$  for which

$$\theta \geq \sum_{i=1}^d |\xi_i|^2 + \sum_{\substack{1 \leq i \leq m \\ d_i = N}} |\xi_i|.$$

It now follows by Theorem 5.5 that for all  $p > N$ ,  $\mathbf{X}^{n,\phi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C_o^{0,p\text{-var}}([0, 1], G)$ -valued r.v.'s. In particular, ODE flows driven by the random paths  $\mathbf{Z}^n$  along  $\text{Lip}^{\gamma+k-1}$  vector fields, for any  $\gamma > N$ ,  $k \geq 1$ , converge in law as  $\text{Diff}^k$ -valued r.v.'s to the corresponding RDE flow driven by  $\mathbf{X}^\phi$  (Corollary 5.7).

*Remark 5.15* Note that the previous Example 5.8 is a special case of Example 5.14 by taking  $v = 0$  and  $\gamma$  the constant path  $\gamma \equiv 0$ . Building on Remark 5.9, one can verify that RDEs driven by  $\mathbf{X}^\phi$  coincide (up to reparametrisation) with the associated Marcus SDEs driven by  $\mathbf{Y}$  with an adjusted drift given by appropriate  $N$ -th level Lie brackets of the driving vector fields (cf. [16] and [18, Section 13.3.4]).

### 5.3.2 The Lévy–Khinchine formula for Lévy rough paths

In this subsection we determine a formula for the characteristic function (in the sense of [11]) of the signature of a Lévy rough path.

Recall that for every  $\mathbf{x} \in WG\Omega_p(\mathbb{R}^d)$ , there exists an element

$$S(\mathbf{x})_{0,T} = (1, S(\mathbf{x})_{0,T}^1, S(\mathbf{x})_{0,T}^2, \dots) \in T((\mathbb{R}^d)) = \prod_{k=0}^{\infty} (\mathbb{R}^d)^{\otimes k},$$

called the *signature* of  $\mathbf{x}$ , where  $S(\mathbf{x})_{0,T}^k$  encodes all the  $k$ -fold iterated integrals of  $\mathbf{x}$ . A fundamental result in rough paths theory is that  $S(\mathbf{x})_{0,T}$  belongs to a certain group  $G(\mathbb{R}^d)$  contained in the set of group-like elements of  $T((\mathbb{R}^d))$  (for the tensor Hopf algebra structure). Furthermore, for every linear map  $f \in \mathbf{L}(\mathbb{R}^d, \mathbf{L}(\mathbb{R}^e, \mathbb{R}^e))$ , the series  $\sum_{k=0}^{\infty} f^{\otimes k}(S(\mathbf{x})_{0,T}^k)$  converges absolutely to an operator  $f(S(\mathbf{x})_{0,T}) \in \mathbf{L}(\mathbb{R}^e, \mathbb{R}^e)$  (which is precisely the flow  $U_{T \leftarrow 0}^{\mathbf{x}}$  associated with the RDE (5.1) upon treating  $f$  as a collection of linear vector fields on  $\mathbb{R}^e$ ).

For a finite-dimensional complex Hilbert space  $H$ , let  $\mathfrak{u}(H) \subset \mathbf{L}(H, H)$  denote the Lie algebra of anti-Hermitian operators on  $H$  and  $\mathcal{U}(H) \subset \mathbf{L}(H, H)$  the group of unitary operators on  $H$ . Note that every  $f \in \mathbf{L}(\mathbb{R}^d, \mathfrak{u}(H))$  naturally induces a map  $f : G(\mathbb{R}^d) \mapsto \mathbf{L}(H, H)$  (which is continuous for the topology on  $G(\mathbb{R}^d)$  introduced in [11]) given by  $f(x) = \sum_{k=0}^{\infty} f^{\otimes k}(x^k)$ , where  $x^k \in (\mathbb{R}^d)^{\otimes k}$  denotes the level- $k$  projection of  $x$ . Note that  $f$  satisfies  $f(x) \in \mathcal{U}(H)$  and  $f(xy) = f(x)f(y)$  for all  $x, y \in G(\mathbb{R}^d)$ , i.e.,  $f : G(\mathbb{R}^d) \mapsto \mathcal{U}(H)$  is a unitary representation of  $G(\mathbb{R}^d)$ .

One of the main results of [11] is that for any  $WG\Omega_p(\mathbb{R}^d)$ -valued random variable  $\mathbf{X}$ , the following characteristic function

$$\begin{aligned} \mathbf{L}(\mathbb{R}^d, \mathfrak{u}(H)) &\mapsto \mathbf{L}(H, H), \\ f &\mapsto \mathbb{E} [f(S(\mathbf{X})_{0,T})], \end{aligned} \quad (5.2)$$

where  $H$  varies over all finite dimensional complex Hilbert spaces, uniquely determines  $S(\mathbf{X})_{0,T}$  as a  $G(\mathbb{R}^d)$ -valued random variable  $G(\mathbb{R}^d)$  (more generally, this result holds for every  $G(\mathbb{R}^d)$ -valued random variable).

*Remark 5.16* Boedihardjo et al. [5] have recently established a conjecture of Hambly–Lyons [24] on the kernel of the map  $S : WG\Omega_p(\mathbb{R}^d) \mapsto T((\mathbb{R}^d))$ . A consequence of the main result of [5] is that for all  $\mathbf{x}, \mathbf{y} \in WG\Omega_p(\mathbb{R}^d)$ ,  $S(\mathbf{x})_{0,T} = S(\mathbf{y})_{0,T} \Leftrightarrow U_{T \leftarrow 0}^{\mathbf{x}} = U_{T \leftarrow 0}^{\mathbf{y}}$  for all collections  $(f_i)_{i=1}^d$  of vector fields in  $\text{Lip}^\gamma(\mathbb{R}^e)$  with  $\gamma > p$  (not necessarily linear). In combination with the results from [11], it follows that for any  $WG\Omega_p(\mathbb{R}^d)$ -valued random variable  $\mathbf{X}$ , knowledge of the map (5.2) uniquely determines the law of every RDE driven by  $\mathbf{X}$ .

We now state the aforementioned formula for the characteristic function of the signature of a Lévy rough path. For a subset  $W \subseteq G$ , path function  $\phi : W \mapsto C_o^{p\text{-var}}([0, 1], G)$ , and a linear map  $f \in \mathbf{L}(\mathbb{R}^d, \mathbf{L}(\mathbb{R}^e, \mathbb{R}^e))$ , we adopt the shorthand notation

$$f_\phi : W \mapsto \mathbf{L}(\mathbb{R}^e, \mathbb{R}^e), \quad f_\phi(x) := f(S(\phi(x))_{0,1}).$$

By interpolation (Lemma 3.3), one can readily verify that  $f_\phi$  is continuous whenever  $\phi$  is  $p$ -approximating and endpoint continuous. Finally, we canonically treat  $\mathfrak{g} =$

$\mathfrak{g}^N(\mathbb{R}^d)$  as a subspace of the tensor algebra  $T(\mathbb{R}^d)$ , so that for any Lie algebra  $\mathfrak{h}$ , every  $f \in \mathbf{L}(\mathbb{R}^d, \mathfrak{h})$  extends uniquely to a linear map  $f : \mathfrak{g} \mapsto \mathfrak{h}$ .

**Theorem 5.17** (Lévy–Khintchine formula). *Let  $\mathbf{X}$  be a Lévy process in  $G$  with triplet  $(A, B, \Pi)$ . Suppose that for some  $1 \leq p < N + 1$ ,  $\|\mathbf{X}\|_{p\text{-var};[0,T]} < \infty$  a.s.. Let  $\phi : \text{supp}(\Pi) \mapsto C_o^{p\text{-var}}([0, 1], G)$  be a  $p$ -approximating, endpoint continuous path function defined on  $\text{supp}(\Pi)$ .*

*Then for every finite-dimensional complex Hilbert space  $H$  and  $f \in \mathbf{L}(\mathbb{R}^d, \mathfrak{u}(H))$ , it holds that the function*

$$f_\phi - \text{id}_H - \sum_{i=1}^m f(u_i)\xi_i : \text{supp}(\Pi) \mapsto \mathbf{L}(H, H)$$

*is  $\Pi$ -integrable, and that*

$$\mathbb{E}[f(S(\mathbf{X}^\phi)_{0,T})] = \exp(T\Psi_{\mathbf{X}}(f)), \quad (5.3)$$

*where*

$$\begin{aligned} \Psi_{\mathbf{X}}(f) := & \sum_{i=1}^m B^i f(u_i) + \frac{1}{2} \sum_{\substack{1 \leq i, j \leq m \\ d_i + d_j \leq N}} A^{i,j} f(u_i) f(u_j) \\ & + \int_G \left[ f_\phi(x) - \text{id}_H - \sum_{i=1}^m \xi_i(x) f(u_i) \right] \Pi(dx). \end{aligned}$$

**Remark 5.18** Note that every pair  $(\mathbf{X}, \phi)$  as in Theorem 5.17 naturally gives rise to a convolution semigroup  $(\mu_t)_{t>0}$  of probability measures on  $G(\mathbb{R}^d)$  (which we recall is a Polish but, if  $d > 1$ , non-locally compact group, [11]) given by  $\mu_t = \text{Law}[S(\mathbf{X}^\phi_{[0,t]})_{0,t}]$ , where  $\mathbf{X}^\phi_{[s,t]} \in C([s, t], G)$  denotes the connecting map applied to the restriction  $\mathbf{X}|_{[s,t]}$ . Moreover, treating  $\phi$  as a map  $\text{supp}(\Pi) \mapsto G(\mathbb{R}^d)$ ,  $x \mapsto S(\phi(x))_{0,1}$ , and every  $f \in \mathbf{L}(\mathbb{R}^d, \mathfrak{u}(H))$  as unitary representation of  $G(\mathbb{R}^d)$ , Theorem 5.17 bears a close resemblance to other forms of the Lévy–Khintchine formula stated in terms of unitary representations of Lie groups (see, e.g., [1, Section 5.5]).

**Remark 5.19** Theorem 5.17 can be seen as an extension of a related result on the expected signature of a Lévy  $p$ -rough path for  $1 \leq p < 3$  ([17, Theorem 53]) in which  $\phi$  is taken as the log-linear path function  $\phi(e^x) = e^{tx}$ ,  $\forall x \in \mathfrak{g}$ , and additional moment assumptions on the Lévy measure are required to ensure existence of the expected signature.

We first record the following estimate which is readily derived from standard Euler approximations to RDEs ([18, Corollary 10.15]).

**Lemma 5.20** *Let  $1 \leq p < \gamma < N + 1$ ,  $\theta$  a scaling function on  $G$ ,  $W \subseteq G$  a subset, and  $\phi$  a path function defined on  $W$  such that  $\lim_{x \rightarrow 1_G} \|\phi(x)\|_{p\text{-var};[0,1]}^\gamma / \theta(x) = 0$ .*

Then for all  $f \in \mathbf{L}(\mathbb{R}^d, \mathbf{L}(\mathbb{R}^e, \mathbb{R}^e))$ , it holds that

$$\lim_{x \rightarrow 1_G} \frac{1}{\theta(x)} \left| f_\phi(x) - \text{id}_{\mathbb{R}^e} - \sum_{i=1}^m \xi_i(x) f(u_i) - \frac{1}{2} \sum_{\substack{1 \leq i, j \leq m \\ d_i + d_j \leq N}} \xi_i(x) \xi_j(x) f(u_i) f(u_j) \right| = 0.$$

*Proof of Theorem 5.17* Without loss of generality, we can assume  $T = 1$ . Let  $V$  be a bounded neighbourhood of  $1_G$  and  $W := \text{supp}(\Pi) \cup V$ . Note that  $\phi$  shrinks on the diagonal (see Remark A.2), so we can find a path function  $\psi : W \mapsto C([0, 1], G)$  which is also  $p$ -approximating and shrinks on the diagonal and such that  $\psi \equiv \phi$  on  $\text{supp}(\Pi)$  (e.g., let  $\psi(x)$  be a geodesic from  $1_G$  to  $x$  for all  $x \in V \setminus \text{supp}(\Pi)$ ).

Let  $X_{n1}, \dots, X_{nn}$  be the iid array constructed in Sect. 2.3 associated to  $\mathbf{X}$ , and let  $\mathbf{X}^n$  be the associated random walk. Due to the shrinking support of the random variables  $Y_{nj}$  from Sect. 2.3, observe that

$$\text{for every } \varepsilon > 0, \mathbb{P}[\mathbf{X}^n \notin \text{supp}(\Pi)^\varepsilon] = 0 \text{ for all } n \text{ sufficiently large,} \quad (5.4)$$

where we recall the notation  $\text{supp}(\Pi)^\varepsilon$  from Section A.1. In particular, for all  $n$  sufficiently large,  $\mathbf{X}^n \in W^0$  a.s., so that  $\mathbf{X}^{n,\psi}$  is well-defined. Observe that, due to (5.4) and Proposition A.4,  $\mathbf{X}^{n,\psi} \xrightarrow{\mathcal{D}} \mathbf{X}^\psi$  as  $C_o([0, 1], G)$ -valued random variables.

Let  $p < p' < N + 1$ . Since  $\|\mathbf{X}\|_{p\text{-var};[0,1]} < \infty$  a.s. by assumption, we deduce from Theorem 5.1, Lemma 5.4, and Proposition 3.5, that

$$\mathbf{X}^{n,\psi} \xrightarrow{\mathcal{D}} \mathbf{X}^\psi \stackrel{\mathcal{D}}{=} \mathbf{X}^\phi \text{ as } C_o^{0,p'\text{-var}}([0, 1], G) \text{ - valued random variables,}$$

where the equality in law follows from the fact that  $\psi \equiv \phi$  on  $\text{supp}(\Pi)$  and  $\mathbf{X} \in \text{supp}(\Pi)^0$  a.s. (see Remark 5.6).

For all  $i \in \{1, \dots, m\}$ , define  $q_i := 2 \wedge (p/d_i)$ , and let  $\theta$  be a scaling function on  $G$  such that  $\theta \equiv \sum_{i=1}^m |\xi_i|^{q_i}$  in a neighbourhood of  $1_G$ . It follows from Lemma 2.5 and part (2) of Theorem 5.1 that  $\theta$  scales the array  $X_{nj}$ .

Since  $\psi$  is  $p$ -approximating, it holds that  $\lim_{x \rightarrow 1_G} \|\psi(x)\|_{p\text{-var};[0,1]}^\gamma / \theta(x) = 0$  for all  $\gamma > p$ . For  $f \in \mathbf{L}(\mathbb{R}^d, \mathbf{u}(H))$ , observe that  $f_\psi$  is a map from  $W$  to the unitary operators on  $H$  (thus bounded) and is continuous on  $\text{supp}(\Pi)$ . Since  $f_\psi \equiv f_\phi$  on  $\text{supp}(\Pi)$ , it now follows from Lemmas 5.20 and 2.6 that

$$\lim_{n \rightarrow \infty} n \mathbb{E}[f_\psi(X_{n1}) - \text{id}_H] = \Psi_{\mathbf{X}}(f),$$

and thus

$$\lim_{n \rightarrow \infty} \mathbb{E}[f_\psi(X_{n1})]^n = \exp(\Psi_{\mathbf{X}}(f)).$$

Since the array  $X_{nj}$  is iid, note that for all  $n \geq 1$

$$\mathbb{E}[f(S(\mathbf{X}^{n,\psi})_{0,1})] = \mathbb{E}[f_\psi(X_{n1})]^n.$$

Since  $\mathbf{X}^{n,\psi} \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $WG\Omega_{p'}(\mathbb{R}^d)$ -valued r.v.'s, and  $\mathbf{x} \mapsto f(S(\mathbf{x})_{0,1})$  is a continuous bounded function on  $WG\Omega_{p'}(\mathbb{R}^d)$ , we obtain (5.3).  $\square$

**Acknowledgements** Part of this work is contained in the author's D.Phil thesis written under the guidance of Prof. Terry Lyons, to whom the author is sincerely grateful. The author would also like to thank Atul Shekhar, Guy Flint, and Prof. Peter Friz for helpful conversations on this topic, and the anonymous referees whose suggestions helped to significantly improve this paper.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix A: Path functions

In this section we introduce and study the basic properties of path functions, which serve to systematically connect the left- and right-limits of a càdlàg path. Throughout the section, let  $(E, d)$  be a metric space and equip  $C([0, T], E)$  and  $D([0, T], E)$  with the uniform and the Skorokhod topology respectively.

**Definition A.1** For a subset  $J \subseteq E \times E$ , we call  $\phi : J \mapsto C([0, 1], E)$  a *path function* defined on  $J$  if

$$\phi(x, y)_0 = x \text{ and } \phi(x, y)_1 = y, \forall (x, y) \in J.$$

We say  $\phi$  is *endpoint continuous* if  $\phi$  is continuous and

$$\phi(x, x)_t = x, \forall (x, x) \in J, \forall t \in [0, 1].$$

For  $p \geq 1$ , we say  $\phi$  is *p-approximating* if for every  $r > 0$  there exists  $C > 0$  such that for all  $(x, y) \in J$  with  $d(x, y) < r$

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

When  $E$  is a Lie group, we say  $\phi$  is *left-invariant* if there exists a subset  $B \subseteq E$  such that  $J = \{(x, y) \mid x^{-1}y \in B\}$  and

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

Note that for a Lie group  $G$  and a subset  $B \subseteq G$ , there is a canonical bijection between functions  $\phi : B \mapsto C([0, 1], G)$ , for which

$$\phi(x)_0 = 1_G \text{ and } \phi(x)_1 = x, \forall x \in B, \quad (\text{A.1})$$

and left-invariant path functions defined on  $J := \{(x, y) \mid x^{-1}y \in B\}$ . Whenever we speak of a path function  $\phi$  defined on a subset  $B \subseteq G$ , we shall mean that  $\phi$  satisfies (A.1) and shall identify  $\phi$  with the corresponding left-invariant path function defined on  $\{(x, y) \mid x^{-1}y \in B\}$ .

### A.1. The connecting map on the Skorokhod space

For a path  $\mathbf{x} \in D([0, T], E)$  and a time  $t \in [0, T]$ , define  $\Delta \mathbf{x}_t := (\mathbf{x}_{t-}, \mathbf{x}_t) \in E \times E$  and  $\|\Delta \mathbf{x}_t\| := d(\mathbf{x}_{t-}, \mathbf{x}_t)$ . We call  $t$  a jump time of  $\mathbf{x}$  if  $\|\Delta \mathbf{x}_t\| > 0$ .

For a subset  $J \subseteq E \times E$  and  $\varepsilon \geq 0$  define the subset of càdlàg paths

$$J^\varepsilon := \{\mathbf{x} \in D([0, T], E) \mid \forall t \in [0, T], \|\Delta \mathbf{x}_t\| > \varepsilon \Rightarrow \Delta \mathbf{x}_t \in J\}.$$

In particular,  $\mathbf{x} \in J^0$  if and only if all the jumps of  $\mathbf{x}$  are in  $J$ . In the case that  $E$  is a Lie group and  $B \subseteq E$ , we set  $B^\varepsilon := J^\varepsilon$  where  $J := \{(x, y) \mid x^{-1}y \in B\}$ .

For a path function  $\phi : J \mapsto C([0, 1], E)$ , we now define a map  $\mathbf{x} \mapsto \mathbf{x}^\phi$  from  $J^0$  to  $C([0, T], E)$ . The construction is similar to the method considered in [17] and [36] of adding fictitious times over which to traverse the jumps.

Consider  $\mathbf{x} \in J^0$ . Let  $t_1, t_2, \dots$  be the jump times of  $\mathbf{x}$  ordered so that  $\|\Delta \mathbf{x}_{t_1}\| \geq \|\Delta \mathbf{x}_{t_2}\| \geq \dots$  with  $t_j < t_{j+1}$  if  $\|\Delta \mathbf{x}_{t_j}\| = \|\Delta \mathbf{x}_{t_{j+1}}\|$ . Let  $0 \leq m \leq \infty$  be the number of jumps of  $\mathbf{x}$ . We call the sequence  $(t_j)_{j=1}^m$  the *canonically ordered jump times* of  $\mathbf{x}$ .

We henceforth fix a strictly decreasing sequence  $(r_i)_{i=1}^\infty$  of positive real numbers such that  $\sum_{i=1}^\infty r_i < \infty$ . Define the sequence  $(n_k)_{k=0}^m$  by  $n_0 = 0$ , and for  $1 \leq k < m+1$  let  $n_k$  be the smallest integer such that  $n_k > n_{k-1}$  and  $r_{n_k} < \|\Delta \mathbf{x}_{t_k}\|$ .

Let  $r := \sum_{k=1}^m r_{n_k}$ . Define the strictly increasing (càdlàg) function

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

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  $r_{n_k}$ .

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

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

Denote by  $\tau_r$  the linear bijection from  $[0, T]$  to  $[0, T + r]$ . We finally define

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

and the associated time change

$$\tau_{\mathbf{x}} := \tau_r^{-1} \circ \tau : [0, T] \mapsto [0, T].$$

We call the map  $\mathbf{x} \mapsto \mathbf{x}^\phi$  from  $J^0$  to  $C([0, T], E)$  the *connecting map*.

### A.2. Measurability and continuity

The main result of this subsection is a continuity property of the connecting map, which we summarise in Proposition A.4.

For a subset  $J \subseteq E \times E$ , we say that a path function  $\phi : J \mapsto C([0, T], E)$  *shrinks on the diagonal* if for every bounded set  $B \subseteq J$  and  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for all  $(x, y) \in B$  with  $d(x, y) < \delta$

$$\sup_{t \in [0, 1]} d(\phi(x, y)_t, y) < \varepsilon.$$

**Remark A.2** Observe that every left-invariant, endpoint continuous path function defined on a subset of a Lie group shrinks on the diagonal.

**Lemma A.3** Consider  $J \subseteq E \times E$  and a path function  $\phi : J \mapsto C([0, 1], E)$  which shrinks on the diagonal. Suppose that  $\phi$  is endpoint continuous on a subset  $K \subseteq J$ .

Let  $\mathbf{x} \in K^0$  and a sequence  $\mathbf{x}(n) \in J^0$  such that  $\mathbf{x}(n) \rightarrow \mathbf{x}$  in the Skorokhod topology as  $n \rightarrow \infty$ , and such that for every  $\varepsilon > 0$ , there exists  $n_0 > 0$  such that  $\mathbf{x}(n) \in K^\varepsilon$  for all  $n \geq n_0$ .

Then  $\lim_{n \rightarrow \infty} d_{\infty; [0, T]}(\mathbf{x}(n)^\phi, \mathbf{x}^\phi) = 0$ .

*Proof* Let  $\varepsilon > 0$ . By uniform continuity of  $\mathbf{x}^\phi$ , there exists  $\eta > 0$  such that

$$\sup_{|s-t| < \eta} d(\mathbf{x}_t^\phi, \mathbf{x}_s^\phi) < \varepsilon.$$

Let  $t_1, t_2, \dots$  be the canonically ordered jump times of  $\mathbf{x}$ , and denote  $[s_i, u_i) := [\tau_{\mathbf{x}}(t_i-), \tau_{\mathbf{x}}(t_i))$ . For another element  $\mathbf{y} \in D([0, T], E)$ , let  $t'_1, t'_2, \dots$  be the jump times of  $\mathbf{y}$  and  $[s'_i, u'_i) := [\tau_{\mathbf{y}}(t'_i-), \tau_{\mathbf{y}}(t'_i))$ . Let  $(n_k)_{k=0}^m$  and  $(n'_k)_{k=1}^{m'}$  be the corresponding sequences for  $\mathbf{x}$  and  $\mathbf{y}$  respectively (where  $m$  and  $m'$  are the number of jumps of  $\mathbf{x}$  and  $\mathbf{y}$ ).

Denote by  $\Lambda^*$  the set of continuous, strictly increasing bijections  $\lambda : [0, T] \mapsto [0, T]$  and let  $\text{id} \in \Lambda^*$  be the identity map. Consider on  $D([0, T], E)$  the Skorokhod metric

$$\sigma(\mathbf{x}, \mathbf{y}) := \inf_{\lambda \in \Lambda^*} \max\{d_{\infty; [0, T]}(\lambda, \text{id}), d_{\infty; [0, T]}(\mathbf{x} \circ \lambda, \mathbf{y})\}.$$

Let  $\delta > 0$  (which we shall send to zero), and suppose that there exists  $\lambda \in \Lambda^*$  such that  $d_{\infty; [0, T]}(\mathbf{x} \circ \lambda, \mathbf{y}) < \delta$  and  $d_{\infty; [0, T]}(\lambda, \text{id}) < \delta$ .

Observe that there exists an integer  $k \geq 1$  such that  $\lambda(t'_i) = t_i$  for all  $i \in \{1, \dots, k\}$ , and, denoting by  $v_1 < \dots < v_k$  (resp.  $v'_1 < \dots < v'_k$ ) the same set of points as  $t_1, \dots, t_k$  (resp.  $t'_1, \dots, t'_k$ ) ordered monotonically, it holds that  $\lambda(t') \in [v_i, v_{i+1})$  for all  $t' \in [v'_i, v'_{i+1})$ . In particular, it holds that  $d(\mathbf{y}_{t'_i-}, \mathbf{x}_{t_i-}), d(\mathbf{y}_{t'_i}, \mathbf{x}_{t_i}) < \delta$  for all  $i \in \{1, \dots, k\}$ .

Moreover, by choosing  $\delta$  sufficiently small, we can assume that  $n_i = n'_i$  for all  $i \in \{1, \dots, k\}$  and that  $k$  is sufficiently large so that, by making  $\sum_{j=k+1}^{\infty} r_j$  sufficiently small, it holds that  $|\tau_{\mathbf{y}}(t') - \tau_{\mathbf{x}}(\lambda(t'))| < \eta$  for all  $t' \in [v'_i, v'_{i+1})$  (this is where we have used the condition  $r_{n_j} < \|\Delta \mathbf{x}_{t_j}\|$ ).

In particular, it holds that that for all  $t' \in [v'_i, v'_{i+1})$

$$\begin{aligned} d(\mathbf{y}_{\tau_{\mathbf{y}}(t')}, \mathbf{x}_{\tau_{\mathbf{y}}(t')}) &\leq d(\mathbf{y}_{\tau_{\mathbf{y}}(t')}, \mathbf{x}_{\tau_{\mathbf{x}}(\lambda(t'))}) + d(\mathbf{x}_{\tau_{\mathbf{x}}(\lambda(t'))}, \mathbf{x}_{\tau_{\mathbf{y}}(t')}) \\ &< d(\mathbf{y}_{t'}, \mathbf{x}_{\lambda(t')}) + \varepsilon \\ &< \delta + \varepsilon. \end{aligned} \quad (\text{A.2})$$

This covers all points not in an interval  $[s'_i, u'_i)$  (note that no assumptions on  $\phi$  were needed yet except that  $\phi(x, y)$  itself is a continuous path for each  $(x, y) \in J$ ).

Now we let  $\mathbf{y} = \mathbf{x}(n)$  for some  $n$ . We may choose  $n$  sufficiently large, such that  $\sigma(\mathbf{x}, \mathbf{y}) < \delta$  and such that  $\Delta \mathbf{y}_{t'_i} \in K$  for all  $i \in \{1, \dots, k\}$ .

Due to the continuity of  $\phi : K \mapsto C([0, 1], E)$  at  $\Delta \mathbf{x}_{t_i} \in K$ , it follows that for all  $w' \in [s'_i, u'_i)$  and  $i \in \{1, \dots, k\}$ , there exists  $w \in [s_i, u_i)$  such that  $|w' - w| < \eta$  and  $d(\mathbf{x}_w, \mathbf{y}_{w'}) < \varepsilon$ , so that

$$d(\mathbf{y}_{w'}, \mathbf{x}_{w'}) < d(\mathbf{y}_{w'}, \mathbf{x}_w) + d(\mathbf{x}_w, \mathbf{x}_{w'}) < 2\varepsilon.$$

Finally, since  $\phi$  shrinks on the diagonal, we may further decrease  $\delta$  if necessary so that for all  $w' \in [s'_j, u'_j)$  and  $j > k$ , it holds that  $|w' - u'_j| < \eta$  and  $d(\mathbf{y}_{w'}, \mathbf{y}_{u'_j}) < \varepsilon$ .

Now  $u'_j = \tau_{\mathbf{y}}(t')$  for some  $t' \in [0, T]$ , and thus, by (A.2),

$$d(\mathbf{y}_{u'_j}, \mathbf{x}_{u'_j}) < \delta + \varepsilon,$$

from which it follows that

$$d(\mathbf{y}_{w'}, \mathbf{x}_{w'}) \leq d(\mathbf{y}_{w'}, \mathbf{y}_{u'_j}) + d(\mathbf{y}_{u'_j}, \mathbf{x}_{u'_j}) + d(\mathbf{x}_{u'_j}, \mathbf{x}_{w'}) < \delta + 3\varepsilon.$$

□

In the following proposition, we equip all topological spaces with their respective Borel  $\sigma$ -algebras.

**Proposition A.4** *Suppose  $E$  is a Polish space,  $K \subseteq J \subseteq E \times E$  are measurable sets, and  $\phi : J \mapsto C([0, 1], E)$  is a measurable path function.*

- (i) *Then  $J^\varepsilon$  is a measurable subset of  $D([0, T], E)$  for all  $\varepsilon \geq 0$ , and the connecting map  $\cdot^\phi : J^0 \mapsto C([0, T], E)$  is measurable.*
- (ii) *Let  $\mathbf{X}$  be a  $D([0, T], E)$ -valued random variable such that  $\mathbf{X} \in K^0$  a.s.. Let  $(\mathbf{X}_n)_{n \geq 1}$  be a collection of  $D([0, T], E)$ -valued random variables such that  $\mathbf{X}_n \in J^0$  a.s. and  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}$  as  $D([0, T], E)$ -valued random variables. Suppose further that for every  $\varepsilon > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{X}_n \notin K^\varepsilon] = 0$ , and that  $\phi$  is endpoint continuous on  $K$  and shrinks on the diagonal in  $J$ .*

*Then  $\mathbf{X}_n \xrightarrow{\mathcal{D}} \mathbf{X}^\phi$  as  $C([0, T], E)$ -valued random variables.*

*Proof* (i) Measurability of  $J^\varepsilon$  is an easy consequence of the measurability of the finite-dimensional projections  $\mathbf{x} \mapsto \mathbf{x}_t$  on  $D([0, T], E)$  ([3, Theorem 12.5]). To show measurability of the connecting map, note that it suffices to show that for every  $s \in [0, T]$ , the map  $\mathbf{x} \mapsto \mathbf{x}^\phi(s)$  is measurable (cf. [3, Example 1.3]), which in turn follows easily from the construction of  $\mathbf{x}^\phi$ .

(ii) Note that the condition  $\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{X}_n \notin K^\varepsilon] = 0$  for all  $\varepsilon > 0$  implies that

$$\mathbf{Y}_n := (\mathbf{X}_n, \mathbf{1}\{\mathbf{X}_n \notin K^1\}, \mathbf{1}\{\mathbf{X}_n \notin K^{1/2}\}, \mathbf{1}\{\mathbf{X}_n \notin K^{1/4}\}, \dots)$$

is a sequence of  $D([0, T], E) \times \{0, 1\}^{\mathbb{N}}$ -valued random variables which converges in law to  $\mathbf{Y} := (\mathbf{X}, 0, 0, \dots)$ . The conclusion now follows from an application of the Skorokhod representation theorem [27, Theorem 3.30] and Lemma A.3.  $\square$

### A.3. $p$ -variation

The main result of this subsection is Proposition A.7, which shows that a  $p$ -approximating path function does not significantly increase the  $p$ -variation of a càdlàg path.

We first require the following lemma, whose proof was inspired by [17, Lemma 22].

**Lemma A.5** *Let  $(E, d)$  be a metric space and  $\mathbf{x} : [0, T] \mapsto E$  a function (not necessarily càdlàg). Let  $(I_n)_{n \geq 1}$  be a countable collection of disjoint open subintervals of  $[0, T]$  and set  $I := \cup_{n \geq 1} I_n$ . Define  $\mathbf{y} : [0, T] \mapsto E$  by  $\mathbf{y}_t = \mathbf{x}_t$  for  $t \in [0, T] \setminus I$ , and  $\mathbf{y}_t = \mathbf{x}_{c_n}$  for  $t \in (c_n, d_n) := I_n$ . Let  $p > 0$  and  $C = 1 + 2 \cdot 2^{(p-1) \vee 0} + 3^{(p-1) \vee 0}$ . Then*

$$\|\mathbf{x}\|_{p\text{-var}; [0, T]}^p \leq (C + 3^{(p-1) \vee 0}) \sum_{n=1}^{\infty} \|\mathbf{x}\|_{p\text{-var}; [c_n, d_n]}^p + C \|\mathbf{y}\|_{p\text{-var}; [0, T]}^p.$$

*Proof* Define the super-additive functions  $\omega_{\mathbf{x}}(s, t) = \|\mathbf{x}\|_{p\text{-var}; [s, t]}^p$  and  $\omega_{\mathbf{y}}(s, t) = \|\mathbf{y}\|_{p\text{-var}; [s, t]}^p$ . Let  $\mathcal{D} = (t_0, t_1, \dots, t_k)$  be a partition of  $[0, T]$ .

Denote by  $J_1 = (a_1, b_1), \dots, J_m = (a_m, b_m)$  those intervals  $I_n$  which contain some partition point  $t_j \in \mathcal{D}$ , ordered so that  $b_j < a_{j+1}$  for all  $j \in \{1, \dots, m-1\}$ . Call a *block* a consecutive run of partition points  $t_j, t_{j+1}, \dots, t_n$  either all in  $J_r$  for some  $r \in \{1, \dots, m\}$ , in which case we call it red, or all outside  $I$ , in which case we call it blue. We call a consecutive pair of partition points  $t_i, t_{i+1} \in \mathcal{D}$  which lie in different blocks either red-red, red-blue, or blue-red depending on their respective blocks (note there are no blue-blue pairs). For convenience of notation, set  $J_0 = (a_0, b_0) := (-\infty, 0)$  and  $J_{m+1} = (a_{m+1}, b_{m+1}) := (T, \infty)$ .

For  $r \in \{1, \dots, m\}$  and a red block  $t_j, t_{j+1}, \dots, t_n$  in  $J_r$  we have

$$\sum_{i=j+1}^n d(\mathbf{x}_{t_{i-1}}, \mathbf{x}_{t_i})^p \leq \omega_{\mathbf{x}}(a_r, b_r).$$

For  $r \in \{0, \dots, m\}$  and a blue block  $t_j, t_{j+1}, \dots, t_n$  between  $J_r, J_{r+1}$  we have

$$\sum_{i=j+1}^n d(\mathbf{x}_{t_{i-1}}, \mathbf{x}_{t_i})^p = \sum_{i=j+1}^n d(\mathbf{y}_{t_{i-1}}, \mathbf{y}_{t_i})^p \leq \omega_{\mathbf{y}}(b_r, a_{r+1}).$$

For  $r \in \{1, \dots, m\}$  and a red-blue pair  $t_i, t_{i+1}$  with  $t_i \in J_r$  and  $t_{i+1}$  between  $J_r, J_{r+1}$ , we have

$$\begin{aligned} d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p &\leq 2^{(p-1) \vee 0} [d(\mathbf{x}_{t_i}, \mathbf{x}_{b_r})^p + d(\mathbf{x}_{b_r}, \mathbf{x}_{t_{i+1}})^p] \\ &\leq 2^{(p-1) \vee 0} [\omega_{\mathbf{x}}(a_r, b_r) + \omega_{\mathbf{y}}(b_r, t_{i+1})]. \end{aligned}$$

For  $r \in \{0, \dots, m-1\}$  and a blue-red pair  $t_i, t_{i+1}$  with  $t_i$  between  $J_r, J_{r+1}$  and  $t_{i+1} \in J_{r+1}$ , we have

$$\begin{aligned} d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p &\leq 2^{(p-1) \vee 0} [d(\mathbf{x}_{t_i}, \mathbf{x}_{a_{r+1}})^p + d(\mathbf{x}_{a_{r+1}}, \mathbf{x}_{t_{i+1}})^p] \\ &\leq 2^{(p-1) \vee 0} [\omega_{\mathbf{y}}(t_i, a_{r+1}) + \omega_{\mathbf{x}}(a_{r+1}, b_{r+1})]. \end{aligned}$$

Finally for  $r \in \{1, \dots, m-1\}$  a red-red pair  $t_i, t_{i+1}$  with  $t_i \in J_r$  and  $t_{i+1} \in J_{r+1}$ , we have

$$\begin{aligned} d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p &\leq 3^{(p-1) \vee 0} [d(\mathbf{x}_{t_i}, \mathbf{x}_{b_r})^p + d(\mathbf{x}_{b_r}, \mathbf{x}_{a_{r+1}})^p + d(\mathbf{x}_{a_{r+1}}, \mathbf{x}_{t_{i+1}})^p] \\ &\leq 3^{(p-1) \vee 0} [\omega_{\mathbf{x}}(a_r, b_r) + \omega_{\mathbf{y}}(b_r, a_{r+1}) + \omega_{\mathbf{x}}(a_{r+1}, b_{r+1})]. \end{aligned}$$

Since  $\omega_{\mathbf{x}}$  and  $\omega_{\mathbf{y}}$  are super-additive, the conclusion now follows from splitting the sum  $\sum_{i=1}^k d(\mathbf{x}_{t_{i-1}}, \mathbf{x}_{t_i})^p$  into blocks and consecutive pairs in different blocks.  $\square$

**Corollary A.6** Let  $p \geq 1$ ,  $J \subseteq E \times E$ ,  $\phi : J \mapsto C([0, T], E)$  a path function, and  $\mathbf{z} \in J^0$ . Suppose that there exists  $C > 0$  such that  $\|\phi(\mathbf{z}_{t-}, \mathbf{z}_t)\|_{p\text{-var}; [0, 1]} \leq Cd(\mathbf{z}_{t-}, \mathbf{z}_t)$  for every jump time  $t$  of  $\mathbf{z}$ . Then

$$\|\mathbf{z}^\phi\|_{p\text{-var}; [0, T]}^p \leq \left(1 + 2^p + 3^{p-1} + C^p(1 + 2^p + 2 \cdot 3^{p-1})\right) \|\mathbf{z}\|_{p\text{-var}; [0, T]}^p.$$

The following result now follows immediately from Corollary A.6 and from the definition of a  $p$ -approximating path function (Definition A.1).

**Proposition A.7** Let  $p \geq 1$ ,  $J \subseteq E \times E$ , and  $\phi : J \mapsto C([0, T], E)$  a  $p$ -approximating path function. There exists a continuous function  $\psi : [0, \infty) \mapsto [0, \infty)$  such that for some  $R, \varepsilon > 0$ ,  $\psi(x) = Rx$  for all  $x \in [0, \varepsilon)$ , and such that  $\|\mathbf{x}^\phi\|_{p\text{-var}; [0, T]} \leq \psi(\|\mathbf{x}\|_{p\text{-var}; [0, T]})$  for all  $\mathbf{x} \in J^0$ .

## Appendix B: Infinite $p$ -variation of Lévy processes in Lie groups

The purpose of this section is to establish conditions under which sample paths of a Lévy process have *infinite*  $p$ -variation. The methods are all well-known for the case

$G = \mathbb{R}^d$ , so we mostly provide indications of how they extend to a general Lie group. Throughout this section, we use the notation from Section 2.

Let  $\mathbf{X}$  be a Lévy process in a Lie group  $G$  with triplet  $(A, B, \Pi)$  and let  $\tau = \inf \{t \geq 0 \mid \mathbf{X}_t \notin U\}$  be the first exit time of  $\mathbf{X}$  from  $U$ . Let  $y_0 \in \mathfrak{g} \setminus \log(U)$  be a distinguished point and consider the  $\mathfrak{g}$ -valued process

$$\mathbf{Y}_t = \begin{cases} \log(\mathbf{X}_t) & \text{if } t \in [0, \tau) \\ \mathbf{Y}_t = y_0 & \text{if } t \geq \tau. \end{cases}$$

For  $\varepsilon > 0$ , define the  $\mathfrak{g}$ -valued process  $\mathbf{Y}_t^\varepsilon := \varepsilon^{-1/2}(\mathbf{Y}_{\varepsilon t})$  for  $t \in [0, 1]$ . Let  $\mathbf{B}$  be a  $\mathfrak{g}$ -valued centred Brownian motion starting from zero with covariance matrix  $(A^{i,j})_{i,j=1}^m$  with respect to the basis  $u_1, \dots, u_m$ .

**Lemma B.1** *It holds that  $\mathbf{Y}^\varepsilon \xrightarrow[\varepsilon \rightarrow 0]{\mathcal{D}} \mathbf{B}$  as  $D_o([0, 1], \mathfrak{g})$ -valued random variables.*

*Proof* Note that for every  $\varepsilon > 0$ ,  $\mathbf{Y}^\varepsilon$  can be considered as a  $\mathfrak{g}$ -valued Markov process (for which every point outside  $\varepsilon^{-1/2} \log(U)$  is absorbing). Writing  $L^\varepsilon$  and  $L_{\mathbf{B}}$  for the generators of  $\mathbf{Y}^\varepsilon$  and  $\mathbf{B}$  respectively, it suffices to show that  $L^\varepsilon f \rightarrow L_{\mathbf{B}} f$  in  $C_0(\mathfrak{g})$  for all  $f \in C_c^\infty(\mathfrak{g})$  (see, e.g., [27, Chapter 17]). This in turn follows from writing the generator of  $\mathbf{X}$  in the log chart and performing a straightforward limiting argument.  $\square$

**Proposition B.2** *Suppose  $A^{i,i} > 0$  for some  $i \in \{1, \dots, m\}$ . Then*

$$\mathbb{P} \left[ \limsup_{t \rightarrow 0} t^{-1/2} |\xi_i(\mathbf{X}_t)| = \infty \right] = 1.$$

*Proof* Let  $c > 0$ . Lemma B.1 implies that there exist  $\delta, \varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\mathbb{P} \left[ \varepsilon^{-1/2} |\xi_i(\mathbf{X}_\varepsilon)| > c \right] > \delta.$$

Observe that, by the CBH formula, there exist a neighbourhood  $V$  of  $1_G$  and a constant  $C > 0$  such that for all  $x, y \in V$

$$|\xi_i(x^{-1}y)| \leq |\xi_i(y) - \xi_i(x)| + C|\xi(x)|^2 + C|\xi(y)|^2. \quad (\text{B.1})$$

For  $k \geq 1$ , define  $\varepsilon_k := 3^{-k} \varepsilon_0$ . Then by Lemma B.1 and the first Borel–Āř Cantelli lemma there exists a strictly increasing sequence  $(k(n))_{n \geq 1}$  such that

$$\mathbb{P} \left[ \limsup_{n \rightarrow \infty} \varepsilon_{k(n)}^{-1/2} C \left( |\xi(\mathbf{X}_{\varepsilon_{k(n)}})|^2 + |\xi(\mathbf{X}_{\varepsilon_{k(n)-1}})|^2 \right) > c/2 \right] = 0. \quad (\text{B.2})$$

On the other hand, since  $\mathbf{X}_{\varepsilon_k, \varepsilon_{k-1}} \stackrel{\mathcal{D}}{=} \mathbf{X}_{\varepsilon_{k-1} - \varepsilon_k}$  and the r.v.'s  $(\mathbf{X}_{\varepsilon_k, \varepsilon_{k-1}})_{k \geq 1}$  are independent, the second Borel–Āř Cantelli lemma yields

$$\mathbb{P} \left[ \limsup_{n \rightarrow \infty} \varepsilon_{k(n)}^{-1/2} |\xi_i(\mathbf{X}_{\varepsilon_{k(n)}, \varepsilon_{k(n)-1}})| > c \right] = 1.$$

Since  $\lim_{t \rightarrow 0} \mathbf{X}_t = 1_G$  a.s., we now readily deduce from (B.1), (B.2), and the definition of  $\varepsilon_k$  that

$$\mathbb{P} \left[ \limsup_{k \rightarrow \infty} \varepsilon_k^{-1/2} |\xi_i(\mathbf{X}_{\varepsilon_k})| > \frac{c}{4\sqrt{3}} \right] = 1.$$

As  $c > 0$  was arbitrary, the conclusion follows.  $\square$

**Corollary B.3** Suppose  $A^{i,i} > 0$  for some  $i \in \{1, \dots, m\}$ . Then

$$\mathbb{P} \left[ \sup_{\mathcal{D} \subset [0,1]} \sum_{t_k \in \mathcal{D}} |\xi_i(\mathbf{X}_{t_k, t_{k+1}})|^2 = \infty \right] = 1.$$

*Proof* By Proposition B.2,  $\limsup_{t \rightarrow 0} t^{-1} |\xi_i(\mathbf{X}_t)|^2 = \infty$  a.s.. Since  $\mathbf{X}$  has stationary and independent increments, the conclusion follows from an application of the Vitali covering argument (see [6, Proposition p. 68], or [18, Theorem 13.69]).  $\square$

The following is a form of the classical Blumenthal–Gettoor index [4] adapted to the setting of Lie groups. Recall the definitions of  $\Gamma_i$  and  $K$  from Section 2.3.

**Proposition B.4** (Blumenthal–Gettoor index). Let  $i \in \{1, \dots, m\}$  and  $q > 0$ . Then

$$\mathbb{P} \left[ \sup_{\mathcal{D} \subset [0,1]} \sum_{t_k \in \mathcal{D}} |\xi_i(\mathbf{X}_{t_k, t_{k+1}})|^q = \infty \right] = 1 \quad (\text{B.3})$$

if either

- (i)  $q \in \Gamma_i$ , or
- (ii)  $i \in K$  and  $q < 1$ .

*Proof* Define  $f \in C_c(G)$  by  $f(x) = 1 - \exp(-|\xi_i(x)|^q)$ . Since  $\mathbf{X}$  has independent and stationary increments, we can readily show (cf. [4, p. 499]) that (B.3) holds whenever

$$\lim_{t \rightarrow 0} t^{-1} \mathbb{E}[f(\mathbf{X}_t)] = \infty. \quad (\text{B.4})$$

It thus suffices to show that (B.4) holds in both cases of (i) and (ii):

- (i) Let  $(\psi_n)_{n \geq 1}$  be a non-decreasing sequence of non-negative functions in  $C^\infty(\mathbb{R})$ , each vanishing on some neighbourhood of zero, and such that  $\lim_{n \rightarrow \infty} \psi_n(x) = |x|^q$  for all  $x \in \mathbb{R}$ . Then for  $f_n(x) := 1 - \exp(-\psi_n(\xi_i(x)))$ , we have

$$\lim_{t \rightarrow 0} t^{-1} \mathbb{E}[f_n(\mathbf{X}_t)] = \int_G f_n(x) \Pi(dx) \geq c \int_G \psi_n(\xi_i(x)) \Pi(dx) \xrightarrow{n \rightarrow \infty} \infty,$$

where the final convergence follows from  $q \in \Gamma_i$ . Since  $0 \leq f_n \leq f$ , we obtain (B.4).

- (ii) Since  $q < 1$ , for every integer  $n \geq 1$  we can find  $\psi_n \in C^\infty(\mathbb{R})$  such that  $|\psi_n(x)| \leq |x|^q$  for all  $x \in \mathbb{R}$  and such that  $\psi_n(x) = nx/\tilde{B}^i$  for all  $x$  in a neighbourhood  $V_n$  of zero. Note that we may suppose  $A^{i,i} = 0$  and  $q \notin \Gamma_i$  (as otherwise the desired result follows by Corollary B.3 or by case (i)). Then for  $f_n(x) := 1 - \exp(-\psi_n(\xi_i(x)))$ , a straightforward calculation shows that

$$\lim_{t \rightarrow 0} t^{-1} \mathbb{E}[f_n(\mathbf{X}_t)] = n \left( 1 + n^{-1} \int_G f_n(x) \Pi(dx) \right) \xrightarrow{n \rightarrow \infty} \infty,$$

where the final convergence follows from  $q \notin \Gamma_i$  and  $|f_n(x)| \leq C|\psi_n(\xi_i(x))|$ . Since again  $f_n \leq f$ , we obtain (B.4). □

## References

1. Applebaum, D.: Probability on compact Lie groups. Probability Theory and Stochastic Modelling. With a Foreword by Herbert Heyer, vol. 70, pp. xxvi+217. Springer, Cham. ISBN: 978-3-319-07841-0; 978-3-319-07842-7. doi:[10.1007/978-3-319-07842-7](https://doi.org/10.1007/978-3-319-07842-7) (2014)
2. Baudoin, F.: An Introduction to the Geometry of Stochastic Flows. Imperial College Press, London, pp. x+140. ISBN: 1-86094-481-7. doi:[10.1142/9781860947261](https://doi.org/10.1142/9781860947261) (2004)
3. Billingsley, P.: Convergence of probability measures. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication, pp. x+277. Wiley, Inc., New York, . ISBN: 0-471-19745-9. doi:[10.1002/9780470316962](https://doi.org/10.1002/9780470316962) (1999)
4. Blumenthal, R.M., Getto, R.K.: Sample functions of stochastic processes with stationary independent increments. J. Math. Mech. **10**, 493–516 (1961)
5. Boedihardjo, H., Geng, Xi., Lyons, T., Yang, D.: The signature of a rough path: uniqueness. Adv. Math. **293** 720–737. ISSN: 0001- 8708. doi:[10.1016/j.aim.2016.02.011](https://doi.org/10.1016/j.aim.2016.02.011) (2016)
6. Bretagnolle, J.:  $p$ -variation de fonctions aléatoires. II. Processus à accroissements indépendants”. Séminaire de Probabilités, VI (Univ. Strasbourg, année universitaire 1970–1971). Lecture Notes in Mathematics. Springer, Berlin, vol. 258, pp. 64–71 (1972)
7. Breuillard, E., Friz, P., Huesmann, M.: From random walks to rough paths. Proc. Am. Math. Soc. **137**(10), 3487–3496. ISSN: 0002-9939. doi:[10.1090/S0002-9939-09-09930-4](https://doi.org/10.1090/S0002-9939-09-09930-4) (2009)
8. Bruned, Y., Chevyrev, I., Friz, P.K., Preiss, R.: A Rough Path Perspective on Renormalization. ArXiv e-prints. [arXiv:1701.01152](https://arxiv.org/abs/1701.01152) [math.PR] (2017)
9. Bruned, Y., Hairer, M., Zambotti, L.: Algebraic renormalisation of regularity structures. ArXiv e-prints. [arXiv:1610.08468](https://arxiv.org/abs/1610.08468) [math.RA] (2016)
10. Cass, T., Ogorodnik, M.: Tail estimates for Markovian rough paths. ArXiv e-prints. To appear in Annals of Probability. [arXiv:1411.5189](https://arxiv.org/abs/1411.5189) [math.PR] (2014)
11. Chevyrev, I., Lyons, T.: Characteristic functions of measures on geometric rough paths. Ann. Probab. **44**(6), 4049–4082. ISSN: 0091-1798. doi:[10.1214/15-AOP1068](https://doi.org/10.1214/15-AOP1068) (2016)
12. Chevyrev, I., Ogorodnik, M.: A support and density theorem for Markovian rough paths. ArXiv e-prints. [arXiv:1701.03002](https://arxiv.org/abs/1701.03002) [math.PR] (2017)
13. Feinsilver, P.: Processes with independent increments on a Lie group. Trans. Am. Math. Soc. **242**, 73–121. ISSN: 0002-9947 (1978)
14. Flint, G., Hambly, B., Lyons, T.: Discretely sampled signals and the rough Hoff process. Stochastic Process. Appl. **126**(9), 2593–2614. ISSN: 0304-4149. doi:[10.1016/j.spa.2016.02.011](https://doi.org/10.1016/j.spa.2016.02.011) (2016)
15. Folland, G.B., Stein, E.M.: Hardy spaces on homogeneous groups. Mathematical Notes, vol. 28, pp. xii+285. Princeton University Press, Princeton, NJ; University of Tokyo Press, Tokyo. ISBN: 0-691-08310-X (1982)
16. Friz, P., Oberhauser, H.: Rough path limits of the Wong–Zakai type with a modified drift term. J. Funct. Anal. **256**(10), 3236–3256. ISSN: 0022-1236. doi:[10.1016/j.jfa.2009.02.010](https://doi.org/10.1016/j.jfa.2009.02.010) (2009)
17. Friz, P., Shekhar, A.: General rough integration, Levy rough paths and a Levy–Kintchine type formula. ArXiv e-prints. To appear in Annals of Probability. [arXiv:1212.5888](https://arxiv.org/abs/1212.5888) [math.PR] (2012)

18. Friz, P.K., Victoir, N.B.: Multidimensional stochastic processes as rough paths. In: Cambridge Studies in Advanced Mathematics, vol. 120. pp. xiv+656. Cambridge University Press, Cambridge. ISBN: 978-0-521- 87607-0 (2010)
19. Friz, P., Victoir, N.: On uniformly subelliptic operators and stochastic area. *Probab. Theory Relat. Fields* **142**(3–4), 475–523. ISSN: 0178-8051. doi:[10.1007/s00440-007-0113-y](https://doi.org/10.1007/s00440-007-0113-y) (2008)
20. Gihman, I.I., Skorohod, A.V.: The theory of stochastic processes. I. Translated from the Russian by S. Kotz, Die Grundlehren der mathematischen Wissenschaften, Band 210, pp. viii+570. Springer, New York (1974)
21. Gubinelli, M.: Ramification of rough paths. *J. Differ. Equ.* **248**(4), 693–721. ISSN: 0022-0396. doi:[10.1016/j.jde.2009.11.015](https://doi.org/10.1016/j.jde.2009.11.015) (2010)
22. Hairer, M., Kelly, D.: Geometric versus non-geometric rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **51**(1), 207–251. ISSN: 0246- 0203. doi:[10.1214/13-AIHP564](https://doi.org/10.1214/13-AIHP564) (2015)
23. Hairer, M.: A theory of regularity structures. *Invent. Math.* **198**(2), 269–504. ISSN: 0020-9910. doi:[10.1007/s00222-014-0505-4](https://doi.org/10.1007/s00222-014-0505-4) (2014)
24. Hambly, B., Lyons, T.: Uniqueness for the signature of a path of bounded variation and the reduced path group. *Ann. Math.* **171**(1), 109–167. ISSN: 0003-486X. doi:[10.4007/annals.2010.171.109](https://doi.org/10.4007/annals.2010.171.109) (2010)
25. Hebisch, W., Sikora, A.: A smooth subadditive homogeneous norm on a homogeneous group. *Stud. Math.* **96**(3), 231–236. ISSN: 0039-3223 (1990)
26. Hunt, G.A.: Semi-groups of measures on Lie groups. *Trans. Am. Math. Soc.* **81**, 264–293. ISSN: 0002-9947 (1956)
27. Kallenberg, O.: Foundations of modern probability. Probability and its Applications (New York), pp. xii+523. Springer, New York. ISBN: 0-387-94957-7 (1997)
28. Kolokoltsov, V.N.: Markov processes, semigroups and generators. de Gruyter Studies in Mathematics, vol. 38, pp. xviii+430. Walter de Gruyter & Co., Berlin. ISBN: 978-3-11-025010-7 (2011)
29. Kunita, H.: Some problems concerning Lévy processes on Lie groups. *Stochastic analysis* (Ithaca, NY, 1993). In: Proceedings of Symposium. Pure American Mathematical Society, Providence, RI, vol. 57, pp. 323–341. doi:[10.1090/pspum/057/1335479](https://doi.org/10.1090/pspum/057/1335479) (1995)
30. Liao, M.: Lévy processes in Lie groups. *Cambridge Tracts in Mathematics*, vol. 162, pp. x+266. Cambridge University Press, Cambridge. ISBN: 0- 521-83653-0. doi:[10.1017/CBO9780511546624](https://doi.org/10.1017/CBO9780511546624) (2004)
31. Lyons, T.J., Caruana, M., Lévy, T.: Differential equations driven by rough paths. *Lecture Notes in Mathematics*, vol. 1908, pp. xviii+109. Springer, Berlin. ISBN: 978-3-540-71284-8; 3-540-71284-4 (2007)
32. Manstavičius, M.:  $p$ -variation of strong Markov processes. *Ann. Probab.* **32**(3A), 2053–2066. ISSN: 0091-1798. doi:[10.1214/009117904000000423](https://doi.org/10.1214/009117904000000423) (2004)
33. McShane, E.J.: 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. Univ. California Press, Berkeley, California, pp. 263–294 (1972)
34. Porter, J.E.: Helly’s selection principle for functions of bounded P-variation. *Rocky Mt. J. Math.* **35**(2), 675–679. ISSN: 0035-7596. doi:[10.1216/rmj.1181069753](https://doi.org/10.1216/rmj.1181069753) (2005)
35. Sussmann, H.J.: Limits of the Wong–Zakai Type with a Modified Drift Term. *Stochastic Analysis*. Academic Press, Boston (1991)
36. Williams, D.R.E.: Path-wise solutions of stochastic differential equations driven by Lévy processes. *Rev. Mat. Iberoamericana* **17**(2), 295–329. ISSN: 0213-2230. doi:[10.4171/RMI/296](https://doi.org/10.4171/RMI/296) (2001)