TAIL ESTIMATES FOR MARKOVIAN ROUGH PATHS

By Thomas Cass[1] and Marcel Ogrodnik[2]

Imperial College London

The accumulated local p-variation functional [*Ann. Probab.* **41** (213) 3026–3050] arises naturally in the theory of rough paths in estimates both for solutions to rough differential equations (RDEs), and for the higher-order terms of the signature (or Lyons lift). In stochastic examples, it has been observed that the tails of the accumulated local p-variation functional typically decay much faster than the tails of classical p-variation. This observation has been decisive, for example, for problems involving Malliavin calculus for Gaussian rough paths [*Ann. Probab.* **43** (2015) 188–239].

All of the examples treated so far have been in this Gaussian setting that contains a great deal of additional structure. In this paper, we work in the context of Markov processes on a locally compact Polish space E, which are associated to a class of Dirichlet forms. In this general framework, we first prove a better-than-exponential tail estimate for the accumulated local p-variation functional derived from the intrinsic metric of this Dirichlet form. By then specialising to a class of Dirichlet forms on the step $\lfloor p \rfloor$ free nilpotent group, which are sub-elliptic in the sense of Fefferman–Phong, we derive a better than exponential tail estimate for a class of Markovian rough paths. This class includes the examples studied in [*Probab. Theory Related Fields* **142** (2008) 475–523]. We comment on the significance of these estimates to recent papers, including the results of Ni Hao [Personal communication (2014)] and Chevyrev and Lyons [*Ann. Probab.* To appear].

1. Introduction. Lyons's rough path theory has allowed a pathwise interpretation to be given to stochastic differential equations of the form

$$dY_t = V(Y_t)\, dX_t, \qquad Y_0 = y_0,$$

where the vector fields $V = (V^1, \ldots, V^d)$ are driven along an $\mathbb{R}^d$-valued rough random signal X. An important feature of Lyons' approach—as compared, say, to the classical framework of Itô—is the relaxation of the condition that X be a semimartingale. There is typically no way of accommodating this feature within Itô's or any comparable theory. Furthermore, there are fundamental classes of random signals where the semimartingale property is either absent, or only present in special cases, for example, Markov processes, fractional Brownian motions and, more

broadly, the family of Gaussian processes. Study of the Gaussian-driven RDEs using rough path analysis has been especially prolific over recent years; we reference [16, 27, 29, 34, 36, 50] and [4] as an illustrative, although by no means exhaustive, list of applications.

Semimartingales have a well-defined quadratic variation process. It is widely appreciated, at least for continuous semimartingales, that control of the quadratic variation provides insight on the moments, tails and deviations of the semimartingale itself. The exponential martingale inequality (see, e.g., [51]) and the Burkhölder–Davis–Gundy inequalities (see, e.g., [14]) are prime examples of this principle in practice. The latter result, in particular, allows one to control the moments of linear differential equations

$$dY_t = AY_t \, dX_t, \qquad Y_0 = y_0,$$

where A is in $\mathrm{Hom}(\mathbb{R}^d, \mathbb{R}^e)$ and X is a semimartingale. A more sophisticated example to which this idea applies is the case when Y is the derivative of the flow of an SDE, which is well known to solve an SDE with linear growth vector fields. In many applications, such as Malliavin calculus, it is crucial to show that this derivative process (and its inverse) has finite moments of all orders.

In rough path theory, by contrast, one deliberately postpones using probabilistic features of X. Indeed, a key advantage is the separation between the deterministic theory, which is used to solve the differential equation, and the probability, which is used to enhance the driving path to a rough path. This separation, however, can—and often, does—introduce complications in probabilistic applications. For instance, in trying to prove moment estimates of the type discussed in the last paragraph using a rough path approach, it is reasonable to try to integrate the natural growth estimate for the solution, which in this case has the form (see [31])

$$(1) \qquad \|\mathbf{Y}\|_{p\text{-var},[0,T]} \le C \exp(C \|\mathbf{X}\|^p_{p\text{-var},[0,T]}),$$

where $\|\mathbf{X}\|_{p\text{-}var,[0,T]}$ denotes p-variation of the rough path enhancement of X. In the case when X is the Gaussian (even Brownian) rough path, this inequality is useless for proving moment estimates because the right-hand side is not integrable; $\|\mathbf{X}\|_{p\text{-var},[0,T]}$ has only Gaussian tail. Nevertheless, it is possible to surmount this problem, as demonstrated by [18]. The key idea is to use a slight sharpening of the estimate (1) to one of the form (see [18])

$$
\|\mathbf{Y}\|_{p\text{-var},[0,T]} \le C \exp\left(\sup_{\substack{D=(t_i) \\ \|\mathbf{X}\|_{p\text{-var},[t_i,t_{i+1}]} \le 1}} \sum_{i:t_i \in D} \|\mathbf{X}\|^p_{p\text{-var},[t_i,t_{i+1}]} \right)
$$

$$(2) \qquad := C \exp[CM(\mathbf{X}, [0,T])].$$

The functional M is called the accumulated local p-variation. Later we will use the term ρ-accumulated local p-variation if $\mathbf{X}$ is not necessarily a rough path but

any path X with values in a metric space (E, ρ); in this case, the accumulated local p-variation will be measured with respect to the underlying metric ρ.

While it may not appear on first inspection that this estimate helps much, in fact, it considerably improves the tail analysis mentioned above. The main result of [18] is the following tail estimate for Gaussian rough paths $\mathbf{X}$:

$$(3) \qquad \mathbb{P}\big(M(\mathbf{X}, [0, T]) > x\big) \leq \exp(-cx^{2/q}),$$

where $q \in [1, 2)$ is a parameter related to the Cameron–Martin Hilbert space of X. It follows as a consequence that the left-hand side of (2) has moments of all orders.

The strategy for proving the estimate (3) in the Gaussian setting is somewhat subtle. The first step is to introduce the so-called p-variation greedy partition by setting

$$\tau_0 = 0 \quad \text{and} \quad \tau_{n+1} = \inf\big\{t \geq \tau_n : \|\mathbf{x}\|_{p\text{-var},[\tau_n,t]} = 1\big\} \wedge T.$$

An integer-valued random variable defined by

$$(4) \qquad N_{p\text{-var}}\big(\mathbf{x}, [0, T]\big) = \sup\big\{n \in \mathbb{N} \cup \{0\} : \tau_n < T\big\}$$

then counts the number of distinct intervals in the partition $(\tau_n)_{n=0}^{\infty}$. Second, a relatively simple argument gives

$$N_{p\text{-var}}\big(\mathbf{x}, [0, T]\big) \leq M\big(\mathbf{x}, [0, T]\big) \leq 2N_{p\text{-var}}\big(\mathbf{x}, [0, T]\big) + 1,$$

and hence the tail of the random variable $M(\mathbf{X}, [0, T])$ can be deduced from that of $N_{p\text{-var}}(\mathbf{X}, [0, T])$. Third, the estimate (3) is proved for $N_{p\text{-var}}(\mathbf{X}, [0, T])$ in place of $M(\mathbf{X}, [0, T])$; the two key tools in doing this are (Borell's) Gaussian isoperimetric inequality (see, e.g., [1, 13]), and the Cameron–Martin embedding theorem of [31].

In this paper, we study this problem for a different class of rough paths: the Markovian rough paths. Rough paths which are themselves Markov, or which are the lifts of such processes, have been studied previously. In [3], for example, the authors start with a reversible $\mathbb{R}^d$-valued continuous Markov process X having a stationary probability measure μ. By assuming a moment condition on the increments of X and by starting X in its stationary distribution, they construct a Lévy-area process as a limit of dyadic piecewise linear approximations to X. The argument uses a forward-backward martingale decomposition, in the spirit of [46], which is applied to a natural sequence of approximations to the area. The reversibility of X and the anti-symmetry of the Lévy-area are used in an attractive way to realise suitable cancellations in this approximating sequence. Earlier work by Lyons and Stoica (see [42]) has also exploited the forward-backward martingale decomposition in the construction of the Lévy-area. An alternative approach, which we will follow in our presentation, was proposed in [28] and [31]. Here $\mathbf{X} = (X, A)$ is constructed not by *enhancing* X as in the initially mentioned approach, but directly as the Markov process associated with (the Friedrich's extension of) a Dirichlet form (see Section 3 for a review of this idea).

There are big obstacles to implementing the Gaussian approach of [18] in this setting. The most important is the lack of a usable substitute for the isoperimetric inequality and, relatedly, the Cameron–Martin embedding theorem (indeed, there is no longer any Cameron–Martin space). Analogous results which exist in the literature (e.g., [1, 15]) do not seem easy to implement here. As a consequence, we have to re-think the whole strategy upon which [18] is founded. In so doing, we gain important insights into the general principles for proving estimates of the type (3). In summary, these are:

1. That it can be useful to determine the greedy partition $(\sigma_n)_{n=0}^{\infty}$ from a metric topology which is weaker than the p-variation rough path topology. Let d denote the metric, and $N_d(\mathbf{x}, [0, T])$ the integer corresponding to the greedy partition under this metric. Then, clearly, $N_d(\mathbf{x}, [0, T]) \leq N_{p\text{-var}}(\mathbf{x}, [0, T])$. This has the immediate advantage of making the proof of the tail estimate for $N_d(\mathbf{X}, [0, T])$ easier to prove than for $N_{p\text{-var}}(\mathbf{X}, [0, T])$. The price one pays is that it is no longer true that

$$\|\mathbf{X}\|_{p\text{-var},[\sigma_n,\sigma_{n+1}]} \leq 1 \qquad \text{for all } n = 0, 1, 2, \dots.$$

Nevertheless, the control of $\mathbf{X}$ in some topology—even a weaker one than p-variation—is often sufficient to dramatically improve the tail behaviour of the random variable $\|\mathbf{X}\|_{p\text{-var},[\sigma_n,\sigma_{n+1}]}$. Similar observations to this have been made before in other contexts, for example, [43] and in support theorems, [9, 25, 41].

2. A natural choice of metric in the regime of this paper is the supremum of the intrinsic metric induced by the Dirichlet form. In the present setting, we can control the tails on N using a combination of large deviations estimates, Gaussian heat kernel bounds and exponential Tauberian theorems. For other examples, a different way of obtaining these bounds will be needed. But the study of tail estimates for the maximum of a stochastic process is a much more widely addressed subject than the corresponding study for p-variation; see, for example, [57]. There are likely to be many more examples which can be approached by adapting these methods.

We have already mentioned some applications. Without giving an exhaustive list, or trying to anticipate all future uses of this work, we briefly summarise what we believe will be the most immediately obvious sources of impact. The chief application of [18] has been in Gaussian Hörmander theory to prove, for example, smoothness and other properties of the density for Gaussian RDEs (see, e.g., [5–7, 17, 34]). A similar approach might be attempted with Markovian signals, but one has to be careful—unlike in the Gaussian setting, the driving Markov process will no longer have a smooth density in general. Nevertheless, it is interesting to consider whether the Itô map preserves the density (and its derivatives—if it has any) under Hörmander's condition. Here, the Malliavin method will radically break down; abstract Wiener analysis will need to be replaced by analysis of the Dirichlet form.

Second, growth estimates involving the accumulated p-variation occur naturally and generically in rough path theory; see [26] for a range of examples. We therefore expect uses of our results to be widespread. In [19], it was observed that $M(\mathbf{X}, [0, T])$ appears in optimal Lipschitz-estimates on the rough path distance between two different RDE solutions. This has uses in fixed-point arguments, for example, in studying interacting McKean–Vlasov-type RDEs.

Another illustration of the use of our result can be found in very interesting recent papers [35] and [20]. In these papers, the authors prove criteria for the law of a geometric rough path to be determined by its expected signature. These criteria are formulated in terms of the power series

$$(5) \qquad \sum_{n=1}^{\infty} \lambda^n \big|\mathbb{E}[\mathbf{X}_{0,T}^n]\big|_n,$$

where $S(\mathbf{X})_{0,T} = \sum_{k=0}^{\infty} \mathbf{X}_{0,T}^k$ denotes the signature of a geometric rough path $\mathbf{X}$, and $|\cdot|_n$ is a suitable norm on $(\mathbb{R}^d)^{\otimes n}$. An important result in [20] is that the radius of convergence of (5) being infinite is sufficient for $\mathbb{E}[S(\mathbf{X})_{0,T}]$ to determine the law of $\mathbf{X}$ uniquely over $[0, T]$. The work of Ni Hao [35] and Friz and Riedel [26] complements this result by proving an upper bound on the signature $S(\mathbf{X})_{0,T}$ in terms of $N_{p\text{-var}}(\mathbf{X}, [0, T])$. In [35], these estimates are then used to show that if $N_{p\text{-var}}(\mathbf{X}, [0, T])$ has a Gaussian tail then the radius of convergence of the series (5) is infinite. In [20], this statement is refined to show that any better-than-exponential tail of $N_{p\text{-var}}(\mathbf{X}, [0, T])$ suffices for the same conclusion, and that a somewhat weaker determination of the law of $\mathbf{X}$ is possible when the tail is only exponential. One example cited in [20] is the class of Markovian rough paths stopped on leaving a domain (the domain is required to have some boundedness properties in order for it to have a well-defined diameter). For this class, they are able to show exponential integrability of $N_{p\text{-var}}(\mathbf{X}, [0, T])$. Our main result, Theorem 5.4, imposes no restriction on the domain of $\mathbf{X}$ and we prove a stronger tail-estimate; more exactly, we prove one which is better-than-exponential in the sense that

$$\mathbb{P}\big(M(\mathbf{X}; [0, T]) > R\big) \leq C \exp\big(-C R^{2(1-1/p)}\big) \qquad \text{for any } p > 2, \text{ where } C = C_p.$$

This is obviously better than just exponential decay, and it has the consequence that one may verify the stronger criterion mentioned in the above work. One immediate application of our results therefore is to broaden substantially the range of examples to which the results of Chevyrev–Lyons–Ni Hao [20, 35] are known to apply.

The outline of the article is as follows. In Section 2, we give a general overview of the results of rough path theory required for our analysis. In Section 3, we review the theory and key results for symmetric Markov processes associated to a certain class of Dirichlet forms. In Section 4, we use large deviations techniques and exponential Tauberian theorems to prove that under the intrinsic metric the

integer associated to the greedy partition of the just mentioned Markov processes has a Gaussian tail. The main work is done in Section 5, where we prove a crucial bound on the ρ-accumulated local p-variation in terms of the aforementioned integer of the greedy partition and the accumulated p-variation of the Markov process between the points of this partition. This result in concert with heat kernel estimates and the results of Section 3 allows us to prove our main theorem, that is, the ρ-accumulated local p-variation of a Markov process has better-than-exponential tails provided that ρ is locally controlled by the intrinsic metric. In Section 6, we gather several examples of Markovian rough paths, that is, Markov processes associated to certain Dirichlet forms that are also rough paths, for which the main result holds.

2. Rough paths. There are now many texts which outline the core content of rough path theory (e.g., [31, 44, 45] and [30]). Here, we focus on gathering together relevant notation.

To start with, assume V is a d-dimensional real vector space. Then a basic role in the theory is played by the truncated tensor algebra which for $N \in \mathbb{N}$ is the set

$$T^N(V) := \{g = (g^0, g^1, \ldots, g^N) : g^k \in V^{\otimes k}, k = 0, 1, \ldots, N\}$$

equipped with the truncated tensor product. Two subsets of $T^N(V)$ of particular interest are

$$\tilde{T} := \tilde{T}^N(V) := \{h \in T^N(V) : g^0 = 1\}$$

and

$$\tilde{\mathfrak{t}} := \tilde{\mathfrak{t}}^N(V) := \{A \in T^N(V) : A^0 = 0\}.$$

It is easy to see that $\tilde{T}$ is a group under truncated tensor multiplication. In fact, it is a Lie group and the vector space $\tilde{\mathfrak{t}}$ is its Lie algebra $\mathrm{Lie}(\tilde{T})$, that is, $\tilde{\mathfrak{t}}$ is tangent space to $\tilde{T}$ at the group identity 1. The diffeomorphisms $\log : \tilde{T} \to \tilde{\mathfrak{t}}$ and $\exp : \tilde{\mathfrak{t}} \to \tilde{T}$ defined respectively by the power series

$$\log(g) = \sum_{k=1}^{N} \frac{(-1)^{k-1}}{k} (g-1)^k \quad \text{and} \quad \exp(A) = \sum_{k=0}^{N} \frac{1}{k!} A^k$$

are mutually inverse, and log defines a global chart on $\tilde{T}$. The map exp coincides with the Lie group exponential, that is, for every A, $\exp(A) = \gamma_A(1)$ where $\gamma_A : \mathbb{R} \to \tilde{T}$ is the unique integral curve through the identity of the left-invariant vector field associated with A.

In the paper, it will be useful to realise the group structure of $\tilde{T}$ on the set $\tilde{\mathfrak{t}}$. To do this, we define a product $* \colon \tilde{\mathfrak{t}} \times \tilde{\mathfrak{t}} \to \tilde{\mathfrak{t}}$ using the functions exp and log as follows:

$$A * B := \log(\exp(A) \exp(B)) \qquad \text{for all } A, B \in \tilde{\mathfrak{t}}.$$

Under this definition, $(\tilde{\mathfrak{t}}, *)$ is again a Lie group with identity element 0, and exp is then a Lie group isomorphism from $(\tilde{\mathfrak{t}}, *)$ to $\tilde{T}$. The differential of exp at 0 then pushes forward tangent vectors in $T_0\tilde{\mathfrak{t}}$ to elements of the vector space $\tilde{\mathfrak{t}}$. This linear isomorphism is easily seen to be the identity map on $\tilde{\mathfrak{t}}$, hence $\text{Lie}(\tilde{\mathfrak{t}}, *) = \tilde{\mathfrak{t}}$ as a vector space. The Lie group exponential map $\text{Lie}(\tilde{\mathfrak{t}}, *) \to (\tilde{\mathfrak{t}}, *)$ also equals the identity map on $\tilde{\mathfrak{t}}$, and the Campbell–Baker–Hausdorff formula (see [22, 52]) can be used to show that the Lie bracket induced by $(\tilde{\mathfrak{t}}, *)$ agrees with $AB - BA$, the commutator Lie bracket derived from the original truncated tensor multiplication.

We let $\mathfrak{g}^N := \mathfrak{g} = \text{Lie}(V)$ be the Lie algebra generated by V. The vector space $\mathfrak{g}$ is an embedded submanifold of $\tilde{\mathfrak{t}}$ and is also a subgroup of $(\tilde{\mathfrak{t}}, *)$ under the product $*$. It follows that $(\mathfrak{g}, *)$ is a Lie group, which we call the step-N nilpotent Lie group with d generators. The Lie algebra associated with $(\mathfrak{g}, *)$ is the vector space $\mathfrak{g}$.

DEFINITION 2.1.   For any $a \in V$, we define B_a to be the unique left-invariant vector field on $(\mathfrak{g}, *)$ associated with $(0, a, 0, \ldots, 0) \in \mathfrak{g}$. Given $A \in \mathfrak{g}$ we then define the horizontal subspace $\mathcal{H}_A$ at $A \in \mathfrak{g}$ to be the vector subspace of $\mathfrak{g}$ given by

$$\mathcal{H}_A = \text{span}\{B_a(A) : a \in V\}.$$

An absolutely continuous curve $\gamma : [0, T] \to \mathfrak{g}$ is then said to be horizontal if $\dot{\gamma}(t) \in \mathcal{H}_{\gamma(t)}$ for almost every $t \in [0, T]$.

REMARK 2.2.   For example, when $N = 2$ a simple calculation shows that

$$B_a(A) = a + \frac{1}{2}[A^1, a], \qquad \text{where } A = (A^1, A^2).$$

We will equip V with a norm and consider paths x belonging to $C^{1\text{-var}}([0, T], V)$, the space of continuous V-valued paths of finite 1-variation $\|\mathbf{x}\|_{1\text{-var};[0,T]}$. The truncated signature $S_N(x)$ of x is defined by

$$S_N(x)_{0,\cdot} := 1 + \sum_{k=1}^{N} \int_{0 < t_1 < \cdots < t_k < \cdot} dx_{t_1} \otimes \cdots \otimes dx_{t_k} =: 1 + \sum_{k=1}^{N} \mathbf{x}_{0,\cdot}^k \in \tilde{T}^N(V).$$

It is well known (see [31]) that $\log S_N(x)_{0,\cdot}$ is a path which takes values in the group $(\mathfrak{g}, *)$. Any horizontal curve starting from 0, the identity in $(\mathfrak{g}, *)$, can be realised as the unique solution to

$$d\gamma_t = B_{dx_t}(\gamma_t), \qquad \gamma_0 = 0.$$

This so-called horizontal lift of x is easily shown to equal $\log S_N(x)_{0,\cdot}$.

A classical theorem of Chow (see, e.g., [33, 48]) shows that any distinct points in $\mathfrak{g}$ can be connected by a horizontal curve (which is smooth in the case $N = 2$). This

gives rise to the Carnot–Carathéodory norm on $(\mathfrak{g}, *)$ as the associated geodesic distance

$$(6) \quad \|g\|_{CC} := \inf\{\|x\|_{1\text{-var};[0,T]} : x \in C^{1\text{-var}}([0, T], V) \text{ and } \log S_N(x)_{0,T} = g\}.$$

The function $\|\cdot\|_{CC}$ has the property of being a *homogeneous norm on* $(\mathfrak{g}, *)$. By this, we mean a map $\|\cdot\| : (\mathfrak{g}, *) \to \mathbb{R}_{\geq 0}$ which vanishes at the identity and is homogeneous in the sense that

$$\|\delta_r g\| = |r|\|g\| \qquad \text{for every } r \in \mathbb{R},$$

wherein $\delta_r : \mathfrak{g} \to \mathfrak{g}$ is the restriction to $\mathfrak{g}$ of the scaling operator $\delta_r : T^N(V) \to T^N(V)$ defined by

$$\delta_r : (g^0, g^1, g^2, \dots, g^N) \to (g^0, rg^1, r^2 g^2, \dots, r^N g^N).$$

In finite dimensions it is a basic fact ([31]) that all such homogeneous norms are Lipschitz equivalent, and the subset of symmetric and subadditive homogeneous norms gives rise to metrics on $(\mathfrak{g}, *)$. The one which we will use most often is the left-invariant Carnot–Carathéodory metric d_{CC} determined from (6) by

$$d_{CC}(g, h) = \|g^{-1} * h\|_{CC}, \qquad g, h \in \mathfrak{g}.$$

For any path $\mathbf{x} : [0, T] \to (\mathfrak{g}, *)$, the group structure provides us with a natural notion of increment given by $\mathbf{x}_{s,t} := \mathbf{x}_s^{-1} * \mathbf{x}_t$. For each α in $(0, 1]$ and p in $[1, \infty)$, we can then let $C^{\alpha\text{-Höl}}([0, T], \mathfrak{g})$ and $C^{p\text{-var}}([0, T], \mathfrak{g})$ be the subsets of the continuous $\mathfrak{g}$-valued paths such that the following, respectively, are finite real numbers:

$$(7) \qquad \|\mathbf{x}\|_{\alpha\text{-Höl};[0,T]} := \sup_{\substack{[s,t] \subseteq [0,T], \\ s \neq t}} \frac{\|\mathbf{x}_{s,t}\|_{CC}}{|t - s|^\alpha},$$

$$(8) \qquad \|\mathbf{x}\|_{p\text{-var};[0,T]} := \left(\sup_{D=(t_j)} \sum_{j:t_j \in D} \|\mathbf{x}_{t_j, t_{j+1}}\|_{CC}^p \right)^{1/p},$$

where, in the latter, the supremum runs over all partitions D of the interval $[0, T]$.

DEFINITION 2.3. For $p \geq 1$, we let

$$WG\Omega_p(V) := WG\Omega_p([0, T], V) := C^{p\text{-var}}([0, T], \mathfrak{g}^{\lfloor p \rfloor}).$$

We call $WG\Omega_p(V)$ the set of weakly[2] geometric p-rough paths.

REMARK 2.4. Note that $C^{1/p\text{-Höl}}([0, T], \mathfrak{g}^{\lfloor p \rfloor}) \subset WG\Omega_p(V)$.

---

[2]The prefix *weakly* here is really a misnomer; what are customarily called weakly geometric rough paths really ought to be called geometric rough paths. We persist with it for the sake of consistency with the literature.

The definitions (7) and (8) can be easily extended for any compact subset $I \subset \mathbb{R}$ by simply replacing $[0, T]$ by I. We will also consider the case where $I = [0, \infty)$, by which we mean the following.

DEFINITION 2.5. For $p \geq 1$, we define $C^{p\text{-var}}([0, \infty), \mathfrak{g})$ to be the subset of the continuous $\mathfrak{g}$-valued paths, $C([0, \infty), \mathfrak{g})$ as follows:

$$C^{p\text{-var}}\big([0, \infty), \mathfrak{g}\big) := \big\{ \mathbf{x} \in C\big([0, \infty), \mathfrak{g}\big) : \forall T \geq 0, \mathbf{x}|_T \in C^{p\text{-var}}\big([0, T], \mathfrak{g}\big) \big\},$$

where $\mathbf{x}|_T$ denotes the restriction of a path $\mathbf{x}$ on $[0, \infty)$ to one on $[0, T]$. We define $C^{1/p\text{-Höl}}([0, \infty), \mathfrak{g})$ similarly.

We will later need the fact that for $\mathbf{x} \in C^{p\text{-var}}([0, T], \mathfrak{g})$ the map

$$(9) \qquad \omega_{\mathbf{x}}(s, t) := \|\mathbf{x}\|_{p\text{-var};[s,t]}^p$$

is a *control*; by this we mean it is a continuous, nonnegative, super-additive function on the simplex $\Delta_T := \{(s, t) \in [0, T] \times [0, T] : 0 \leq s \leq t \leq T\}$ which is zero on the diagonal (see [31], page 80).

3. Markov processes induced by Dirichlet forms.

We now recall some basic facts from the theory of Dirichlet forms and the corresponding probabilistic study of symmetric Markov processes. The most prominent references for our setting include [10, 24, 32, 38, 53–56] and [47].

We will consider Markov processes constructed on a locally compact Polish space E. When working with examples from rough paths theory, we will specialise to $E = \mathfrak{g} = \mathfrak{g}^N$, as considered in the last section.

We assume throughout that E is equipped with a Radon measure μ that has full support. We let $\mathcal{E}$ denote a Dirichlet form with domain $D(\mathcal{E}) \subseteq L^2(E, \mu) = L^2(\mu)$, which we assume to be symmetric, strongly local and strongly regular in the sense of [56]. For suitable $f, g \in D(\mathcal{E})$, a Dirichlet form of this type can be written as

$$\mathcal{E}(f, g) = \int_E d\Gamma(f, g),$$

where Γ is a positive semidefinite bilinear form on $D(\mathcal{E})$ taking values in the space of signed Radon measure, and is called the energy measure associated with $\mathcal{E}$.

To every such Dirichlet form $\mathcal{E}$, one can associate a nonpositive self-adjoint operator $\mathcal{L}$. Classical theory then provides the route from $\mathcal{L}$ to a semi-group of contractions on $L^2(\mu)$, which we will denote $(P_t)_{t \geq 0}$, and thence to an associated E-valued Markov process. The assumption of strong regularity implies that the intrinsic metric, which we recall below, associated with $\mathcal{E}$ is a genuine metric on E, and that the metric topology coincides with the original topology on E. We denote this intrinsic metric by d; it is defined for all $x, y \in E$ by

$$(10) \qquad \begin{aligned} &d(x, y) \\ &\quad = \sup\{f(x) - f(y) : f \in \mathcal{F}_{\text{loc}}, f : E \to \mathbb{R} \text{ continuous}, d\Gamma(f, f) \leq d\mu\}, \end{aligned}$$

wherein $\mathcal{F}_{\text{loc}} := \{f \in L^2(\mu) : \Gamma(f, f) \text{ is a Radon measure}\}$. The following list enumerates the basic conditions we impose on the Dirichlet space $(\mathcal{E}, D(\mathcal{E}), L^2(E, \mu))$.

CONDITION 3.1. *Let $\mathcal{E}$ be a symmetric strongly local, strongly regular Dirichlet form. Denote by $B_d(x, r) = B(x, r)$ the ball of radius r around $x \in E$ with respect to d, the intrinsic metric associated to $\mathcal{E}$. We will assume that the following properties are satisfied:*

(I) (*Completeness property*) *The metric space (E, d) is complete.*
(II) (*Doubling property*) *There exists $M > 0$ such that*

$$\forall r > 0, \qquad x \in E : \mu\big(B(x, 2r)\big) \leq 2^M \mu\big(B(x, r)\big).$$

(III) (*Weak Poincaré inequality*) *There exists $C_P > 0$ such that for all $r > 0$, $x \in E$ and $f \in D(\mathcal{E})$ we have*

$$\int_{B(x,r)} |f - \bar{f}_r|^2 \, d\mu \leq C_P r^2 \int_{B(x,2r)} d\Gamma(f, f),$$

where

$$\bar{f}_r = \mu\big(B(x, r)\big)^{-1} \int_{B(x,r)} f \, d\mu.$$

With our later applications in mind, it is fruitful to note that these conditions are satisfied, for example, by the Dirichlet form (see [31, 56])

$$\mathcal{E}^a(f, g) := \sum_{i,j=1}^{d} \int_{\mathbb{R}^N} a^{ij}(x) \frac{\partial}{\partial x_i} f(x) \frac{\partial}{\partial x_j} g(x) \, d\mu(x),$$

where a is a measurable map from $\mathbb{R}^N$ to a class of symmetric uniformly elliptic matrices and μ is chosen to be the usual Lebesgue measure. Uniform ellipticity is not necessary: If μ is a smooth measure and a is a smooth map with values in the class of symmetric positive semi-definite matrices such that $\mathcal{E}^a$ is sub-elliptic in the Fefferman–Phong sense, then, too, the above conditions are satisfied (see [8, 12, 38, 56])—they are even satisfied for $\mathcal{E}^b$ if b is merely uniformly sub-elliptic with respect to such an $\mathcal{E}^a$.

We will revisit these examples later when we discuss the applications of our results in the rough paths framework; for now we will continue in the more general setting of a Dirichlet space that satisfies the above conditions.

Under the assumptions (I), (II), (III), the semi-group $(P_t)_{t\geq 0}$ referred to above is easily seen by Sobolev estimates (see, e.g., [21]) to admit a kernel representation so that

$$(P_t f)(x) = \int f(y) p(t, x, y) \mu(dy).$$

The heat kernel p can be shown to satisfy the following upper bound.

THEOREM 3.2 (Sturm [56], Corollary 4.2). *The heat kernel p associated with the Dirichlet form $\mathcal{E}$ satisfies, for $\varepsilon > 0$ fixed,*

$$(11) \qquad p(t, x, y) \leq \frac{C}{\sqrt{\mu(B(x, t^{1/2}))\mu(B(y, t^{1/2}))}} \exp\left(-\frac{d(x, y)^2}{(4 + \varepsilon)t}\right),$$

for some constant C that only depends on the doubling and Poincaré constants of $\mathcal{E}$.

REMARK 3.3. As commented in [56], Corollary 4.2, it follows by applying the doubling property that $\mu(B(x, \sqrt{t})) \leq \mu(B(y, \sqrt{t}))2^M(1 + \frac{d(x,y)}{\sqrt{t}})^M$ and, therefore, the previous theorem also gives

$$(12) \qquad p(t, x, y) \leq \frac{C_U}{\mu(B(x, t^{1/2}))} \exp\left(-\frac{d(x, y)^2}{(4 + \varepsilon)t}\right).$$

The constant C_U here only depends on ε, M and C_P.

The heat kernel p allows for a consistent family of finite-dimensional distributions, and thus determines an E-valued (strong) Markov process $X^x = (X^x_t)_{t \geq 0}$ with $X^x_0 = x$. An important observation using Kolmogorov's criterion (see [28], Theorem 13) is that, for any $p > 2$, X^x has a version with sample paths in $C_d^{1/p\text{-Höl}}([0, \infty), E)$; that is, for every $T < \infty$

$$\left|X^x\right|_{1/p\text{-Höl};[0,T],d} := \sup_{s \neq t \in [0,T]} \frac{d(X^x_s, X^x_t)}{|t - s|^{1/p}} < \infty$$

μ-almost surely. This, in the usual way, also implies that the p-variation is finite almost surely

$$\left|X^x\right|_{p\text{-var};[0,T],d} := \sup_{D = \{0 = t_0 < t_1 < \cdots < t_n = T\}} \left(\sum_{i=1}^n d(X^x_{t_{i-1}}, X^x_{t_i})^p\right)^{1/p} < \infty.$$

In fact, much more can be shown; the following theorem is an assembly of results from [31], Proposition E.19, which we will need subsequently.

THEOREM 3.4. *Suppose $\mathcal{E}$ is a Dirichlet form satisfying Condition 3.1. Let $x \in E$ and $p > 2$. There exists a version X^x of the Markov processes associated with $\mathcal{E}$ which belongs to $C^{1/p\text{-Höl}}([0, \infty), E)$. If $\mathbb{P}^x$ denotes the probability measure on $C([0, \infty), E)$ given by the law of X^x, then for any $T > 0$ there exists a finite constant $C = C(C_P, C_D, T)$ such that*

$$\sup_{x \in E} \mathbb{P}^x\left(\sup_{[s,t] \subseteq [0,T]} \frac{d(X_s, X_t)}{|t - s|^{1/p}} > r\right) \leq C \exp\left(-\frac{r^2}{C}\right),$$

wherein $X_t : C([0, \infty), E) \to E$ denotes the canonical evaluation map $X_t(\omega) = \omega(t)$.

PROOF. The proof in [31] of the second statement works in the case where μ coincides with the Haar measure on $\mathfrak{g}$. The proof in the general setting works along the same lines; however, one needs to note that it is by exploiting the doubling property later in the calculation that one eliminates the dependence on the starting point in the constant. To be more precise, one easily sees that for fixed $s < t$ in $[0, 1]$ and $\eta > 0$ we have

$$\mathbb{E}^x\left[\exp\left(\eta \frac{d(X_s, X_t)^2}{t - s}\right)\right] \leq \sup_{x \in E} \mathbb{E}^x[\exp(\eta \tilde{d}(x, \tilde{X}_1)^2)],$$

where $\tilde{X}. \equiv X_{(t-s).}$ and $\tilde{d} = \frac{d}{\sqrt{t-s}}$. Writing $\tilde{p}(t, x, y)$ for the heat kernel associated with $\tilde{X}$, and $\tilde{B}$ for the balls associated with $\tilde{d}$, we find that for $\eta \in (0, 1/4)$

$$\mathbb{E}^x[\exp(\eta \tilde{d}(x, \tilde{X}_1)^2)] = \int_E \exp(\eta \tilde{d}(x, y)^2) \tilde{p}(1, x, y) \mu(dy)$$

$$\leq C_u \int_0^\infty \frac{\mu(\tilde{B}(x, r))}{\mu(\tilde{B}(x, 1))} \frac{d}{dr}\left(-\exp\left(-\left(\frac{1}{(4+\varepsilon)} - \eta\right)r^2\right)\right) dr$$

$$\leq C_u \int_0^1 \frac{d}{dr}\left(-\exp\left(-\left(\frac{1}{(4+\varepsilon)} - \eta\right)r^2\right)\right) dr$$

$$+ C_u \int_1^\infty (2r)^M \frac{d}{dr}\left(-\exp\left(-\left(\frac{1}{(4+\varepsilon)} - \eta\right)r^2\right)\right) dr$$

$$< \infty,$$

where we used the heat kernel bound in (12), a change-of-variables, integration by parts, the doubling property and again integration by parts. $\square$

Henceforth, we always work with the version of the process given by this theorem. An important further remark is that for $r > 0$ we can scale the time-parameter of X^x to form a new process $X^{r,x} = (X^x_{rt})_{t \geq 0}$. This new process is the Markov process associated to the Dirchlet form $\mathcal{E}^r := r\mathcal{E}$; its intrinsic distance equals $r^{-1/2}d$ and $\mathcal{E}^r$ again satisfies the doubling property and weak Poincaré inequality with the same constants as $\mathcal{E}$.

REMARK 3.5. When we wish to highlight the dependence of the law of X^x on $\mathcal{E}$ we will write $\mathbb{P}^{x,\mathcal{E}}, \mathbb{E}^{x,\mathcal{E}}$ etc. The scaling property then in particular shows that the $\mathbb{P}^{x,\mathcal{E}}$ distribution of $(X_{rt})_{t \geq 0}$ equals the $\mathbb{P}^{x,r\mathcal{E}}$ distribution of $(X_t)_{t \geq 0}$.

4. A large deviations result. Throughout this section, we fix a Dirichlet form $\mathcal{E}$ satisfying Condition 3.1, and again use d to denote its intrinsic metric. Given a path x in $C([0, \infty), E)$ and $r > 0$ we can define inductively a nondecreasing sequence $(\sigma_n^r)_{n=0}^\infty = (\sigma_n^r(x))_{n=0}^\infty$ by setting $\sigma_0^r = 0$, and then

$$(13) \qquad \sigma_n^r := \inf\{t \geq \sigma_{n-1}^r : d(x_{\sigma_{n-1}^r}, x_t) \geq r\}$$

for $n \in \mathbb{N}$.

DEFINITION 4.1. For any $T \geq 0$, we define the functional $N_0^r(\cdot) = N_0^r(\cdot, [0, T]) : C([0, \infty), E) \to \mathbb{N} \cup \{0\}$ by

$$N_0^r(x, [0, T]) = \sup\{n : \sigma_n^r < T\}.$$

REMARK 4.2. When $r = 1$, we will omit the superscripts and write $\sigma_n, N_0(\cdot)$ and so forth. Note that $N_0^r(x, [0, T]) < \infty$ implies that the set

$$\{\sigma_j : j = 0, 1, \ldots, N_0^r(x, [0, T])\} \cup \{T\}$$

forms a partition of the interval $[0, T]$.

It is our goal in this section to analyse the tail behaviour of the integer valued random variables $N_0^r(X^x, [0, T])$, when X^x is the Markov process associated to $\mathcal{E}$ described in Section 3. Our approach will be motivated by the following well-known example.

EXAMPLE 4.3 (Brownian motion). Let $E = \mathbb{R}$ and $B = (B_t)_{t \geq 0}$ a one-dimensional standard Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this setting, the sequence in (13) is given by

$$\sigma_0 := 0, \qquad \sigma_{n+1} := \inf\{t \geq \sigma_n : |B_t - B_{\sigma_n}| \geq 1\}.$$

It is a classical result (see, e.g., [40]) that the Laplace transform of $\sigma := \sigma_1$ satisfies

(14) $$\mathbb{E}[e^{-\lambda \sigma}] = \cosh(\sqrt{2\lambda})^{-1} \leq 2e^{-\sqrt{2\lambda}}.$$

If we let $\xi_k := \sigma_k - \sigma_{k-1}$ for $k = 1, \ldots, n$ and note that $\{\xi_k : k = 1, \ldots, n\}$ are i.i.d. with each ξ_k equal in distribution to σ, then using $\sum_{k=1}^n \xi_k = \sigma_n$, it follows that for all $\theta > 0$

$$\mathbb{P}(N_0(B, [0, 1]) \geq n) = \mathbb{P}(\sigma_n < 1) \leq e^\theta \mathbb{E}[e^{-\theta \sigma}]^n \leq 2^n e^\theta e^{-n\sqrt{2\theta}}.$$

The last expression can be minimized by the choice $\theta = 2^{-1} n^2$, which immediately yields the estimate $\mathbb{P}(N_0(B, [0, 1]) \geq n) \leq 2^n e^{-\frac{n^2}{2}} \leq c_1 e^{-c_2 n^2}$, for some c_1 and c_2 in $(0, \infty)$ which do not depend on n.

This example makes clear the importance of the Laplace transform when analysing the tail behaviour of $N_0^r(x, [0, T])$. What is important is not to have a closed-form expression as in (14), but instead to have an upper bound controlling its asymptotic behaviour as $\lambda \to \infty$.

4.1. *Tails for* $N_0^r(x, [0, T])$.   We will adopt the notation of Theorem 3.4, that is, $\mathbb{P}^x$ will be the law of X^x on $C([0, \infty), E)$ and $\mathbb{E}^x$ the corresponding expectation operator. For $t \geq 0$, we continue to denote the evaluation maps by $X_t : C([0, \infty), E) \to E$ so that $X_t(\omega) = \omega(t)$. Let $\sigma =: \sigma_1(X)$, the random variable in (13) with $r = 1$, and denote its Laplace transform under the probability measure $\mathbb{P}^x$ by

$$(15) \qquad M(\lambda; x) := M(\lambda; \mathcal{E}, x) := \mathbb{E}^x\left[e^{-\lambda \sigma}\right] = \int_{C([0,\infty), E)} e^{-\lambda \sigma(\omega)} \mathbb{P}^x(d\omega).$$

We will now state a version of De Bruijn's exponential Tauberian theorem. This well-known result relates the asymptotic behaviour of the log Laplace transform, $\log M(\lambda; \mathcal{E}, x)$ as $\lambda \to \infty$ and the log short-time probability $\log \mathbb{P}^{\mathcal{E},x}(\sigma \leq t)$ as $t \to 0+$.

LEMMA 4.4 (Exponential Tauberian theorem).   *Let $c > 0$. The following two statements are equivalent*:

1. $-\log M(\lambda; x) \sim c\sqrt{\lambda}$, *as* $\lambda \to \infty$;
2. $-\log \mathbb{P}^x(\sigma \leq t) \sim \frac{c^2}{4t}$, *as* $t \to 0+$.

PROOF.   This is an immediate consequence of applying Theorem 4.12.9 in [11], making the choice $B = \frac{c^2}{4}$ and $\phi(\lambda) = \frac{1}{\lambda}$ in the notation of that theorem. $\square$

We will not need the full strength of this equivalence. Instead, we will need the following statement which relates the asymptotic oscillations of the two functions. We give a short proof for completeness and refer the reader to [11] for much greater detail on results of this type.

LEMMA 4.5.   *Suppose there exists $c > 0$ for which*

$$(16) \qquad\qquad \limsup_{t \to 0+} t \sup_{x \in E} \log \mathbb{P}^x(\sigma \leq t) \leq -c,$$

then

$$\limsup_{\lambda \to \infty} \lambda^{-\frac{1}{2}} \sup_{x \in E} \log M(\lambda; x) \leq -\sqrt{c}.$$

PROOF.   Set $\mathbb{P}^x(\sigma \leq t) =: \mu_x(t)$. First, note that it is sufficient to show that the assumption (16) implies

$$\limsup_{\lambda \to 0+} \lambda \sup_{x \in E} \log M\left(\frac{1}{\lambda^2}; x\right) \leq -\sqrt{c}.$$

Second, we observe that for $\xi, \lambda > 0$,

$$M\left(\frac{1}{\lambda^2}; x\right) = \int_0^{\frac{\lambda}{\xi}} \exp\left(-\frac{t}{\lambda^2}\right) d\mu_x(t) + \int_{\frac{\lambda}{\xi}}^\infty \exp\left(-\frac{t}{\lambda^2}\right) d\mu_x(t)$$

(17)

$$\leq \mu_x\left(\frac{\lambda}{\xi}\right) + \exp\left(-\frac{1}{\xi\lambda}\right).$$

Finally, we use this bound and exploit the well-known fact that for any two sequences $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ of positive real numbers we have

$$\limsup_{n\to\infty} \frac{1}{n} \log(a_n + b_n) = \max\left\{\limsup_{n\to\infty} \frac{1}{n} \log a_n, \limsup_{n\to\infty} \frac{1}{n} \log b_n\right\}.$$

In the setting of (17), this gives

$$\limsup_{\lambda\to 0+} \lambda \sup_{x\in E} \log M\left(\frac{1}{\lambda^2}; x\right) \leq \limsup_{\lambda\to 0+} \lambda \log\left(\sup_{x\in E} \mu_x\left(\frac{\lambda}{\xi}\right) + \exp\left(-\frac{1}{\xi\lambda}\right)\right)$$

$$\leq \max\{-\xi c, -\xi^{-1}\},$$

where the last line uses the hypothesis (16). Because the function $(0, \infty) \ni \xi \mapsto \max\{-\xi c, -\xi^{-1}\}$ attains its global minimum $-\sqrt{c}$ at $\xi^\star = c^{-\frac{1}{2}}$, we obtain

$$\limsup_{\lambda\to 0+} \lambda \sup_{x\in E} \log M\left(\frac{1}{\lambda^2}; x\right) \leq -\sqrt{c}$$

which completes the proof. $\square$

The following lemma will make the previous result applicable to our setting.

LEMMA 4.6. *Denote by* $\mathbb{P}^x$ *and* σ, *respectively, the probability measure and stopping time defined in the statement of Lemma* 4.5. *There exist constants* $c_1, c_2 \in (0, \infty)$, *which depend only on the doubling and Poincaré constants associated with* $\mathcal{E}$ *such that for all* $t \in (0, 1/4]$

$$\mathbb{P}^x(\sigma \leq t) \leq c_1 \exp\left(-\frac{4c_2}{t}\right).$$

We defer the proof for a moment to note a useful consequence, namely that

$$\limsup_{t\to 0+} t \sup_{x\in E} \log \mathbb{P}^x(\sigma \leq t) = -4c_2 < 0.$$

This allows us to apply Lemma 4.5 and immediately deduce the following corollary.

COROLLARY 4.7. *Let $M(\lambda; x)$ denote the Laplace transform* (15) *of the stopping time σ. Then under the condition of Lemma* 4.6 *we have*

$$\limsup_{\lambda \to \infty} \sqrt{\lambda}^{-1} \sup_{x \in E} \log M(\lambda; x) \leq -2\sqrt{c_2}.$$

There consequently exists a constant $\lambda_0 \in (0, \infty)$ such that

$$(18) \qquad \sup_{x \in E} M(\lambda; x) \leq \exp(-\sqrt{c_2 \lambda}) \qquad \text{for all } \lambda \geq \lambda_0.$$

PROOF OF LEMMA 4.6. We follow [2], Proposition 6.5, where a similar upper bound is obtained in the case of uniformly elliptic diffusions. By using the Gaussian upper estimate in Theorem 3.2, we will adapt the proof for the class of Markov processes introduced earlier. First, we note that

$$\mathbb{P}^x(\sigma \leq t) \leq \mathbb{P}^x\left(\sigma \leq t, d(X_t, x) < \frac{1}{2}\right) + \mathbb{P}^x\left(d(X_t, x) \geq \frac{1}{2}\right).$$

Second, using Theorem 3.2 (with fixed $\varepsilon > 0$) and the remark which follows it, we see that the second term satisfies

$$
\begin{aligned}
\mathbb{P}^x\left(d(X_t, x) \geq \frac{1}{2}\right) &= \int_{B(x, \frac{1}{2})^c} p(t, x, y) \mu(dy) \\
&\leq \int_{B(x, \frac{1}{2})^c} \frac{C}{\mu(B(x, \sqrt{t}))} \exp\left(-\frac{d(x, y)^2}{(4 + \varepsilon)t}\right) \mu(dy) \\
&= \int_{1/2}^{\infty} \frac{C}{\mu(B(x, \sqrt{t}))} \exp\left(-\frac{r^2}{(4 + \varepsilon)t}\right) d\mu(B(x, r)) \\
&\leq \int_{1/2}^{\infty} \frac{C\mu(B(x, r))}{\mu(B(x, \sqrt{t}))} \frac{2r}{(4 + \varepsilon)t} \exp\left(-\frac{r^2}{(4 + \varepsilon)t}\right) dr \\
&\leq \int_{1/2}^{\infty} C 2^{M+1} \frac{1}{\sqrt{t}} \left(\frac{r}{\sqrt{t}}\right)^{M+1} \exp\left(-\frac{1}{(4 + \varepsilon)} \left(\frac{r}{\sqrt{t}}\right)^2\right) dr \\
&= \int_{\frac{1}{2\sqrt{t}}}^{\infty} c_1 v^{M+1} \exp(-c_2 v^2) \, dv \\
&\leq c_3 e^{-\frac{c_4}{t}},
\end{aligned}
$$

where the constants c_3 and c_4 depend only on the doubling and Poincaré constants of $\mathcal{E}$. For the first term, observe that

$$\mathbb{P}^x\left(\sigma \leq t, d(X_t, x) < \frac{1}{2}\right) \leq \int_0^t \mathbb{P}^x\left(\sigma \in ds, d(X_t, X_{\sigma_1}) \geq \frac{1}{2}\right)$$

$$= \int_0^t \mathbb{E}^x\left[1_{\{\sigma \in ds\}}\mathbb{P}^{X_\sigma}\left(d(X_{t-\sigma}, X_0) \geq \frac{1}{2}\right)\right]$$

$$\leq \int_0^t \mathbb{E}^x\left[1_{\{\sigma \in ds\}}\mathbb{P}^{X_s}\left(d(X_{t-s}, X_0)\right) \geq \frac{1}{2}\right].$$

By the same argument as in (19), we know there exist constants c_5 and c_6 which, again, depend only on the doubling and Poincaré constants of $\mathcal{E}$, such that

$$\sup_{r \leq t} \mathbb{P}^x\left(d(X_r, X_0) \geq \frac{1}{2}\right) \leq c_5 e^{-\frac{c_6}{t}}.$$

Together these bounds imply the desired result. $\quad\square$

REMARK 4.8.   We draw the reader's attention to a similar result which has been proved in [8]. There are important differences, both in the statement of the result and the proof techniques, which are significant for the later applications in the paper. For example, our main theorem, Theorem 5.4, will be based on a family of estimates derived from the above. The proof will rely on a delicate scaling argument, for which it is necessary to track carefully the dependence of the estimates on the parameters such as the starting point x and the Poincaré and doubling constants. The proof in [8] relies crucially on Takeda's inequality and properties of subsolutions to equations associated to the herein presented Dirichlet forms. Our proof by contrast is more elementary as it only relies on the upper heat kernel bound associated with the symmetric Markov process.

We can now prove the needed tail estimates for the random variables $N_0^r(X, [0, T])$ under $\mathbb{P}^x$. To do so we make the following simplifying observation.

REMARK 4.9.   The distribution of $\sigma^r = \sigma_1^r$ under $\mathbb{P}^{\mathcal{E},x}$ equals the distribution of $r^2\sigma = r^2\sigma_1^1$ under $\mathbb{P}^{r^2\mathcal{E},x}$. This is a consequence of the scaling property highlighted in Remark 3.5.

PROPOSITION 4.10.   *Let $\mathcal{E}$ be the Dirichlet form introduced in Section 3, and assume that X^x is the E-valued Markov process, defined on some probability space, associated with this form. Let $\mathbb{P}^x = \mathbb{P}^{\mathcal{E},x}$ be the (Borel) probability measure on $C([0, \infty), E)$ which is the law of the Markov X^x associated to $\mathcal{E}$, and*

let $c_2, \lambda_0 \in (0, \infty)$ be the constants in (18). For every $r > 0$, the random variable $N_0^r(\cdot, [0, T]) \colon C([0, \infty), E) \to \mathbb{N} \cup \{0\}$ in Definition 4.1 satisfies

$$(20) \qquad \mathbb{P}^x\big(N_0^r(X, [0, T]) \geq n\big) \leq \exp\Big(-\frac{c_2 n^2 r^2}{4T}\Big)$$

for all $n \geq 2T\lambda_0^{1/2} r^{-2} c_2^{-1/2}$.

PROOF. As previously, we write $\sigma_n^r = \sum_{k=1}^n \xi_k^r$, where $\xi_k^r = \sigma_k^r - \sigma_{k-1}^r$ and we aim to estimate the probability in (20). To do so, we first note that for $\lambda > 0$ we have

$$(21) \qquad \mathbb{P}^x\big(N_0^r(X, [0, T]) \geq n\big) \leq e^{\lambda T} \mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \sum_{k=1}^n \xi_k^r}\big].$$

Using the scaling property in the manner of Remark 4.9 gives

$$M_r(\lambda; \mathcal{E}, x) := \mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \xi_1^r}\big] = \mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \sigma_1^r}\big] = \mathbb{E}^{r^2 \mathcal{E}, x}\big[e^{-\lambda r^2 \sigma}\big] = M(\lambda r^2; r^2 \mathcal{E}, x),$$

where, for clarity, we have emphasised the dependence on the Dirichlet form. The inequality (18) in Corollary 4.7 then yields

$$(22) \qquad \sup_{x \in E} M_r(\lambda; \mathcal{E}, x) \leq \exp(-\sqrt{c_2 \lambda} r) \qquad \text{for all } \lambda \geq \lambda_0 r^{-2};$$

the same constant c_2 features here because the doubling and Poincaré constants for $r^2 \mathcal{E}$ are the same as those for $\mathcal{E}$. Combining the strong Markov property at the stopping time σ_{n-1}^r with an easy induction yields the estimate

$$\mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \sum_{k=1}^n \xi_k^r}\big] = \mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \sum_{k=1}^{n-1} \xi_k^r} \mathbb{E}^{\mathcal{E}, X_{\sigma_{n-1}^r}}\big[e^{-\lambda \sigma_1^r}\big]\big]$$

$$(23) \qquad\qquad \leq \mathbb{E}^{\mathcal{E}, x}\big[e^{-\lambda \sum_{k=1}^{n-1} \xi_k^r}\big] \sup_{x \in E} M_r(\lambda; \mathcal{E}, x)$$

$$\leq \sup_{x \in E} M_r(\lambda; \mathcal{E}, x)^n.$$

Using (23) together with (22) and (21) gives that

$$\mathbb{P}^x\big(N_0^r(X, [0, T]) \geq n\big) \leq \exp(\lambda T - n\sqrt{c_2 \lambda} r),$$

for all $\lambda \geq \lambda_0 r^{-2}$. The right-hand side of this bound is minimized by the choice $\lambda = T^{-2} 4^{-1} c_2 n^2 r^2$, resulting in the optimized bound

$$\mathbb{P}^x\big(N_0^r(X, [0, T]) \geq n\big) \leq \exp\Big(-\frac{c_2 n^2 r^2}{4T}\Big),$$

which holds provided $T^{-2} 4^{-1} c_2 n^2 r^2 \geq \lambda_0 r^{-2}$, that is, if $n \geq 2T\lambda_0^{1/2} r^{-2} c_2^{-1/2}$. □

5. Tail estimates for the accumulated local p-variation. The law $\mathbb{P}^x$ of the Markov process X^x constructed in Section 3 is, for any $p > 2$ and $T \geq 0$, supported in $C_d^{1/p\text{-Höl}}([0, T], E) \subset C_d^{p\text{-var}}([0, T], E) \subset C([0, T], E)$. This observation allows us to go beyond the analysis of the previous section and address the tail behaviour of the *accumulated local p-variation*. We first recall the definition of this functional (cf. [18]). Later we will wish to compare this functional for different metrics on E, so we reference the metric in the definition.

DEFINITION 5.1 (Accumulated local p-variation). Let $p \geq 1$ and suppose ρ is a metric on E. We define the accumulated local p-variation to be the function $M(\cdot, [0, T], \rho) = M(\cdot) : C_\rho^{p\text{-var}}([0, \infty), E) \to \mathbb{R}_{\geq 0}$ given by

$$(24) \qquad M(x, [0, T], \rho) := \sup_{\substack{D=\{0=t_0 < t_1 < \cdots < t_n = T\} \\ \omega_x(t_i, t_{i+1}) \leq 1}} \sum_{i=1}^n \omega_x(t_{i-1}, t_i),$$

where ω_x is the control induced by the p-variation of x w.r.t. ρ, that is, for $s \leq t$

$$\omega_x(s, t) = |x|^p_{p\text{-var};[s,t],\rho}.$$

The supremum in (24) is taken over the set of all partitions D of the interval $[0, T]$ such that ω_x, when evaluated between two consecutive points in D, is bounded above by unity.

We will now show that the accumulated local p-variation of a path x over $[0, T]$ can be bounded by the number of points in an arbitrary partition of that interval and the accumulated p-variation between the points of this partition.

LEMMA 5.2. *Let $p \geq 1, T > 0$, suppose ρ is a metric on E and assume $x \in C_\rho^{p\text{-var}}([0, \infty), E)$. We use ω_x to denote the control induced by the p-variation of x w.r.t. ρ as introduced in Definition 5.1. If $\{0 = \sigma_0 < \sigma_1 < \cdots < \sigma_N = T\}$ is an arbitrary partition of $[0, T]$, then we can bound $M(x, [0, T], \rho)$ from above using the following estimate*:

$$(25) \qquad M(x, [0, T], \rho) \leq N - 1 + \sum_{j=1}^N \omega_x(\sigma_{j-1}, \sigma_j).$$

PROOF. Suppose $D = \{0 = t_0 < t_1 < \cdots < t_n = T\}$ is an arbitrary partition of $[0, T]$, such that any two consecutive points $s < t$ in D satisfy $\omega_x(s, t) \leq 1$. We define the function $\Phi : \{0, 1, \ldots, n\} \to \{0, 1, \ldots, N\}$ by

$$\Phi(i) = \max\{k \in \mathbb{N} \cup \{0\} : \sigma_k \leq t_i\} \qquad \text{for } i = 0, 1, \ldots, n,$$

and then let A denote the subset

$$A = \{k < N - 1 : \exists i \text{ with } \Phi(i) = k\} \subseteq \{0, 1, \ldots, N - 2\}.$$

For each $k \in A$, we define

$$m_k = \min\{i : \Phi(i) = k\} \quad \text{and} \quad n_k = \max\{i : \Phi(i) = k\},$$

whereupon it is an easy consequence that we have $\sigma_k \leq t_{m_k} < t_{m_k+1} < \cdots < t_{n_k} < \sigma_{k+1}$, and hence

$$\sum_{j=m_k}^{n_k} \omega_x(t_j, t_{j+1}) \leq \omega_X(\sigma_k, \sigma_{k+1}) + 1.$$

To finish, we note that

$$\sum_{i=1}^{n} \omega_x(t_{i-1}, t_i) \leq \sum_{k \in A} \sum_{j=m_k}^{n_k} \omega_x(t_j, t_{j+1}) + \omega_x(\sigma_{N-1}, \sigma_N)$$

$$\leq \sum_{k \in A} \left[\omega_x(\sigma_k, \sigma_{k+1}) + 1 \right] + \omega_x(\sigma_{N-1}, \sigma_N)$$

$$\leq \sum_{k=0}^{N-2} \left[\omega_x(\sigma_k, \sigma_{k+1}) + 1 \right] + \omega_x(\sigma_{N-1}, \sigma_N)$$

$$\leq \sum_{k=1}^{N} \omega_x(\sigma_{k-1}, \sigma_k) + N - 1,$$

and since the right-hand side of the previous estimate no longer depends on D, we can take the supremum over all D satisfying the constraint in Definition 5.1. The conclusion (25) then follows immediately. □

We are now ready to prove the main result. Before doing so, we introduce the following notion.

CONDITION 5.3. *Let d and ρ be two metrics on E. We say that ρ is locally controlled by d if there exist $r_0 > 0$ and $C > 0$ such that for every $e \in E$ we have $B_d(e, r) \subseteq B_\rho(e, Cr)$. for all $r < r_0$.*

In other words, d and ρ satisfying this condition have the property that

$$(26) \qquad \rho(x, y) \leq Cd(x, y) \qquad \text{for all } x, y \in E \text{ such that } d(x, y) < r_0.$$

THEOREM 5.4. *Let $\mathcal{E}$ be the Dirichlet form satisfying Condition 3.1 of Section 3, with associated intrinsic distance d. Given $x \in E$ and $p > 2$, assume X^x is the E-valued Markov process associated to $\mathcal{E}$ which is described in Theorem 3.4, and let $\mathbb{P}^{\mathcal{E},x} = \mathbb{P}^x$ be the (Borel) probability measure on $C([0, \infty), \mathfrak{g})$ under which the canonical process X has the same distribution as X^x. We continue to use $c_2, \lambda_0 \in (0, \infty)$ to denote the constants in (18). Suppose ρ is any metric on E which is locally controlled by d (Condition 5.3) with $C, r_0 > 0$ such that*

$B_d(e, r) \subseteq B_\rho(e, Cr)$ *for all* $e \in E$, *and all* $r \leq r_0$. *Then there exist finite constants* $C_1, C_2, C_3 > 0$, *which depend only on* p, N, T, *and on the doubling and Poincaré constants associated to* $\mathcal{E}$, *such that for any* $r < \frac{r_0}{2}$ *we have*

$$(27) \quad \mathbb{P}^{\mathcal{E},x}\big(M(X, [0, T], \rho) > R\big) \leq \exp(-C_1 r^2 R^2) + R r^{-p} C_2 \exp(-C_3 R r^{2-p})$$

for all $R \geq (16\lambda_0 c_2^{-1})^{1/2} r^{-2}$. *In particular, by choosing* $r := R^{-1/p} < \frac{r_0}{2}$ *in* (27), *this yields a better-than-exponential tail for the accumulated local p-variation functional; that is for some finite $C > 0$, which depends only on p, N, T and on the doubling and Poincaré constants associated to $\mathcal{E}$, we have*

$$(28) \qquad \mathbb{P}^{\mathcal{E},x}\big(M(X, [0, T], \rho) > R\big) \leq C \exp\big(-C R^{2(1-1/p)}\big)$$

for all $R \geq \max\{(16\lambda_0 c_2^{-1})^{p(2p-4)^{-1}}, (\frac{r_0}{2})^{-p}\}$.

REMARK 5.5. At first glance, the reader may be surprised that the quality of the tail seems to improve for larger p. This comes about because of the specific way in which scaling is used in the proof (see below). This does not give rise to a contradiction because, unlike for p-variation, there is no way of ordering the accumulated p and q variation for different values of p and q. Indeed, if $1 \leq p < q$, while it is true that

$$\sum_{i=1}^{n} |X|_{q\text{-var};[t_{i-1},t_i],\rho}^q \leq \sum_{i=1}^{n} |X|_{p\text{-var};[t_{i-1},t_i],\rho}^p$$

$$\text{if } |X|_{p\text{-var};[t_{i-1},t_i],\rho} \leq 1 \text{ for } i = 1, \ldots, n,$$

the set of partitions over which one optimises these functionals to form the accumulated variation is larger for q than for p.

As p tends to infinity, the tail approaches Gaussian. If $E = \mathfrak{g}$, $\rho = d_{CC}$ and $\mathcal{E}(f, g) = \sum_{i=1}^{d} \int_{\mathfrak{g}} B_i f\, B_i g\, dm$, then $X = \mathbf{B}$ is the Brownian rough path. For this example, we know from [18] that Borell's inequality yields the full Gaussian tail for the accumulated p-variation for any p. The decay rate here is, in all cases, better than exponential, and hence suffices for the important examples referred to in the introduction. But it remains open and unclear whether isoperimetric arguments can be used in this setting of general Markov processes to obtain a Gaussian tail.

PROOF. We will prove the main estimate (27) by using the family of estimates in Lemma 5.2, when the partition is taken to be

$$\big\{0 = \sigma_0^r < \sigma_1^r < \cdots < \sigma_{N_0(X,[0,T])}^r < \sigma_{N_0(X,[0,T])+1}^r := T\big\}$$

with the σ_j^r's given as (13) and we have re-defined $\sigma_{N_0(X,[0,T])+1}^r := T$ for notational convenience. We will assume that $T = 1$ and write $M(X; \rho)$ and $N_0^r(X)$ in lieu of $M(X; [0, T], \rho)$ and $N_0^r(X, [0, T])$, respectively. The assumption $T = 1$

involves no loss of generality because of the scaling property. First, note from Lemma 5.2 that we can bound $M(X; \rho)$ by

$$N_0^r(X) + \sum_{j=0}^{N_0^r(X)} \omega_X(\sigma_j^r, \sigma_{j+1}^r).$$

It follows that for any $R > 0$ and $r > 0$

$$\{\omega : M(X(\omega); \rho) > R\}$$

$$\subset \left\{\omega : N_0^r(X(\omega)) > \frac{R}{2}\right\} \cup \left\{\omega : \sum_{j=0}^{N_0^r(X(\omega))} \omega_X(\sigma_j^r, \sigma_{j+1}^r) > \frac{R}{2}\right\}.$$

A simple estimate then gives

$$(29) \quad \mathbb{P}^x(M(X; \rho) > R) \leq \mathbb{P}^x\left(N_0^r(X) > \frac{R}{2}\right) + \mathbb{P}^x\left(\sum_{j=0}^{N_0^r(X)} \omega_X(\sigma_j^r, \sigma_{j+1}^r) > \frac{R}{2}\right)$$

for all $R > 0$ and $r > 0$.

By Proposition 4.10,

$$\mathbb{P}^x\left(N_0^r(X) > \frac{R}{2}\right) \leq \exp\left(-\frac{c_2 R^2 r^2}{16}\right)$$

for all $R \geq (16\lambda_0 c_2^{-1})^{1/2} r^{-2}$. It remains to treat the second term on the right-hand side in (29). To this end, we note the following elementary inequality:

$$\omega_X(\sigma_i^r, \sigma_{i+1}^r) \leq |X|_{1/p\text{-Höl};[\sigma_i^r, \sigma_{i+1}^r], \rho}^p (\sigma_{i+1}^r - \sigma_i^r).$$

We assume $r < r_0$ so that in particular we have $\rho(X_s, X_t) \leq C d(X_s, X_t)$ whenever $[s, t] \subset [\sigma_i^r, \sigma_{i+1}^r]$. We then notice that for any $h > 0$

$$|X|_{1/p\text{-Höl};[\sigma_i^r, \sigma_{i+1}^r], \rho}^p \leq \sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [\sigma_i^r, \sigma_{i+1}^r]}} \frac{\rho(X_s, X_t)^p}{|t-s|} + \sup_{\substack{s \neq t, |t-s| > h, \\ [s,t] \subset [\sigma_i^r, \sigma_{i+1}^r]}} \frac{\rho(X_s, X_t)^p}{|t-s|}$$

$$\leq \sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [\sigma_i^r, \sigma_{i+1}^r]}} \frac{C^p d(X_s, X_t)^p}{|t-s|} + \frac{(2Cr)^p}{h},$$

where the last line uses the definition of σ_i^r and σ_{i+1}^r. Using the equality $\sum_{i=0}^{N_0^r(X)} (\sigma_{i+1}^r - \sigma_i^r) = 1$, we thus have for any $h > 0$

$$\sum_{i=0}^{N_0^r(X)} \omega_X(\sigma_i^r, \sigma_{i+1}^r) \leq \sum_{i=0}^{N_0^r(X)} \left[\sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [\sigma_i^r, \sigma_{i+1}^r]}} \frac{C^p d(X_s, X_t)^p}{|t-s|} (\sigma_{i+1}^r - \sigma_i^r)\right] + \frac{2^p C^p r^p}{h}$$

$$\leq \sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [0,1]}} \frac{C^p d(X_s, X_t)^p}{|t-s|} + \frac{2^p C^p r^p}{h}.$$

Applying this estimate with the choice $h = 2^{p+2}R^{-1}r^pC^p$, we obtain

$$\sum_{i=0}^{N_0^r(X)} \omega_X(\sigma_i^r, \sigma_{i+1}^r) \leq \sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [0,1]}} \frac{C^p d(X_s, X_t)^p}{|t-s|} + \frac{R}{4}$$

and consequently it suffices to bound

$$\mathbb{P}^x \left(\sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [0,1]}} \frac{d(X_s, X_t)^p}{|t-s|} \geq \frac{R}{4C^p} \right).$$

To do so, note that if the interval $[s, t] \subseteq [0, 1]$ satisfies $|t - s| < h$, it must be contained in at least one interval of the form

$$\big[(k-1)h, (k+1)h\big] \qquad \text{for some } k = 1, \ldots, \lceil h^{-1} \rceil.$$

Therefore,

(30)
$$\mathbb{P}^x \left(\sup_{\substack{s \neq t, |t-s| \leq h, \\ [s,t] \subset [0,1]}} \frac{d(X_s, X_t)^p}{|t-s|} \geq \frac{R}{4C^p} \right)$$

$$\leq \sum_{k=1}^{\lceil h^{-1} \rceil} \mathbb{P}^x \left(\sup_{[s,t] \subseteq [(k-1)h, (k+1)h]} \frac{d(X_s, X_t)^p}{|t-s|} \geq \frac{R}{4C^p} \right).$$

We will now show that each term in this sum possesses the desired bound, that is, there exists a positive constant $c > 0$ such that

(31) $$\mathbb{P}^x \left(\sup_{[s,t] \subseteq [(k-1)h, (k+1)h]} \frac{d(X_s, X_t)^p}{|t-s|} \geq \frac{R}{4C^p} \right) \leq c \exp\left(-\frac{Rr^{2-p}}{c} \right).$$

Because there are only $\lceil h^{-1} \rceil \leq Rr^{-p}$ terms in the sum, it will follow that we can bound the left-hand side of (30) by

$$Rr^{-p} c \exp\left(-\frac{1}{c} Rr^{2-p} \right).$$

To prove (31), we exploit the scaling property in Remark 3.5 to see that

(32)
$$\sup_{[s,t] \subseteq [(k-1)h, (k+1)h]} \frac{d^{\mathcal{E}}(X_s, X_t)^p}{|t-s|} \qquad \text{under } \mathbb{P}^{\mathcal{E},x}$$

$$\overset{\mathcal{D}}{=} \sup_{[s,t] \subseteq [(k-1), (k+1)]} \frac{d^{h\mathcal{E}}(X_s, X_t)^p}{|t-s|} \cdot h^{p/2-1} \qquad \text{under } \mathbb{P}^{h\mathcal{E},x},$$

where $\overset{\mathcal{D}}{=}$ denotes equality in distribution, and we have again emphasized the dependency on $\mathcal{E}$ of $\mathbb{P}^x$ and the intrinsic metric d. We then conclude by applying (32)

with $h = 2^{p+2} R^{-1} r^p C^p$ to give

$$\sup_{y \in E} \mathbb{P}^{\mathcal{E}, y} \left(\sup_{[s,t] \subseteq [(k-1)h, (k+1)h]} \frac{d^{\mathcal{E}}(X_s, X_t)^p}{|t - s|} \geq \frac{R}{4C^p} \right)$$

$$= \sup_{y \in E} \mathbb{P}^{h\mathcal{E}, y} \left(\sup_{[s,t] \subseteq [(k-1), (k+1)]} \frac{d^{h\mathcal{E}}(X_s, X_t)^p}{|t - s|} \geq \frac{R^{p/2} r^{p - p^2/2}}{(2C)^{p^2/2}} \right)$$

$$\leq c_3 \exp\left(-\frac{R r^{2-p}}{c_3 2^p} \right).$$

The last step here results from applying Theorem 3.4 and noting that c_3 is independent of h as it only depends on the doubling and Poincaré constants associated with $h\mathcal{E}$. These constants coincide with those of $\mathcal{E}$ for each $h > 0$. $\quad\square$

6. Examples: $\mathfrak{g}$-valued Markovian rough paths. As we discussed in the Introduction, the motivation for this paper comes from estimates which arise naturally in the theory of rough paths. In this section, we specialise the results we have derived to this setting. The state space E is henceforth taken to be $\mathfrak{g} = \mathfrak{g}^N(\mathbb{R}^d)$, as presented in Section 2, and we construct our processes there. The Dirichlet form we work with is given by

$$
\begin{aligned}
\mathcal{E}(f, g) &= \sum_{i,j=1}^{d} \int_{\mathfrak{g}} a^{ij}(x) B_i f(x) B_j g(x) \mu(dx) \\
&= \sum_{i,j=1}^{d} \int_{\mathfrak{g}} a^{ij}(x) B_i f(x) B_j g(x) v(x) m(dx).
\end{aligned}
$$
(33)

In this formula, $a : \mathfrak{g} \to \mathcal{S}_d$ is a fixed measurable map into the space of $d \times d$ positive semi-definite matrices and $B_i : i = 1, \ldots, d$ denotes the canonical left-invariant vector fields which coincide with the standard basis vectors at the origin. The measure μ on $\mathfrak{g}$ is assumed to be absolutely continuous with respect to the Haar measure, and to possess a smooth positive density v.

Our natural regularity assumption on $\mathcal{E}$ is that it is sub-elliptic. This means that there exist constants $C > 0$ and $\varepsilon > 0$ such that for every open subset $U \subset \mathfrak{g}$ and every $f \in C_c^\infty(U)$

$$\|f\|_\varepsilon^2 \leq C[\mathcal{E}(f, f) + \|f\|_{L^2(\mu)}^2],$$
(34)

where, for $s > 0$, $\|f\|_s^2 := \int_U |\hat{f}(u)|^2 (1 + |u|^2)^s \mu(du)$ denotes the usual fractional Sobolev norm of order s and $\hat{f}$ is the Fourier transform of f. The authors of [23] have shown, for the case of smooth a, that $\mathcal{E}$ is sub-elliptic if and only if the balls with respect to the intrinsic metric d of $\mathcal{E}$ and the Euclidean metric $|\cdot|_{\text{Euc}}$ on $\mathbb{R}^m$,

where $m = \dim \mathfrak{g}^N(\mathbb{R}^d)$, are locally equivalent. More precisely, for some positive constants C and r_0 and for all $x \in E$, we have that

$$(35) \qquad B_d\big(x, C^{-1}r\big) \subset B_{|\cdot|_{\mathrm{Euc}}}(x, r) \subset B_d\big(x, Cr^\varepsilon\big) \qquad \text{for all } 0 < r < r_0.$$

The case of nonsmooth a is discussed in [12, 56], where it is also shown that Condition 3.1, our basic set of three conditions, is satisfied for this sub-elliptic class. We refer the reader to the important references [8, 12, 24, 37–39, 49] for further discussion of sub-elliptic operators. By applying our main theorem in this setting, we obtain the following as an immediate corollary.

COROLLARY 6.1. *Assume that $\mathcal{E}$ is the Dirichlet form satisfying the Fefferman–Phong sub-ellipticity condition* (34) *on $\mathfrak{g}$, which satisfies Condition* 3.1. *For $x \in \mathfrak{g}$, let $\mathbf{X}^x$ be the $\mathfrak{g}$-valued Markov process associated to $\mathcal{E}$ which starts from x. Then, for any $p > 2$, the accumulated local p-variation of $\mathbf{X}^x$ with respect to the intrinsic metric of $\mathcal{E}$ on $\mathfrak{g}$ has better than exponential tails. Using* (35), *the same conclusion holds for the accumulated local p-variation of $\mathbf{X}^x$ with respect to the metric induced by $|\cdot|_{\mathrm{Euc}}$ on $\mathfrak{g}$.*

An important special case of the above set of examples is the class of Dirichlet forms for which the matrix a in (33) is assumed to satisfy the upper bound

$$(36) \qquad\qquad \forall y \in \mathbb{R}^d : \sup_{x \in \mathfrak{g}} y^T a(x) y v(x) \le \Lambda |y|^2,$$

for some $\Lambda \ge 1$. This generalizes the class of processes studied by Friz and Victoir in [28], where $v \equiv 1$ and (36) is assumed to be complemented by a commensurate lower bound. The correct assumption is the Fefferman–Phong sub-ellipticity condition identified above. Note that (36) implies

$$d_{\mathrm{CC}}(x, y) \le \Lambda^{1/2} d(x, y) \qquad \text{for all } x, y \in \mathfrak{g}.$$

In particular, d_{CC} is locally controlled by d. As a result, we obtain the following corollary.

COROLLARY 6.2. *Assume that the Dirichlet form $\mathcal{E}$ in* (33) *satisfies the Fefferman–Phong condition* (34) *and that a is bounded above in the sense of* (36) *on $\mathfrak{g} = \mathfrak{g}^N$. For $x \in \mathfrak{g}$ let $\mathbf{X}^x$ be the $\mathfrak{g}$-valued Markov process associated to $\mathcal{E}$ which starts from x. Then, for any $p > 2$, the accumulated local p-variation of $\mathbf{X}^x$ with respect to the Carnot–Carathéodory metric on $\mathfrak{g}$ has better than exponential tails.*

REFERENCES

[1] BAKRY, D. and LEDOUX, M. (1996). Lévy–Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator. *Invent*. *Math*. **123** 259–281. MR1374200

[2] BASS, R. F. (1998). *Diffusions and Elliptic Operators*. Springer, New York. MR1483890

[3] BASS, R. F., HAMBLY, B. M. and LYONS, T. J. (2002). Extending the Wong–Zakai theorem to reversible Markov processes. *J. Eur. Math. Soc*. (*JEMS*) **4** 237–269. MR1924401

[4] BAUDOIN, F. and OUYANG, C. (2015). On small time asymptotics for rough differential equations driven by fractional Brownian motions. In *Large Deviations and Asymptotic Methods in Finance*. *Springer Proc*. *Math. Stat*. **110** 413–438. Springer, Cham. MR3375191

[5] BAUDOIN, F., OUYANG, C. and TINDEL, S. (2014). Upper bounds for the density of solutions to stochastic differential equations driven by fractional Brownian motions. *Ann*. *Inst. Henri Poincaré Probab. Stat*. **50** 111–135. MR3161525

[6] BAUDOIN, F., OUYANG, C. and ZHANG, X. (2015). Varadhan estimates for rough differential equations driven by fractional Brownian motions. *Stochastic Process*. *Appl*. **125** 634–652. MR3293297

[7] BAUDOIN, F., OUYANG, C. and ZHANG, X. (2016). Smoothing effect of rough differential equations driven by fractional Brownian motions. *Ann*. *Inst. Henri Poincaré Probab. Stat*. **52** 412–428. MR3449308

[8] BENDIKOV, A. and SALOFF-COSTE, L. (2005). On the regularity of sample paths of sub-elliptic diffusions on manifolds. *Osaka J. Math*. **42** 677–722. MR2166728

[9] BEN AROUS, G., GRĂDINARU, M. and LEDOUX, M. (1994). Hölder norms and the support theorem for diffusions. *Ann*. *Inst. Henri Poincaré Probab. Stat*. **30** 415–436. MR1288358

[10] BEURLING, A. and DENY, J. (1959). Dirichlet spaces. *Proc*. *Natl. Acad. Sci. USA* **45** 208–215. MR0106365

[11] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1987). *Regular Variation*. *Encyclopedia of Mathematics and Its Applications* **27**. Cambridge Univ. Press, Cambridge. MR0898871

[12] BIROLI, M. and MOSCO, U. (1995). A Saint-Venant type principle for Dirichlet forms on discontinuous media. *Ann*. *Mat. Pura Appl*. (4) **169** 125–181. MR1378473

[13] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent*. *Math*. **30** 207–216. MR0399402

[14] BURKHOLDER, D. L. (1966). Martingale transforms. *Ann*. *Math. Stat*. **37** 1494–1504. MR0208647

[15] CAPITAINE, M., HSU, E. P. and LEDOUX, M. (1997). Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces. *Electron*. *Commun. Probab*. **2** 71–81 (electronic). MR1484557

[16] CASS, T. and FRIZ, P. (2010). Densities for rough differential equations under Hörmander's condition. *Ann. of Math*. (2) **171** 2115–2141. MR2680405

[17] CASS, T., HAIRER, M., LITTERER, C. and TINDEL, S. (2015). Smoothness of the density for solutions to Gaussian rough differential equations. *Ann*. *Probab*. **43** 188–239. MR3298472

[18] CASS, T., LITTERER, C. and LYONS, T. (2013). Integrability and tail estimates for Gaussian rough differential equations. *Ann*. *Probab*. **41** 3026–3050. MR3112937

[19] CASS, T. and LYONS, T. (2015). Evolving communities with individual preferences. *Proc*. *Lond. Math. Soc*. (3) **110** 83–107. MR3299600

[20] CHEVYREV, I. and LYONS, T. (2015). Characteristic functions of measures on geometric rough paths. *Ann*. *Probab*. To appear.

[21] DAVIES, E. B. (1989). *Heat Kernels and Spectral Theory*. *Cambridge Tracts in Mathematics* **92**. Cambridge Univ. Press, Cambridge. MR0990239

[22] DYNKIN, E. B. (1947). Calculation of the coefficients in the Campbell–Hausdorff formula. *Dokl. Akad. Nauk SSSR* **57** 323–326. MR0021940

[23] FEFFERMAN, C. and PHONG, D. H. (1983). Subelliptic eigenvalue problems. In *Conference on Harmonic Analysis in Honor of Antoni Zygmund*, *Vol. I*, *II* (*Chicago*, *Ill.*, 1981). *Wadsworth Math. Ser.* 590–606. Wadsworth, Belmont, CA.

[24] FEFFERMAN, C. L. and SÁNCHEZ-CALLE, A. (1986). Fundamental solutions for second order subelliptic operators. *Ann. of Math.* (2) **124** 247–272. MR0855295

[25] FRIZ, P., LYONS, T. and STROOCK, D. (2006). Lévy's area under conditioning. *Ann. Inst. Henri Poincaré Probab. Stat.* **42** 89–101. MR2196973

[26] FRIZ, P. and RIEDEL, S. (2013). Integrability of (non-)linear rough differential equations and integrals. *Stoch. Anal. Appl.* **31** 336–358. MR3021493

[27] FRIZ, P. and RIEDEL, S. (2014). Convergence rates for the full Gaussian rough paths. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 154–194. MR3161527

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

[29] FRIZ, P. and VICTOIR, N. (2010). Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 369–413. MR2667703

[30] FRIZ, P. K. and HAIRER, M. (2014). *A Course on Rough Paths*. Springer, Cham. MR3289027

[31] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths*: *Theory and Applications*. *Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669

[32] FUKUSHIMA, M., ŌSHIMA, Y. and TAKEDA, M. (1994). *Dirichlet Forms and Symmetric Markov Processes*. de Gruyter, Berlin. MR1303354

[33] GROMOV, M. (1996). Carnot-Carathéodory spaces seen from within. In *Sub-Riemannian Geometry*. (A. Bellaiche and J. Risler, eds.) 79–323. Birkhäuser, Basel. MR1421823

[34] HAIRER, M. and PILLAI, N. S. (2013). Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths. *Ann. Probab.* **41** 2544–2598. MR3112925

[35] HAO, N. (2014). The decay rate of the expected signature of a Gaussian process. Personal communication.

[36] INAHAMA, Y. (2016). Short time kernel asymptotics for rough differential equation driven by fractional Brownian motion. *Electron. J. Probab.* **21** Paper No. 34, 29. MR3492938

[37] JERISON, D. (1986). The Poincaré inequality for vector fields satisfying Hörmander's condition. *Duke Math. J.* **53** 503–523. MR0850547

[38] JERISON, D. and SÁNCHEZ-CALLE, A. (1987). Subelliptic, second order differential operators. In *Complex Analysis*, *III* (*College Park*, *Md.*, 1985–86). *Lecture Notes in Math.* **1277** 46–77. Springer, Berlin. MR0922334

[39] JERISON, D. S. and SÁNCHEZ-CALLE, A. (1986). Estimates for the heat kernel for a sum of squares of vector fields. *Indiana Univ. Math. J.* **35** 835–854. MR0865430

[40] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York. MR1121940

[41] LYONS, T. (2004). Differential equations driven by rough paths (handwritten Saint-Flour notes).

[42] LYONS, T. and STOICA, L. (1999). The limits of stochastic integrals of differential forms. *Ann. Probab.* **27** 1–49. MR1681146

[43] LYONS, T. and ZEITOUNI, O. (1999). Conditional exponential moments for iterated Wiener integrals. *Ann. Probab.* **27** 1738–1749. MR1742886

[44] LYONS, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14** 215–310. MR1654527

[45] LYONS, T. J., CARUANA, M. and LÉVY, T. (2007). *Differential Equations Driven by Rough Paths*. Springer, Berlin. MR2314753

[46] LYONS, T. J. and ZHANG, T. S. (1994). Decomposition of Dirichlet processes and its application. *Ann. Probab.* **22** 494–524. MR1258888

[47] MA, Z. M. and RÖCKNER, M. (1992). *Introduction to the Theory of (Nonsymmetric) Dirichlet Forms*. Springer, Berlin. MR1214375

[48] MONTGOMERY, R. (2002). *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. Amer. Math. Soc., Providence, RI. MR1867362

[49] NAGEL, A., STEIN, E. M. and WAINGER, S. (1985). Balls and metrics defined by vector fields. I. Basic properties. *Acta Math.* **155** 103–147. MR0793239

[50] RIEDEL, S. (2014). Talagrand's transportation-cost inequality and applications to (rough) path spaces. Preprint. Available at arXiv:1403.2585.

[51] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales. Vol.* 1. Cambridge Univ. Press, Cambridge. MR1796539

[52] STRICHARTZ, R. S. (1987). The Campbell–Baker–Hausdorff–Dynkin formula and solutions of differential equations. *J. Funct. Anal.* **72** 320–345.

[53] STROOCK, D. W. (1988). Diffusion semigroups corresponding to uniformly elliptic divergence form operators. In *Séminaire de Probabilités, XXII* 316–347. Springer, Berlin. MR0960535

[54] STURM, K.-T. (1994). Analysis on local Dirichlet spaces. I. Recurrence, conservativeness and L^p-Liouville properties. *J. Reine Angew. Math.* **456** 173–196. MR1301456

[55] STURM, K.-T. (1995). Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. *Osaka J. Math.* **32** 275–312. MR1355744

[56] STURM, K. T. (1996). Analysis on local Dirichlet spaces. III. The parabolic Harnack inequality. *J. Math. Pures Appl.* (9) **75** 273–297. MR1387522

[57] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer, Heidelberg. MR3184689

DEPARTMENT OF MATHEMATICS
IMPERIAL COLLEGE LONDON
HUXLEY BUILDING, 180 QUEENS GATE
LONDON, SW7 2AZ
UNITED KINGDOM
E-MAIL: thomas.cass@imperial.ac.uk
        m.ogrodnik12@imperial.ac.uk