Tail estimates for Markovian rough paths

Thomas Cass and Marcel Ogrodnik

November 18, 2014

Abstract

We work in the context of Markovian rough paths associated to a class of uniformly subelliptic Dirichlet forms ([25]) and prove an almost-Gaussian tail-estimate for the accummulated local p-variation functional, which has been introduced and studied in [17]. We comment on the significance of these estimates to a range of currently-studied problems, including the recent results of Chevyrev and Lyons in [18].

1 Introduction

Lyons's rough path theory has allowed a pathwise interpreation to be given to stochastic differential equations of the form

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

where the vector fields $V = \left(V^1, ..., V^d\right)$ 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, e.g. 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 [26], [14], [30], [24], [31], [42] 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., [43]) and the Burkholder-Davis-Gundy inequalities (see, e.g., [12]) are prime examples of this principle in practice. The latter result in particular, allows one to control the moments of linear differential equations

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

where A is in $\text{Hom}\left(\mathbb{R}^d, \mathbb{R}^e\right)$ 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 [27])

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

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 [17]. The key idea is to use a slight sharpening of the estimate (1) to one of the form (see [17])

$$||\mathbf{Y}||_{p\text{-}var,[0,T]} \leq C \exp\left(\sup_{\substack{D=(t_i) \\ ||\mathbf{X}||_{p\text{-}var,[t_i,t_{i+1}]}\leq 1}} \sum_{i:t_i\in D} ||\mathbf{X}||^p_{p\text{-}var,[0,T]}\right) := C\exp\left[CM(\mathbf{X},[0,T])\right]. \tag{2}$$

The functional M is called the accumulated local p-variation. 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 [17] is the following tail estimate for Gaussian rough paths $\mathbf{X}$:

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

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

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

An integer-valued random variable defined by

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

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

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

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., [11], [1]), and the Cameron-Martin embedding theorem of [27].

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 an 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 [39], 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 [37]) has also exploited the forward-backward martingale decomposition in the construction of the Lévy-area. An alternative approach, which we will follow most closely in our presentation, was proposed in [25] and [27]. 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 Dirchlet form (see Section 3 for a review of this idea).

There are big obstacles to implementing the Gaussian approach of [17] 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], [13]) do not seem easy to implement here. As a consequence we have to re-think the whole strategy upon which [17] 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
$$||\mathbf{X}||_{p\text{-}var,[\sigma_n,\sigma_{n+1}]} \leq 1 \text{ for all } n = 0,1,2...$$

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, e.g. [38] and in support theorems, [35], [8], [22].

2. A natural choice of metric in the Markovian regime of this paper is the supremum-metric for rough paths, and this is likely to be so for other classes of random rough paths, too. 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 e.g. [49]. 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 [17] has been in Gaussian Hörmander theory to prove, for example, smoothness and other properties of the density for Gaussian RDEs (see, e.g., [30], [6], [7], [5], [15]). 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 accummumlated p-variation occur naturally and generically in rough path theory; see [23] for a range of examples. We therefore expect uses of our results to be widespread. In [16] 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, e.g. in studying interacting McKean-Vlasov-type RDEs.

Upon finishing this paper we became aware of the very interesting and recent paper [18], where the authors prove a criterion for the law of a geometric rough path to be determined by its expected signature. This criterion requires one to check that $N_{p\text{-var}}(\mathbf{X},[0,T])$ has an exponential tail, when $\mathbf{X}$ is a realisation of some probability measure on the space of geometric rough paths. One example they cite 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 5.3, applies to Lyons's and Chevyrev's problem. But our conditions impose no restriction on the domain of $\mathbf{X}$ and we prove a stronger tail-estimate. More precisely, we prove one which is almost-Gaussian in the sense that

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

and not just exponential. One immediate application of our results therefore is to broaden substantially the range of examples to which the work of [18] is 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 Markovian rough paths. In Section 4 we use large deviations techniques and exponential Tauberian theorems to prove that under the supremum metric the integer associated to the greedy partition of a Markovian rough path has a Gaussian tail. The main work is done in Section 5, where we prove a crucial bound on the accumulated local p-variation in terms of the aforementioned integer of the greedy partition and the accumulated p-variation of the Markovian rough path 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, i.e., the accumulated local p-variation of a Markovian rough path has almost-Gaussian tails.

Acknowledgement 1.1 *The research of the first-named author is supported by EPSRC grant EP/M00516X/1. The first-named author is also grateful to Bruce Driver for conversations related to this work.*

2 Rough paths

There are now many texts which outline the core content of rough path theory (e.g., [34], [36], [27] and [21]), 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

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

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

$$\tilde{T}:=\tilde{T}^N(V):=\left\{h\in T^N(V):g^0=1\right\}\text{ and }\tilde{\mathfrak{t}}:=\tilde{\mathfrak{t}}^N(V):=\left\{A\in T^N(V):A^0=0\right\}.$$

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 $\mathrm{Lie}\left(\tilde{T}\right)$, i.e. $\tilde{\mathfrak{t}}$ 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

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

are mutually inverse, and log defines a global chart on $\tilde{T}$. The map exp coincides with the Lie group exponential, i.e. 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 $* : \tilde{\mathfrak{t}} \times \tilde{\mathfrak{t}} \to \tilde{\mathfrak{t}}$ using the functions exp and log as follows

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

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 $\mathrm{Lie}(\tilde{\mathfrak{t}}, *) = \tilde{\mathfrak{t}}$ as a vector space. The Lie group exponential map $\mathrm{Lie}(\tilde{\mathfrak{t}}, *) \to (\tilde{\mathfrak{t}}, *)$ also equals the identity map on $\tilde{\mathfrak{t}}$, and the Campbell-Baker-Hausdorff formula (see [20], [44]) 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 $\mathfrak{g}^N := \mathfrak{g} = \mathrm{Lie}(V)$ 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 $(0, a, 0, ..., 0) \in \mathfrak{g}$. 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*

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

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*

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

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 $\|\mathbf{x}\|_{1-\mathrm{var};[0,T]}$. The truncated signature $S_N(x)$ of x is defined by

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

It is well-known (see [27]) that $\log S_N(x)_{0,\cdot}$ is a path which takes values in the group $(\mathfrak{g}, *)$. Any horizontal curve starting from the 0, the identity in $(\mathfrak{g}, *)$, can be realised as the unique solution to

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

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

A classical theorem of Chow (see, e.g., [29], [41]) 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

$$\|g\|_{CC} := \inf\left\{\|\mathbf{x}\|_{1-\text{var};[0,T]} : x \in C^{\text{1-var}}([0,T],V) \text{ and } S_N(x)_{0,T} = g\right\}. \tag{5}$$

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

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

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

$$\delta_r : \left(1, g^1, g^2, ..., g^N\right) \to \left(1, rg^1, r^2g^2, ..., r^Ng^N\right).$$

In finite dimensions it is a basic fact ([27]) 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 (5) by

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

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

$$\|\mathbf{x}\|_{\alpha-\text{Höl};[0,T]} := \sup_{\substack{[s,t]\subseteq[0,T],\\ s\neq t}} \frac{\|\mathbf{x}_{s,t}\|_{CC}}{|t-s|^\alpha}, \tag{6}$$

$$\|\mathbf{x}\|_{p-\text{var};[0,T]} := \left(\sup_{D=(t_j)} \sum_{j:t_j\in D} \|\mathbf{x}_{t_j,t_{j+1}}\|_{CC}^p\right)^{1/p}, \tag{7}$$

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*

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

We call $WG\Omega_p(V)$ the set of weakly[1] *geometric p-rough paths.*

[1] 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 perist with it for the sake of consistency with the literature.

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

The definitions (6) and (7) 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* $C^{p-var}([0,\infty),\mathfrak{g})$ *to be the subset of the continuous* $\mathfrak{g}$*-valued paths,* $C([0,\infty),\mathfrak{g})$ *as follows*

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

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

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

$$\omega_{\mathbf{x}}(s,t):=\|\mathbf{x}\|^p_{p-\text{var};[s,t]} \tag{8}$$

is a *control*; by this we mean it is a continuous, non-negative, 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 [27, p.80]).

3 Markovian rough paths

Throughout the rest of this paper we will work on the finite dimensional vector space $V=\mathbb{R}^d$, equipped with the Euclidean norm. Formulating the construction here allows one to exploit the rich analysis associated to self-adjoint subelliptic differential operators, and the corresponding probabilistic study of symmetric Markov processes. The most prominent references for our setting include [9], [45], [46], [47], [48], [28] and [40]. We will work with the Dirichlet form which for smooth compactly supported functions $f,g\in C_c^\infty(\mathfrak{g})$ is defined by

$$\mathcal{E}^a(f,g)=\sum_{i,j=1}^d\int_{\mathfrak{g}}a^{ij}(h)B_if(h)B_jg(h)\,dm(h)=:\int_{\mathfrak{g}}\Gamma^a(f,g)\,dm. \tag{9}$$

Here $B_i=B_{e_i}$ is the unique left-invariant vector field on $(\mathfrak{g},*)$ associated with the i^{th} standard basis vector $e_i\in\mathbb{R}^d\subseteq\text{Lie}\left(\mathbb{R}^d\right)$; m denotes the (bi-invariant) Haar measure on $(\mathfrak{g},*)$, which coincides with the Lebesgue measure on the vector space $\mathfrak{g}$; $a=\left(a^{ij}\right)_{i,j\in\{1,\ldots,d\}}$ is a measurable map from $\mathfrak{g}$ to $\mathcal{S}_d$, the space of $d\times d$ symmetric matrices; and, Γ^a is the so-called carré du champ operator. We will need to impose the following uniform upper and lower bounds on a.

Definition 3.1 *For* $\Lambda\geq 1$ *we define* $\Xi(\Lambda)$ *to be the class of such measurable maps* $a:\mathfrak{g}\to\mathcal{S}_d$ *having the property that*

$$\forall y\in\mathbb{R}^d:\Lambda^{-1}|y|^2\leq\sup_{x\in\mathfrak{g}}y^Ta(x)y\leq\Lambda|y|^2. \tag{10}$$

Remark 3.2 *When $a(\cdot) \equiv I_d$, the identity matrix, we write $\mathcal{E}, \Gamma$ etc. in place of $\mathcal{E}^a, \Gamma^a$.*

We can introduce the operator $\mathcal{L}^a : C_c^\infty(\mathfrak{g}) \to L^2(m)$ by

$$\mathcal{L}^a f := \sum_{i,j=1}^{d} B_j \left[a^{ij}(\cdot) B_i f(\cdot) \right](h),$$

where the right-hand-side is defined in $L^2(m)$ by weak integration. Integrating-by-parts in (9), and using the left-invariance of the B_j vector fields together with the right-invariance of m ensures that

$$\mathcal{E}^a(f,g) = -\langle \mathcal{L}^a f, g \rangle_{L^2(m)}.$$

The operator $\mathcal{L}^a : C_c^\infty(\mathfrak{g}) \to L^2(m)$ is then symmetric, but not self-adjoint since $\mathcal{E}^a(f) := \mathcal{E}^a(f,f)$ is defined a priori only on $C_c^\infty(\mathfrak{g}) \subset L^2(m)$. Nevertheless, the (non-negative) bilinear form $\mathcal{E}^a(f,g)$ has the property of being closable, and is also strongly local and strongly regular in the sense of [48, Corollary 4.2] (see also [27]). It is therefore possible to construct a self-adjoint extension using the classical Friedrichs procedure (see for instance [33]). The idea is first to find a suitable domain on which to extend $\mathcal{E}^a$, and then out of this constructing a domain for extending $\mathcal{L}^a$.

In a little more detail, we first take the completion of $C_c^\infty(\mathfrak{g})$ with respect to the norm

$$||f||_{\mathcal{E}^a} := \left(||f||^2_{L^2(m)} + \mathcal{E}^a(f) \right)^{1/2}.$$

This completion may then be identified with a naturally embedded Hilbert subspace of $L^2(m)$ – the Dirichlet domain – which we denote $D(\mathcal{E}^a)$. This then allows for the construction of a dense linear subspace $D(\mathcal{L}^a)$ of $L^2(m)$ on which there exists a self-adjoint extension of $\mathcal{L}^a$, which is customarily also denoted $\mathcal{L}^a$. With the self-adjoint extension in hand, classical theory provides the route from $\mathcal{L}^a$ to a semi-group of contractions on $L^2(m)$, which we will denote $(P_t^a)_{t\geq 0}$, and thence to an associated Markov process. It is often insightful to make comparisons of the probabilistic features of different Markov processes produced by this procedure by comparing their Dirichlet forms. To this end, we recall the intrinsic distance associated with $\mathcal{E}^a$, which is defined by

$$d^a(x,y) = \sup \left\{ f(x) - f(y) : f \in \mathcal{F}_{\mathrm{loc}} \text{ and } f \text{ continuous}, \Gamma^a(f,f) \leq 1 \right\}, \tag{11}$$

where $\mathcal{F}_{\mathrm{loc}} := \left\{ f \in L^2(dm) : \Gamma^a(f,f) \in L^1_{\mathrm{loc}}(dm) \right\}$. The following lemma is proved in [27] based partly on results in [45, p.285]

Lemma 3.3 *Let $\Lambda \geq 1$, then $D(\mathcal{E}^a) = D(\mathcal{E})$ for every $a \in \Xi(\Lambda)$ and*

$$\frac{1}{\Lambda} \mathcal{E}(f) \leq \mathcal{E}^a(f) \leq \Lambda \mathcal{E}(f), \text{ for all } f \in D(\mathcal{E}).$$

The intrinsic distance d associated with $\mathcal{E}$ coincides with the Carnot Carathéodory metric on $\mathfrak{g}$, and furthermore d and d^a are Lipschitz equivalent

This semi-group $(P_t^a)_{t\geq 0}$ referred to above is easily seen by Sobolev estimates (see, e.g., [19]) to admit a kernel representation of the form

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

The heat kernel p^a can be shown to satisfy the following Aronson-type estimate (see [48, Corollary 4.2]).

Theorem 3.4 *Let $a \in \Xi(\Lambda)$. The heat kernel p^a associated with the Dirichlet form $\mathcal{E}^a$ satisfies, for $\epsilon > 0$ fixed,*

$$p^a(t,x,y) \leq \frac{C}{\sqrt{t^{\dim_H \mathfrak{g}}}} \exp\left(-\frac{d^a(x,y)^2}{(4+\epsilon)t}\right). \tag{12}$$

for some constant $C = C(\epsilon, \Lambda)$. Here $\dim_H \mathfrak{g}$ denotes the so called homogeneous dimension of $\mathfrak{g}$.

Lemma 3.3 allows us to compare these transition densities for different a very effectively. In particular, it is immediate from the Lipschitz equivalence of d and d^a that we have the following result.

Corollary 3.5 *Let $a \in \Xi(\Lambda)$ and $\epsilon > 0$ fixed, then the heat kernel p^a satisfies*

$$p^a(t,x,y) \leq \frac{C}{\sqrt{t^{\dim_H \mathfrak{g}}}} \exp\left(-\frac{d(x,y)^2}{(4+\epsilon)\Lambda t}\right).$$

with constant $C = C(\epsilon, \Lambda)$.

The heat kernel p^a allows for a consistent family of finite-dimensional distributions and thus determines a $\mathfrak{g}$-valued (strong) Markov process $(\mathbf{X}_t^{a,x})_{t\geq 0}$ with $a \in \Xi(\Lambda)$ and $\mathbf{X}_0^{a,x} = x \in \mathfrak{g}$. Using Kolmogorov's criterion (see [25, Theorem 13]) it can be shown that $\mathbf{X}^{a,x}$ has a version with continuous sample paths, i.e. $\mathbf{X}^{a,x} \in C^{1/p-\text{Höl}}([0,\infty),\mathfrak{g})$. In fact, much more can be shown; the following theorem is an assembly of results from [25, Theorem 13] which we will need subsequently.

Theorem 3.6 *Let $\Lambda \geq 1$ and suppose $N \geq 2$ is a natural number. Assume $a \in \Xi(\Lambda)$ and $x \in \mathfrak{g} = \mathfrak{g}^N$. For any $p > 2$ there exists a version $\mathbf{X}^{a,x}$ which takes values in $C^{1/p-H\ddot{o}l}([0,\infty),\mathfrak{g})$. In particular when $p \in [N, N+1)$ we have that $\mathbf{X}^{a,x}$ restricted to $[0,T]$ is in $WG\Omega_p([0,T], V)$. Letting $\mathbb{P}^{a,x}$ denote the probability measure on $C([0,\infty),\mathfrak{g})$ given by the law of $\mathbf{X}^{a,x}$. Then there exists a finite constant $C = C(\alpha, \Lambda, T, N)$ such that*

$$\sup_{x\in\mathfrak{g}} \mathbb{P}^{a,x}\left(\sup_{[s,t]\subseteq[0,T]} \frac{d^a(\mathbf{X}_s, \mathbf{X}_t)}{|t-s|^\alpha} > r\right) \leq C \exp\left(-\frac{r^2}{C}\right),$$

wherein $\mathbf{X}_s(\omega) = \omega(s)$ *for* $s \geq 0$ *denotes the evaluation maps on* $C([0,\infty),\mathfrak{g})$*. Moreover* $\mathbf{X}^{a,x}$ *satisfies the following weak scaling property for all* $r > 0$

$$\left(\mathbf{X}_t^{a^r,x} : t \geq 0\right) \stackrel{\mathcal{D}}{=} \left(\delta_r \mathbf{X}_{tr^{-2}}^{a,\delta_{r^{-1}}(x)} : t \geq 0\right), \tag{13}$$

where δ_r *denotes the natural scaling operation on* $\mathfrak{g}$*,* $\stackrel{\mathcal{D}}{=}$ *denotes equality in distribution and* $a^r(x) := a(\delta_{r^{-1}}x) \in \Xi(\Lambda)$*.*

4 A large deviations result

Let d_{CC} be the Carnot-Carathéodory metric on $\mathfrak{g}$ that was introduced in Section 2. Given any $\mathbf{x}$ in $C([0,\infty),\mathfrak{g})$ we can define inductively a non-decreasing sequence $(\sigma_n)_{n=0}^{\infty} = (\sigma_n(\mathbf{x}))_{n=0}^{\infty}$ by setting $\sigma_0 = 0$, and then for $n \in \mathbb{N}$

$$\sigma_n := \inf\left\{t \geq \sigma_{n-1} : d_{CC}\left(\mathbf{x}_{\sigma_{n-1}}, \mathbf{x}_t\right) \geq 1\right\}. \tag{14}$$

Definition 4.1 *For any* $T \geq 0$ *we define the functional* $N_0(\cdot) = N_0(\cdot,[0,T]) : C([0,\infty),\mathfrak{g}) \to \mathbb{N} \cup \{0\}$ *by*

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

Remark 4.2 *Note that* $N_0(\mathbf{x},[0,T]) < \infty$ *implies that the set*

$$\{\sigma_j : j = 0,1,...,N_0(\mathbf{x},[0,T])\} \cup \{T\}$$

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 variable $N_0(\mathbf{X}^{a,x},[0,T])$, when $\mathbf{X}^{a,x}$ is a Markovian rough path of the type described in Section 3. Our approach will be motivated by the following well-known example.

Example 4.3 (Brownian motion) *Let* $B = (B_t)_{t\geq 0}$ *a one-dimensional standard Brownian motion on some probability space* $(\Omega,\mathcal{F},\mathbb{P})$*. In this setting, the sequence in (14) is given by*

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

It is a classical result (see, e.g., [32]) that the Laplace transform of $\sigma := \sigma_1$ *satisfies*

$$\mathbb{E}\left[e^{-\lambda\sigma}\right] = \cosh\left(\sqrt{2\lambda}\right)^{-1} \leq 2e^{-\sqrt{2\lambda}}. \tag{15}$$

If we let $\xi_k := \sigma_k - \sigma_{k-1}$ *for* $k = 1,...,n$ *and note that* $\{\xi_k : k = 1,...,n\}$ *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* $\theta > 0$

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

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 $N_0(\mathbf{X}^{a,x},[0,T])$. What is important – as we will show – is not to have a closed-form expression as in (15), but to have an upper bound controlling its asymptotic behaviour as $\lambda\to\infty$.

4.1 Tails for $N_0\left(\mathbf{X}^{a,x},[0,T]\right)$

From now on we fix $\Lambda\geq 1$, $N\geq 2$ and let $\mathfrak{g}=\mathfrak{g}^N$. We will adopt the notation of Theorem 3.6, i.e. for $a\in\Xi(\Lambda)$ and $x\in\mathfrak{g}$, $\mathbf{X}^{a,x}\in C([0,\infty),\mathfrak{g})$ will be the strong Markov process associated with $\mathcal{E}^a$, $\mathbb{P}^{a,x}$ will be the law of on $C\left([0,\infty),\mathfrak{g}\right)$ and $\mathbb{E}^{a,x}$ the corresponding expectation. For $t>0$ we continue to denote the evaluation maps by

$$\mathbf{X}_t:C([0,\infty),\mathfrak{g})\to C([0,\infty),\mathfrak{g}),\ \ \mathbf{X}_t\left(\omega\right)=\omega\left(t\right)\ \text{ for } t\geq 0.$$

We introduce the following notation for the Laplace transform of $\sigma=:\sigma_1\left(\mathbf{X}\right)$ under $\mathbb{P}^{a,x}$:

$$M(\lambda;a,x):=\mathbb{E}^{a,x}\left[e^{-\lambda\sigma}\right]=\int_{C([0,\infty),\mathfrak{g})}e^{-\lambda\sigma(\omega)}\mathbb{P}^{a,x}\left(d\omega\right),\tag{16}$$

ready for stating a version of De Bruijn's exponential Tauberian theorem. This result, in a precise way, relates the asymptotic behaviour of the log Laplace transform, $\log M(\lambda;a,x)$ as $\lambda\to\infty$, and the log short-time probability $\log\mathbb{P}^{a,x}\left(\sigma\leq t\right)$ as $t\to 0+$.

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

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

Proof. This is an immediate consequence of applying Theorem 4.12.9 in [10], making the choice $B=\frac{c^2}{4}$ and $\phi(\lambda)=\frac{1}{\lambda}$ in the notation of that theorem. ■

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 [10] for further discussion of results of this type.

Lemma 4.5 *Assume that $\Lambda\geq 1$, $a\in\Xi(\Lambda)$ and $x\in\mathfrak{g}$, and let $\mathbb{P}^{a,x}$ denote the law of the rough path $\mathbf{X}^{a,x}$ on $C\left([0,\infty),\mathfrak{g}\right)$. Let $\sigma=\sigma_1$ be the stopping time defined in (14) and suppose there exists $c>0$ for which*

$$\limsup_{t\to 0+}t\sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}}\log\mathbb{P}^{a,x}(\sigma\leq t)\leq -c.\tag{17}$$

Then

$$\limsup_{\lambda\to\infty}\lambda^{-\frac{1}{2}}\sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}}\log M(\lambda;a,x)\leq -2\sqrt{c}.$$

Proof. Set $\mathbb{P}^{a,x}(\sigma \leq t) =: \mu_{a,x}(t)$. And note that it is sufficient to show that the assumption (17) implies

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

Observe that for $\xi, \lambda > 0$,

$$\begin{aligned} M\left(\frac{1}{\lambda^2}; a, x\right) &= \int_0^{\frac{\lambda}{\xi}} \exp\left(-\frac{t}{\lambda^2}\right) d\mu_{a,x}(t) + \int_{\frac{\lambda}{\xi}}^{\infty} \exp\left(-\frac{t}{\lambda^2}\right) d\mu_{a,x}(t) \\ &\leq \mu_{a,x}\left(\frac{\lambda}{\xi}\right) + \exp\left(-\frac{1}{\xi\lambda}\right). \end{aligned} \tag{18}$$

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

$$\limsup_{n\to\infty} \frac{1}{n} \log(a_n + b_n) \leq \limsup_{n\to\infty} \frac{1}{n} \log a_n + \limsup_{n\to\infty} \frac{1}{n} \log b_n.$$

In the setting of (18) this gives

$$\begin{aligned} \limsup_{\lambda\to 0+} \lambda \sup_{a,x} \log M\left(\frac{1}{\lambda^2}; a, x\right) &\leq \limsup_{\lambda\to 0+} \lambda \sup_{a,x} \log\left(\mu_{a,x}\left(\frac{\lambda}{\xi}\right)\right) - \frac{1}{\xi} \\ &\leq -\xi c - \frac{1}{\xi}, \end{aligned}$$

where the last line uses the hypothesis (17). Because the function $(0, \infty) \ni \xi \mapsto -(\xi c + \frac{1}{\xi})$ attains its global maximum $-2\sqrt{c}$ at $\xi^\star = c^{-\frac{1}{2}}$, we obtain

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

which completes the proof. ■

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

Lemma 4.6 *Denote by $\mathbb{P}^{a,x}$ 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 d, Λ and N, such that for all $t \in (0, 1]$*

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

As a consequence of this lemma we see that

$$\limsup_{t\to 0+} t \sup_{a \in \Xi(\Lambda)} \sup_{x \in \mathfrak{g}} \log \mathbb{P}^{a,x}(\sigma \leq t) = -c_2 < 0,$$

which allows us to apply lemma 4.5 and immediately deduce the following corollary.

Corollary 4.7 *Let $M(\lambda; a, x)$ denote the Laplace transform (16) of the stopping time σ. Then under the condition of lemma 4.6 we have*

$$\limsup_{\lambda\to\infty} \sqrt{\lambda}^{-1} \sup_{a\in\Xi(\Lambda)} \sup_{x\in\mathfrak{g}} \log M(\lambda; a, x) \leq -2\sqrt{c_2}.$$

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 Corollary 3.5 we will adapt the proof for the class of Markovian rough paths introduced earlier. First we note that

$$\mathbb{P}^{a,x}(\sigma \leq t) \leq \mathbb{P}^{a,x}\left(\sigma \leq t, d_{CC}(\mathbf{X}_t, x) < \frac{1}{2}\right) + \mathbb{P}^{a,x}\left(d_{CC}(\mathbf{X}_t, x) \geq \frac{1}{2}\right).$$

For $g \in \mathfrak{g}$ and $\delta > 0$ we let $B(g, \delta)$ denote the open d_{CC}-ball of radius δ centred at g. Then using Corollary 3.5 (with fixed $\epsilon > 0$), we see that the second term satisfies

$$\begin{aligned}
\mathbb{P}^{a,x}\left(d_{CC}(\mathbf{X}_t, x) \geq \frac{1}{2}\right) &= \int_{B\left(x,\frac{1}{2}\right)^c} p^a(t, x, y)\, dy \\
&\leq \int_{\frac{1}{2}}^{\infty} \frac{c_1}{\sqrt{t^{\dim_H \mathfrak{g}}}} \exp\left(-c_2\frac{r^2}{t}\right) r^{\dim_H \mathfrak{g}-1}\, dr \\
&= \int_{\frac{1}{2\sqrt{t}}}^{\infty} c_1 v^{\dim_H \mathfrak{g}-1} \exp\left(-c_2 v^2\right)\, dv \\
&\leq c_3 e^{-\frac{c_4}{t}},
\end{aligned} \tag{19}$$

where the constants c_3 and c_4 depend only on d, Λ and N. For the first term, observe that

$$\begin{aligned}
\mathbb{P}^{a,x}\left(\sigma_1 \leq t, d_{CC}(\mathbf{X}_t, x) < \frac{1}{2}\right) &\leq \int_0^t \mathbb{P}^{a,x}\left(\sigma_1 \in ds, d_{CC}(\mathbf{X}_t, \mathbf{X}_{\sigma_1}) \geq \frac{1}{2}\right) \\
&= \int_0^t \mathbb{E}^{a,x}\left[1_{\{\sigma\in ds\}}\mathbb{P}^{a,\mathbf{X}_{\sigma_1}}\left(d_{CC}(\mathbf{X}_{t-\sigma_1}, \mathbf{X}_0) \geq \frac{1}{2}\right)\right] \\
&\leq \int_0^t \mathbb{E}^{a,x}\left[1_{\{\sigma\in ds\}}\mathbb{P}^{a,\mathbf{X}_s}\left(d_{CC}(\mathbf{X}_{t-s}, \mathbf{X}_0)\right) \geq \frac{1}{2}\right].
\end{aligned}$$

By the same argument as in (19), we know there exist constants c_5 and c_6 (which, again, depend only on d, Λ and N) such that

$$\sup_{r\leq t} \mathbb{P}^{a,x}\left(d_{CC}(\mathbf{X}_r, \mathbf{X}_0) \geq \frac{1}{2}\right) \leq c_5 e^{-\frac{c_6}{t}}.$$

Together these bounds imply the desired result ■

We can now prove the needed tail estimates for the law of the random variable $N_0(\mathbf{X}, [0, T])$ under $\mathbb{P}^{a,x}$.

Proposition 4.8 *Let $\Lambda \geq 1$, $a \in \Xi(\Lambda)$. Assume that $\mathbf{X}^{a,x}$ is the $\mathfrak{g}-$valued Markov process, defined on some probability space, associated with the Dirichlet form (9). Let $\mathbb{P}^{a,x}$ be the (Borel) probability measure on $C([0,\infty),\mathfrak{g})$. Then the random variable $N_0(\cdot,[0,T]) : C([0,\infty),\mathfrak{g}) \to \mathbb{N}\cup\{0\}$ in Definition 4.1 has a Gaussian tail under $\mathbb{P}^{a,x}$; i.e. there exist positive constants c_1 and c_2, which do not depend on n, such that*

$$\mathbb{P}^{a,x}(N_0(\mathbf{X},[0,T]) \geq n) \leq c_1 \exp\left(-c_2 n^2\right). \tag{20}$$

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

$$\mathbb{P}^{a,x}(N_0(\mathbf{X},[0,T]) \geq n) \leq e^{\lambda T}\mathbb{E}^{a,x}\left[e^{-\lambda\sum_{k=1}^n \xi_k}\right]. \tag{21}$$

By definition $M(\lambda;a,x) = \mathbb{E}^{a,x}\left[e^{-\lambda\xi_1}\right] = \mathbb{E}^{a,x}\left[e^{-\lambda\sigma_1}\right]$, combining the Strong Markov Property at the stopping time σ_{n-1} with an easy induction yields the estimate

$$\begin{aligned}\mathbb{E}^{a,x}\left[e^{-\lambda\sum_{k=1}^n \xi_k}\right] &= \mathbb{E}^{a,x}\left[e^{-\lambda\sum_{k=1}^{n-1}\xi_k}\mathbb{E}^{a,\mathbf{X}_{\sigma_{n-1}}}\left[e^{-\lambda\sigma_1}\right]\right] \\ &\leq \mathbb{E}^{a,x}\left[e^{-\lambda\sum_{k=1}^{n-1}\xi_k}\right]\sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}} M(\lambda;a,x) \\ &\leq \sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}} M(\lambda;a,x)^n\end{aligned} \tag{22}$$

Corollary 4.7 shows the existence of $c > 0$ and $\eta > 0$ such that for all $\lambda \geq \eta$ we have

$$\sqrt{\lambda}^{-1}\sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}}\log M(\lambda;a,x) \leq -c,$$

which is tantamount to

$$\sup_{a\in\Xi(\Lambda)}\sup_{x\in\mathfrak{g}} M(\lambda;a,x) \leq \exp\left(-c\sqrt{\lambda}\right).$$

Combining this with (22) and (21) gives, for all $\lambda \geq \eta$,

$$\mathbb{P}^{a,x}(N_0(\mathbf{X},[0,T]) \geq n) \leq \exp\left(\lambda T - cn\sqrt{\lambda}\right).$$

Because the right hand side is minimized by the choice $\lambda = T^{-2}4^{-1}c^2n^2$, we have for all $n \geq \frac{2T\sqrt{\eta}}{c}$

$$\mathbb{P}^{a,x}(N_0(\mathbf{X},[0,T]) \geq n) \leq \exp\left(-\frac{c^2n^2}{4T}\right).$$

The conclusion (20), that $N_0(\mathbf{X},[0,T])$ has a Gaussian tail under $\mathbb{P}^{a,x}$, follows at once. ■

5 Tail estimates for the accumulated local p-variation

The law of the sub-elliptic Markov process $\mathbb{P}^{a,x}$ constructed in Section 3 is, for any $p > 2$ and $T \geq 0$, supported in $C^{1/p-\text{Höl}}([0,T],\mathfrak{g}) \subset C^{p-\text{var}}([0,T],\mathfrak{g}) \subset C([0,T],\mathfrak{g})$. 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. [17])

Definition 5.1 (accumulated local p-variation) *Let $p \geq 1$. We define the accumulated local p-variation to be the function $M(\cdot,[0,T]) = M(\cdot) : C^{p-var}([0,\infty),\mathfrak{g}) \to \mathbb{R}_{\geq 0}$ by*

$$M(\mathbf{x},[0,T]) := \sup_{\substack{D=(t_i) \\ \omega_{\mathbf{x}}(t_i,t_{i+1})\leq 1}} \sum_i \omega_{\mathbf{x}}(t_i,t_{i+1}),$$

where $\omega_{\mathbf{x}}(s,t) \equiv ||\mathbf{x}||^p_{p\text{-}var;[s,t]}$ is the control induced by $\mathbf{x}$, and the supremum is taken over all partitions D of the interval $[0,T]$ such that $\omega_{\mathbf{x}}$, when evaluated between two consecutive points in D, is bounded by one.

We will now show that the accumulated local p-variation of $\mathbf{x}$ over $[0,T]$ can be bounded by the sum of $N_0(\mathbf{x},[0,T])$ and the accumulated p-variation between the times associated with the sequence σ_i, $i = 0,1,...,N_0(\mathbf{x},[0,T])$.

Lemma 5.2 *Let $p \geq 1$, assume $\mathbf{x} \in C^{p-var}([0,\infty),\mathfrak{g})$ and suppose $\omega_{\mathbf{x}}$ is the control induced by $\mathbf{x}$. Let the sequence $(\sigma_n(\mathbf{x}))_{n=0}^{\infty} = (\sigma_n)_{n=0}^{\infty}$ and the non-negative integer $N_0(\mathbf{x},[0,T])$ be given as in (14) and Definition 4.1, respectively. Then we can bound $M(\mathbf{x},[0,T])$, the accumulated local $p-$variation, using the following estimate*

$$M(\mathbf{x},[0,T]) \leq N_0(\mathbf{x},[0,T]) + \sum_{j=1}^{N_0(\mathbf{x},[0,T])} \omega_{\mathbf{x}}(\sigma_{j-1},\sigma_j) + \omega_{\mathbf{x}}\left(\sigma_{N_0(\mathbf{x},[0,T])},T\right). \tag{23}$$

Proof. To deal with the end point in a notationally efficient way, we redefine $\sigma_{N_0(\mathbf{x},[0,T])+1}$ so that $\sigma_{N_0(\mathbf{x},[0,T])+1} = T$ (Note: this is, strictly, an abuse of notation in light of the definition in formula (14)). Suppose then that $D = \{t_i : i = 0,1,...,n\}$ is an arbitrary partition of $[0,T]$, such that any two consecutive points $s < t$ in D satisfy $\omega_{\mathbf{x}}(s,t) \leq 1$. Let

$$\left\{t_{k_j} : j = 0,1,...,N_0(\mathbf{x},[0,T]), N_0(\mathbf{x},[0,T])+1\right\} \subset D$$

be a sub-partition where the integers $k_j \in \{0,1,...,n\}$ are defined by setting $k_0 = 0$, $k_{N_0(\mathbf{x},[0,T])+1} = n$, and

$$k_j := \max\{l \in \mathbb{N} : t_l < \sigma_j\} \tag{24}$$

for $j = 1,...,N_0(\mathbf{x},[0,T])$. The crucial step in proving the estimate (23) is the following

$$\sum_{i=1}^{n} \omega_{\mathbf{x}}(t_{i-1},t_i) = \sum_{i=0}^{k_1-1} \omega_{\mathbf{x}}(t_i,t_{i+1}) + \sum_{j=1}^{N_0(\mathbf{x},[0,T])} \left(\omega_{\mathbf{x}}\left(t_{k_j},t_{k_j+1}\right) + \sum_{i=k_j+1}^{k_{j+1}-1} \omega_{\mathbf{x}}(t_i,t_{i+1})\right). \tag{25}$$

To further control this term we make two observations. First, for all $j = 1, ..., N_0(\mathbf{x}, [0,T])$, we have $\omega_{\mathbf{x}}\left(t_{k_j}, t_{k_j+1}\right) \leq 1$ by the definition of D. Second, the super-additivity of $\omega_{\mathbf{x}}$ and the definition (24) of k_j yields

$$\sum_{i=k_j+1}^{k_{j+1}-1} \omega_{\mathbf{x}}(t_i, t_{i+1}) \leq \omega_{\mathbf{x}}(\sigma_j, \sigma_{j+1}), \text{ for } j = 0, 1, ..., N_0(\mathbf{x}, [0,T]).$$

Using these two observation in (25) results in the bound

$$\sum_{i:t_i \in D} \omega_{\mathbf{x}}(t_i, t_{i+1}) = \sum_{i=1}^{n} \omega_{\mathbf{x}}(t_i, t_{i+1}) \leq N_0(\mathbf{x}, [0,T]) + \sum_{j=0}^{N_0(\mathbf{x},[0,T])} \omega_{\mathbf{x}}(\sigma_j, \sigma_{j+1}).$$

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 (23) then follows immediately. ■

We are now ready to prove the main result.

Theorem 5.3 *Let $\Lambda \geq 1$, $a \in \Xi(\Lambda)$. Assume that $\mathbf{X}^{a,x}$ is the $\mathfrak{g}-$valued Markov process, defined on some probability space, associated with the Dirichlet form (9). Let $\mathbb{P}^{a,x}$ be the (Borel) probability measure on $C([0,\infty), \mathfrak{g})$ under which the coordinate process $\mathbf{X}$ has the same law as $\mathbf{X}^{a,x}$. Assume $p > 2$, and write $M(\cdot, [0,T])$ for the accumulated local $p-$variation given in Definition 5.1. Since $\mathbb{P}^{a,x}$ is supported in $C^{1/p-H\ddot{o}l}([0,\infty), \mathfrak{g})$, $M(\mathbf{X}, [0,T])$ is defined $\mathbb{P}^{a,x}$- almost surely. Moreover there exist constants c_1 and c_2, which depend only on Λ, p, N and T, such that*

$$\mathbb{P}^{a,x}(M(\mathbf{X}; [0,T]) > R) \leq c_1 \exp\left(-c_2 R^{1+2/p}\right) \tag{26}$$

for all $R > 0$.

Proof. We will assume that $T = 1$ and write $M(\mathbf{X})$ and $N_0(\mathbf{X})$ in lieu of $M(\mathbf{X}; [0,T])$ and $N_0(\mathbf{X}; [0,T])$, respectively. The assumption $T = 1$ involves no loss of generality because of the scaling property (13). The assertion that $\mathbb{P}^{a,x}$ is supported in $C^{1/p-\text{Höl}}([0,\infty), \mathfrak{g})$ is proved in [27], see also our presentation in Theorem 3.6.

To prove the main estimate (26) we first note it is a straight-forward consequence of lemma 5.2 that

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

where again $N_0(\mathbf{X})$ and the sequence $(\sigma_n)_{n=0}^{N_0(\mathbf{X})}$ are as given as in (14) with $T = 1$, and we redefine $\sigma_{N_0(\mathbf{x})+1}$ to equal $T = 1$. A simple estimate then gives

$$\mathbb{P}^{a,x}(M(\mathbf{X}) > R) \leq \mathbb{P}^{a,x}\left(N_0(\mathbf{X}) > \frac{R}{2}\right) + \mathbb{P}^{a,x}\left(\sum_{j=0}^{N_0(\mathbf{X})} \omega_{\mathbf{X}}(\sigma_j, \sigma_{j+1}) > \frac{R}{2}\right). \tag{27}$$

It follows from Proposition 4.8 that $N_0(\mathbf{X})$ has a Gaussian tail under $\mathbb{P}^{a,x}$, and so it remains to focus on the second term on the right in (27). To this end, first note the following elementary inequality

$$\omega_{\mathbf{X}}(\sigma_i, \sigma_{i+1}) \leq \|\mathbf{X}\|^p_{1/p-\text{Höl};[\sigma_i,\sigma_{i+1}]}(\sigma_{i+1} - \sigma_i).$$

Then for any $c > 0$ we notice that

$$\begin{aligned}\|\mathbf{X}\|^p_{1/p-\text{Höl},[\sigma_i,\sigma_{i+1}]} &\leq \sup_{\substack{s\neq t, |t-s|\leq c, \\ [s,t]\subset[\sigma_i,\sigma_{i+1}]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} + \sup_{\substack{s\neq t, |t-s|>c, \\ [s,t]\subset[\sigma_i,\sigma_{i+1}]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} \\ &\leq \sup_{\substack{s\neq t, |t-s|\leq c, \\ [s,t]\subset[\sigma_i,\sigma_{i+1}]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} + 2^p c^{-1},\end{aligned}$$

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

$$\begin{aligned}\sum_{i=0}^{N_0(\mathbf{X})} \omega_{\mathbf{X}}(\sigma_i, \sigma_{i+1}) &\leq \sum_{i=0}^{N_0(\mathbf{X})} \left[\sup_{\substack{s\neq t, |t-s|\leq c, \\ [s,t]\subset[\sigma_i,\sigma_{i+1}]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} (\sigma_{i+1} - \sigma_i) \right] + 2^p c^{-1} \\ &\leq \sup_{\substack{s\neq t, |t-s|\leq c, \\ [s,t]\subset[0,1]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} + 2^p c^{-1}.\end{aligned}$$

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

$$\sum_{i=0}^{N_0(\mathbf{X})} \omega_{\mathbf{X}}(\sigma_i, \sigma_{i+1}) \leq \sup_{\substack{s\neq t, |t-s|\leq h, \\ [s,t]\subset[0,1]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} + \frac{R}{4}$$

and consequently it suffices to bound

$$\mathbb{P}^{a,x}\left(\sup_{\substack{s\neq t, |t-s|\leq h, \\ [s,t]\subset[0,1]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} \geq \frac{R}{4} \right).$$

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

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

Therefore,

$$\mathbb{P}^{a,x}\left(\sup_{\substack{s\neq t, |t-s|\leq h, \\ [s,t]\subset[0,1]}} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} \geq \frac{R}{4} \right) \leq \sum_{k=1}^{\lceil h^{-1} \rceil} \mathbb{P}^{a,x}\left(\sup_{[s,t]\subseteq[(k-1)h,(k+1)h]} \frac{\|\mathbf{X}_{s,t}\|^p_{CC}}{|t-s|} \geq \frac{R}{4} \right). \tag{28}$$

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

$$\mathbb{P}^{a,x}\left(\sup_{[s,t]\subseteq[(k-1)h,(k+1)h]}\frac{\|\mathbf{X}_{s,t}\|_{CC}^p}{|t-s|}\geq\frac{R}{4}\right)\leq c\exp\left(-\frac{1}{c}R^{1+2/p}\right). \tag{29}$$

Because there are only $\lceil h^{-1}\rceil \leq R$ terms in the sum, it will follow that we can bound the left hand side of (28) by

$$RC\exp\left(-\frac{1}{C}R^{1+2/p}\right)\leq c_1\exp\left(-c_2R^{1+2/p}\right)$$

for some appropriately chosen constants c_1 and c_2 which do not depend on R. To prove (29) we exploit the scaling property (13) with $r := h^{-1/2} = 2^{-1-p/2}R^{1/2}$ and the homogeneity of the CC-norm to see that

$$\|\mathbf{X}^{a,x}\|^p_{1/p-\text{Höl};[(k-1)h,(k+1)h]}\overset{\mathcal{D}}{=}\frac{1}{r^p}\left\|\mathbf{X}^{a^r,\delta_r x}\right\|^p_{1/p-\text{Höl};[(k-1),(k+1)]}.$$

We then conclude with

$$\begin{aligned}\sup_{y\in\mathfrak{g}}\mathbb{P}^{a,y}\left(\|\mathbf{X}\|^p_{1/p-\text{Höl};[(k-1)h,(k+1)h]}\geq\frac{R}{4}\right)&=\sup_{y\in\mathfrak{g}}\mathbb{P}^{a^r,y}\left(\|\mathbf{X}\|^p_{1/p-\text{Höl};[(k-1),(k+1)]}\geq\frac{R^{1+p/2}}{4\cdot 2^{p+p^2/2}}\right)\\&\leq\sup_{a\in\Xi(\Lambda)}\sup_{y\in\mathfrak{g}}\mathbb{P}^{a,y}\left(\|\mathbf{X}\|^p_{1/p-\text{Höl};[(k-1),(k+1)]}\geq\frac{R^{1+p/2}}{2^{2+p+p^2/2}}\right)\\&\leq c\exp\left(-\frac{R^{1+2/p}}{c2^{4/p+2+p}}\right)\end{aligned}$$

where the last step results from applying Theorem 3.6. ■

References

[1] Dominique Bakry and Michel Ledoux. Lévy-Gromov's isoperimetric inequality for an infinite dimensional diffusion generator. *Inventiones mathematicae*, 123(1):259–281, 1996.

[2] Richard F. Bass. *Diffusions and elliptic operators*. Probability and its Applications. Springer-Verlag, New York, 1998.

[3] Richard F. Bass, Ben M. Hambly, and Terry Lyons. Extending the Wong-Zakai theorem to reversible Markov processes. *Journal of the European Mathematical Society*, 4(3):237–269, 2002.

[4] Fabrice Baudoin and Cheng Ouyang. On small time asymptotics for rough differential equations driven by fractional Brownian motions. Arxiv preprint. 2014.

[5] Fabrice Baudoin, Cheng Ouyang, and Samy Tindel. Upper bounds for the density of solutions to stochastic differential equations driven by fractional Brownian motions. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(1):111–135, 2014.

[6] Fabrice Baudoin, Cheng Ouyang, and Xuejing Zhang. Smoothing effect of rough differential equations driven by fractional Brownian motions. Arxiv preprint. 2013.

[7] Fabrice Baudoin, Cheng Ouyang, and Xuejing Zhang. Varadhan Estimates for rough differential equations driven by fractional Brownian motions. Arxiv preprint. 2013.

[8] Gérard Ben Arous, Mihai Gradinaru, and Michel Ledoux. Hölder norms and the support theorem for diffusions. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 30(3):415–436, 1994.

[9] Arne Beurling and Jacques Deny. Dirichlet spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 45(2):208–215, 1959.

[10] Nicholas H. Bingham, Charles M. Goldie, and Jozef L. Teugels. *Regular variation.* Cambridge University Press, Cambridge, 1987.

[11] Christer Borell. The Brunn-Minkowski inequality in Gauss Space. *Inventiones Mathematicae*, 30(2):207–216, 1975.

[12] Donald L. Burkholder. Martingale transforms. *The Annals of Mathematical Statistics*, 37(6):1494–1504, 1966.

[13] Mireille Capitaine, Elton P. Hsu, and Michel Ledoux. Martingale representation and a simple proof of logarithmic Sobolev inequalities on path spaces. *Electron. Comm. Probab*, 2:71–81, 1997.

[14] Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander's condition. *Annals of Mathematics*, 171(3):2115–2141, 2010.

[15] Thomas Cass, Martin Hairer, Christian Litterer, and Samy Tindel. Smoothness of the density for solutions to Gaussian Rough Differential Equations. *to appear in Annals of Probability*, 2014.

[16] Thomas Cass and Terry Lyons. Evolving Communities with Individual Preferences. *to appear in Proceedings of the London Mathematical Society*, 2014.

[17] Thomas Cass, Terry Lyons, and Christian Litterer. Integrability and Tail Estimates for Gaussian Rough Differential Equations. *Annals of Probability*, 41(4):3026–3050, 2013.

[18] Ilya Chevyrev and Terry Lyons. Characteristic functions of measures on geometric rough paths. Arxiv preprint. 2014.

[19] Edward Brian Davies. *Heat Kernels and Spectral Theory*. Cambridge University Press, Cambridge, 1989.

[20] Eugene Borisovich Dynkin. Calculation of the coefficients in the CampbellHausdorff formula. *Doklady Akademii Nauk SSSR*, 57:323–326, 1947.

[21] Peter Friz and Martin Hairer. *A Course on Rough Paths*. Universitext. Springer-Verlag, Berlin, 2014.

[22] Peter Friz, Terry Lyons, and Daniel Stroock. Lévy's area under conditioning. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 42(1):89–101, 2006.

[23] Peter Friz and Sebastian Riedel. Integrability of (Non-)Linear Rough Differential Equations and Integrals. *Stochastic Analysis and Applications*, 31(2):336–358, 2013.

[24] Peter Friz and Sebastian Riedel. Convergence rates for the full Gaussian rough paths. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 50(1):1–314, 2014.

[25] Peter Friz and Nicolas Victoir. On uniformly subelliptic operators and stochastic area. *Probability Theory and Related Fields*, 142(3-4):475–523, 2008.

[26] Peter Friz and Nicolas Victoir. Differential equations driven by Gaussian signals. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 46(2):369–413, 2010.

[27] Peter Friz and Nicolas Victoir. *Multidimensional Stochastic Processes as Rough Paths : Theory and Applications*. Cambridge University Press, Cambridge, 2010.

[28] M Fukushima, Y Oshima, and M Takeda. *Dirichlet forms and symmetric Markov processes*. W. de Gruyter, Berlin, 2010.

[29] Mikhael Gromov. Carnot-Carathéodory spaces seen from within. In A. Bellaiche and J. Risler, editors, *Sub-Riemannian Geometry*. Birkhäuser, Basel, 1996.

[30] Martin Hairer and Natesh S. Pillai. Regularity of laws and ergodicity of hypoelliptic SDEs driven by rough paths. *Annals of Probability*, 41(4):2544–2598, 2013.

[31] Yuzuru Inahama. Short time kernel asymptotics for rough differential equation driven by fractional Brownian motion. Arxiv preprint. 2014.

[32] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, Berlin, 2nd edition, 1991.

[33] Tosio Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin, 1995.

[34] Terry Lyons. Differential equations driven by rough signals. *Revista Matematica Iberoamericana*, 14(2):215–310, 1998.

[35] Terry Lyons. Differential equations driven by rough paths (handwritten Saint-Flour notes), 2004.

[36] Terry Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer-Verlag, Berlin, 2007.

[37] Terry Lyons and Lucretiu Stoica. The limits of stochastic integrals of differential forms. *Annals of Probability*, 27(1):1–49, 1999.

[38] Terry Lyons and Ofer Zeitouni. Conditional exponential moments for iterated Wiener integrals. *Annals of Probability*, 27(4):1738–1749, 1999.

[39] Terry Lyons and T. S. Zhang. Decomposition of Dirichlet processes and its application. *Annals of Probability*, 22(1):494–524, 1994.

[40] Zhi-Ming Ma and Michael Röckner. *Introduction to the Theory of (Non-Symmetric) Dirichlet Forms*. Springer, Berlin, 1992.

[41] Richard Montgomery. *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. American Mathematical Society, 2002.

[42] Sebastian Riedel. Talagrand's transportation-cost inequality and applications to (rough) path spaces. Arxiv preprint. 2014.

[43] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge University Press, Cambridge, 2000.

[44] Robert S. Strichartz. The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. *Journal of Functional Analysis*, 72(2):320–345, 1987.

[45] Daniel W. Stroock. Diffusion semigroups corresponding to uniformly elliptic divergence form operators. *Séminaire de Probabilités de Strasbourg*, 22:316–347, 1988.

[46] Karl-Theodor Sturm. Analysis on local Dirichlet spaces. I. Recurrence, conservativeness and Lp-Liouville properties. *Journal für die reine und angewandte Mathematik*, 456:173–196, 1994.

[47] Karl-Theodor Sturm. Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. *Osaka J. Math.*, 32:275–312, 1995.

[48] Karl-Theodor Sturm. Analysis on local Dirichlet spaces. III. The parabolic Harnack inequality. *Journal de Mathématiques Pures et Appliquées*, 1996.

[49] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes*. Springer-Verlag, Berlin, 2014.