Probab. Theory Relat. Fields (2008) 142:475–523
DOI 10.1007/s00440-007-0113-y

On uniformly subelliptic operators and stochastic area

Peter Friz · Nicolas Victoir

Received: 16 March 2007 / Revised: 17 October 2007 / Published online: 27 November 2007

Abstract Let X^a be a Markov process with generator $\sum_{i,j} \partial_i \left(a^{ij} \partial_j \cdot\right)$ where a is a uniformly elliptic symmetric matrix. Thanks to the fundamental works of T. Lyons, stochastic differential equations driven by X^a can be solved in the "rough path sense"; that is, pathwise by using a suitable stochastic area process. Our construction of the area, which generalizes previous works of Lyons–Stoica and then Lejay, is based on Dirichlet forms associated to subellitpic operators. This enables us in particular to discuss large deviations and support descriptions in suitable rough path topologies. As typical rough path corollary, Freidlin–Wentzell theory and the Stroock–Varadhan support theorem remain valid for stochastic differential equations driven by X^a.

Mathematics Subject Classification (2000) 60H99

1 Introduction

Let $V = (V_1, \ldots, V_d)$ be a collection of sufficiently nice vector fields on $\mathbb{R}^e$ and consider the stochastic differential equation in the Stratonovich sense $\mathrm{d}Y = V(Y)\,\mathrm{d}B$, $Y(0) = y_0 \in \mathbb{R}^e$, driven by a d-dimensional Brownian motion, a diffusion with generator $\frac{1}{2}\sum_{i=1}^{d} \partial_i^2$. We try to understand what happens when B is replaced by a d-dimensional diffusion process $X = X^a$ with uniformly elliptic generator in

Peter Friz is a Leverhulme Fellow.

P. Friz (✉)
Department of Pure Mathematics and Mathematical Statistics, University of Cambridge,
Wilberforce Road, Cambridge, CB3 0WB, UK
e-mail: P.K.Friz@statslab.cam.ac.uk

N. Victoir
Hong Kong, Hong Kong

divergence form $\sum_{i,j=1}^{d} \partial_i \left(a^{ij} \partial_j \cdot\right)$. Of course, $\mathrm{d}Y = V(Y)\,\mathrm{d}X$ still makes sense as Stratonovich equation if a is smooth but this breaks down when a is only assumed to be measurable. Such an assumption is not only standard in the theory of partial differential equations but also a basic example in the theory of Dirichlet forms [12] and the construction of the corresponding diffusion process X^a is well-known, e.g., [12,28] .

We recall that one can construct X^a as weak limit of semi-martingales $X^{a(\varepsilon)}$ along a sequence of mollifier approximations $\{a(\varepsilon) : \varepsilon > 0\}$. It is a natural question [16] if the sequence of SDE solutions driven by $X^{a(\varepsilon)}$ converges. One can also replace X^a by piecewise linear approximations $X^a(n)$ and ask if the resulting ODE solutions converge. It turns out they all converge to the same limiting object which can be constructed intrinsically as solution to the *rough differential equation* [17,19] of form $\mathrm{d}Y = V(Y)\,\mathrm{d}\mathbf{X}$.A stochastic area process A^a is now considered part of the driving signal $\mathbf{X} = (X^a, A^a)$. The construction of A^a was carried out by subtle forward-backward martingale arguments in [20], together with a convergence statement for piecewise linear approximations. It is verified in [15] that convergence takes place in suitable rough path metrics . By the fundamental continuity result of rough path theory this implies the convergence of ODE solutions driven by $X^a(n)$, i.e., a Wong–Zakai theorem.

In contrast to [15,16,20] we emphasize and exploit the Markovian nature of (X^a, A^a). The basic observation is that for smooth a we are dealing with semi-martingales X^a so that the stochastic area process should be given in terms of Itô stochastic integrals,

$$t \mapsto A_t^a \equiv \frac{1}{2} \int_0^t \left(X^a \otimes \mathrm{d}X^a - \mathrm{d}X^a \otimes X^a\right) \in so(d).$$

It is a simple exercise in Itô calculus[1] to see that the process (X^a, A^a) is Markov with (uniformly subelliptic) generator of form

$$L^a = \sum_{i,j=1}^{d} U_i \left(a^{ij} U_j \cdot\right). \tag{1}$$

The vector fields $U_1, \ldots, U_d$ are defined in (3) and play the rôle of coordinate vector fields $\partial_1, \ldots, \partial_d$ on $g^2\left(\mathbb{R}^d\right) \equiv \mathbb{R}^d \oplus so(d)$, which is given the structure of a Lie group G. Of course, L^a is understood in a weak sense and the correct mathematical object is the Dirichlet form[2]

[1] Once can proceed as follows. First write $X = X^a$ as solution to a Stratonovich SDE involving a smooth square-root of a. In combination with the fact the the lift of X, denoted by Y say, is obtained by solving the Stratonovich equation $\mathrm{d}Y = \sum_{i=1}^{d} U_i(Y) \circ \mathrm{d}X^i$ along the left-invariant vectorfields $U_1, \ldots, U_d$ on $g_2\left(\mathbb{R}^d\right)$ as defined in (3), a few lines of Itô calculus identify the generator of the lift.

[2] Lebesgue measure on $g^2\left(\mathbb{R}^d\right)$ coincides with Haar measure m on G. Then $U_i^* = -U_i$ where * denotes the formal adjoint with respect to m.

$$\mathcal{E}^a(f,g) = \sum_{i,j=1}^{d} \int_G \mathrm{d}m\, a^{ij} U_i f U_j g. \tag{2}$$

We can thus use the highly developed analytic machinery of Dirichlet forms [5,12]; the collections of results in [30], in conjunction with [27], applies directly to (2). Leaving precise references to those papers, the relevant results in [30] are based on the seminal works of De Giorgi, Nash, Moser for the elliptic case and the various extensions to subelliptic/Hörmander type operators as studied in papers by Rothschild, Stein, Jerison, Sánchez-Calle, Nagel, Waigner and many others.

This paper is organized as follows. In Sects. 2 and 3 we spezialise the toolbox of Dirichlet forms to our situation and settle the notation. In Sect. 4 we show that the $g^2(\mathbb{R}^d)$-valued Markov process $\mathbf{X}^a$ has, just as Brownian motion and Lévy area, $(1/2-\varepsilon)$-Hölder regularity with respect to Carnot–Caratheodory distance on $g^2(\mathbb{R}^d)$. It follows that, a.e., sample path $\mathbf{X}^a(\omega)$ is a geometric Hölder rough path in the sense of Lyons [17,10]. In fact, the Hölder norm of $\mathbf{X}^a$ is seen to have Gaussian tail which answers a question raised in Lyons' St. Flour lecture [18]. In Sect. 5 we study both weak approximations, $a_n \to a$, a.e., is seen to imply $\mathbf{X}^{a_n} \to \mathbf{X}^a$ in distribution, and a strong Wong–Zakai type theorem. The latter shows that our stochastic area associated to X^a coincides with the area constructed by Lyons and Stoica [20] and we improve on results in [15,16]. In Sect. 6 we note that an RDE solution jointly with its driving signal $\mathbf{X}^a$ is Markov and describe its generator, using stochastic Taylor expansions for random RDEs obtained in [11]. In Sect. 7 we prove a sample path large deviation principle for $\mathbf{X}^a$ making crucial use of Ramírez's result [25]. As a typical rough paths corollary, we obtain Freidlin–Wentzell type large deviations for stochastic differential equations driven by $\mathbf{X}^a$ in the rough path sense. Finally, in Sect. 8 we revert to the case where $\mathbf{X}^a$ is the lift of X^a (that is, a is defined on $\mathbb{R}^d$ rather than $g^2(\mathbb{R}^d)$) and prove that $\mathbf{X}^a$ has full support in suitable Hölder topologies. As a typical rough paths corollary, we obtain a Stroock–Varadhan type support theorem for stochastic differential equations driven by $\mathbf{X}^a$ in the rough path sense. Such a support description was conjectured by T. Lyons in [18].

Notation 1 *Although the key notations are introduced in the main text as appropriate we feel the reader will be helped by this brief summary. The space of real antisymmetric $d\times d$ matrices is denoted by $so(d)$ and is given the standard Euclidean structure with $\cdot$ denoting the scalar product. The corresponding norm is denoted by $|\cdot|$. It will cause no confusion to use $\cdot$ and $|\cdot|$ also for standard scalar product and Euclidean norm on $\mathbb{R}^d$. The vector space $g^2(\mathbb{R}^d) = \mathbb{R}^d \oplus so(d)$ will be given a nilpotent Lie algebra structure so that the corresponding Lie group can and will be realized on the same space, $(g^2(\mathbb{R}^d), *, 0)$. Points in $g^2(\mathbb{R}^d)$ are denoted by $x, y, z, \ldots$ and may be written out in coordinates as $((x^{1;i}), (x^{2;jk}) : i, j, k = 1, \ldots, d$ with $j < k)$. We also write $x^1 = \pi_1(x)$, $x^2 = \pi_2(x)$ for the projections to $\mathbb{R}^d$, $so(d)$ respectively. Haar measure on $g^2(\mathbb{R}^d)$ coincides with Lebesgue measure on $\mathbb{R}^d \oplus so(d)$ and is denoted by m, in integrations we write $\mathrm{d}m$, $\mathrm{d}m(x)$ or simply $\mathrm{d}x$. We use $\langle\cdot,\cdot\rangle$ for the scalar product in $L^2(g^2(\mathbb{R}^d), m)$ and the corresponding L^2-norm is written as $|\cdot|_{L^2}$ or $|\cdot|_{L^2(D)}$ for $D \subset g^2(\mathbb{R}^d)$. The Lie group $g^2(\mathbb{R}^d)$ has a dilation structure $\delta_\lambda(x) \mapsto (\lambda\pi_1(x), \lambda^2\pi_2(x))$,* *carries a Carnot–Carathedory continuous norm $x \mapsto \|x\|$, homogenuous in the sense that $\|\delta_\lambda(x)\| = |\lambda| \|x\|$, and equivalent to $|\pi_1(x)| + |\pi_2(x)|^{1/2}$. It induces the left invariant Carnot–Caratheodory distance $d(x, y) = \|x^{-1} * y\|$under which $g^2(\mathbb{R}^d)$ is a metric (in fact: geodesic) space. This distance coincides with the intrinsic metric from a reference Dirichlet form $\mathcal{E}$. A family of Dirichlet forms $\{\mathcal{E}^a : a \in \Xi(\Lambda)\}$, where $\Xi(\Lambda)$ denotes a class of certain diffusion matrices with ellipticity constant Λ, gives rise to a family of intrinsic metrics on $g^2(\mathbb{R}^d)$, denoted by d^a, all Lipschitz equivalent to d. Stochastic process with values in $g^2(\mathbb{R}^d)$ are denoted by capital bold letter such as $\mathbf{X}$ or $\mathbf{X}^{a,x}$ to indicate dependence on $a \in \Xi(\Lambda)$ and starting point. The $so(d)$-valued area process $A := \pi_2(\mathbf{X})$ will be of interest. A fixed path in $C([0, 1], g^2(\mathbb{R}^d))$ may be written as $\mathbf{x} = \mathbf{x}(\cdot)$ or ω, the latter is used when $C([0, 1], g^2(\mathbb{R}^d))$ is equipped with a Borel measure such as the law of $\mathbf{X}^{a,x}$ for which we write $\mathbb{P}^{a,x}$. L^p-norms with respect to $\mathbb{P}^{a,x}$ are denoted by $\|\cdot\|_{L^p(\mathbb{P}^{a,x})}$. A path $\mathbf{x} \in C([0, 1], g^2(\mathbb{R}^d))$ has increments $\mathbf{x}_{s,t} = \mathbf{x}_s^{-1} * \mathbf{x}_t =: (\mathbf{x}^1_{s,t}, \mathbf{x}^2_{s,t})$. Note $\mathbf{x}^1_t - \mathbf{x}^1_s = \mathbf{x}^1_{s,t}$ but $\mathbf{x}^2_t - \mathbf{x}^2_s \neq \mathbf{x}^2_{s,t} = \mathbf{x}^2_t - \mathbf{x}^2_s - [\mathbf{x}^1_s, \mathbf{x}^1_{s,t}]/2$. (Semi-)norms and distances are defined naturally on this path space over $g^2(\mathbb{R}^d)$. In particular,*

$$\|\mathbf{x}\|_{\alpha-\text{Höl}} = \sup_{0\le s<t\le 1} \frac{\mathrm{d}(\mathbf{x}_s, \mathbf{x}_t)}{|t-s|^\alpha} = \sup_{0\le s<t\le 1} \frac{\|\mathbf{x}_{s,t}\|}{|t-s|^\alpha} \sim \sup_{0\le s<t\le 1} \frac{|\mathbf{x}^1_{s,t}| + |\mathbf{x}^2_{s,t}|^{1/2}}{|t-s|^\alpha}.$$

and

$$d_{\alpha-\text{Höl}}(\mathbf{x}, \mathbf{y}) = \sup_{0\le s<t\le 1} \frac{\mathrm{d}(\mathbf{x}_{s,t}, \mathbf{y}_{s,t})}{|t-s|^\alpha}.$$

We write $d_0 \equiv d_{0-\text{Höl}}$ and $d_\infty(\mathbf{x}, \mathbf{y}) = \sup_{0\le t\le 1} d(\mathbf{x}_t, \mathbf{y}_t)$. Care must be taken since d_0 and d_∞ are not Lipschitz equivalent. We avoid the double bar notation, i.e., $\|\cdot\|_{(\ldots)}$, for semi-norms resp. distances on the path space over some Euclidean space $\mathbb{R}^e$, $e \in \mathbb{N}$. For instance, when $y \in C([0, 1], \mathbb{R}^e)$ we write

$$|y|_{\alpha-\text{Höl}} = \sup_{0\le s<t\le 1} \frac{|y_t - y_s|}{|t-s|^\alpha} = \sup_{0\le s<t\le 1} \frac{|y_{s,t}|}{|t-s|^\alpha}.$$

Vector fields (usually on some Euclidean space $\mathbb{R}^e$, $e \in \mathbb{N}$) are denoted by V and usually assumed to be in some regularity class Lip^p *which means bounded derivatives up to order $\lfloor p \rfloor$, and Hölder regularity of the $\lfloor p \rfloor$th derivative with exponent $p - \lfloor p \rfloor$. In particular, such vector fields are bounded. The (smooth but unbounded) invariant vector fields on $g^2(\mathbb{R}^d)$ are denoted by U_i. A dissection D of $[0, 1]$ is a collection $\{0 = t_0 < t_1 < \cdots < t_{\#D-1} < t_{\#D} = 1\}$. Its mesh is defined as $|D| = \sup_{i=1,\ldots,\#D} t_i - t_{i-1}$. Given $t \in [0, 1]$ we write t_D for its lower neighbour in D that is $t_D = \max\{t_i \in D : t_i \le t\}$. Similarly, t^D denotes the upper neighbour in D. Constants which appears in statement are typically indexed by the statement number. To indicate changing constant in proofs we sometimes number them with upper indices. (This will* *cause no confusion with powers.) We try to be explicit about the dependence of all constant with the exception of* $d = \dim\left(\mathbb{R}^d\right)$.

2 Analysis on the group

Let $g^2\left(\mathbb{R}^d\right)$ be the free step-2 nilpotent Lie algebra over $\mathbb{R}^d$, that is $g^2\left(\mathbb{R}^d\right) = \mathbb{R}^d \oplus so\,(d)$ ($so\,(d)$ being the space of antisymmetric $d \times d$ matrices) with Lie bracket

$$[x, y] \equiv \left[\left(x^1, x^2\right), \left(y^1, y^2\right)\right] = x^1 \otimes y^1 - y^1 \otimes x^1.$$

Due to nilpotency and the Campbell–Baker–Hausdorff formula, we can and will realize the associated Lie group on the same space $g^2\left(\mathbb{R}^d\right) = \mathbb{R}^d \oplus so\,(d)$ with product

$$x * y = x + y + \frac{1}{2}[x, y]$$

and unit element 0. Lebesgue-measure dx on $\mathbb{R}^d \oplus so\,(d)$ is the (left- and right-invariant) Haar measure m; in symbols dx = dm, see [37] for instance. For $i = 1, \ldots, d$ we define left-invariant vector fields by

$$U_i(x) = \partial_i + \frac{1}{2}\left(\sum_{1 \leq j < i \leq d} x^{1;j}\partial_{j,i} - \sum_{1 \leq i < j \leq d} x^{1;j}\partial_{i,j}\right) \tag{3}$$

where ∂_i denotes the coordinate vector field on $\mathbb{R}^d$ and $\partial_{i,j}$ with $i < j$ the coordinate vector field on $so\,(d)$, identified with its upper diagonal elements. A simple computation shows that $\left[U_i, U_j\right] = \partial_{i,j}$ and all higher brackets are zero. Since Hörmander's condition is satisfied, we call $\nabla = (U_1, \ldots, U_d)$ the *hypoelliptic gradient*. A (symmetric, regular, strongly local) Dirichlet form on $L^2\left(g^2\left(\mathbb{R}^d\right), \mathrm{d}x\right)$ is defined by

$$\mathcal{E}\,(f, g) = \int\limits_{g^2(\mathbb{R}^d)} \nabla f \cdot \nabla g \,\mathrm{d}m$$

with domain $\mathcal{F} := D\,(\mathcal{E}) := \left\{f \in L^2 : \mathcal{E}\,(f, f) < \infty\right\}$, closure of smooth compactly support functions with respect to

$$\|f\|_{\mathcal{F}} = \left(\mathcal{E}\,(f, f) + \langle f, f\rangle_{L^2(g^2(\mathbb{R}^d))}\right)^{1/2}.$$

This is a very standard setting, see [12,37], and as pointed out in the introduction, $\mathcal{E}$ is the Dirichlet form associated to the Markov process Brownian Motion plus its canonical Levy area. The Dirichlet form $\mathcal{E}$ is based on the *carré du champ operator*

$$\Gamma(f,g) = \nabla f \cdot \nabla g = \sum_{i=1}^{d} U_i f(.)\, U_i g(.),$$

which can be defined for all $f, g \in \mathcal{F}_{\text{loc}} = \left\{ f \in L^2 : \Gamma(f,f) \in L^1_{\text{loc}}(\mathrm{d}m) \right\}$. The associated *energy measure* is simply $\mathrm{d}\Gamma(f,g) := \Gamma(f,g)\,\mathrm{d}m$. Given $x, y \in g^2(\mathbb{R}^d)$ the (left-invariant) *Carnot–Caratheodory* or *control* distance $d(x,y)$ is defined as the length of the shortest path from x to y which remains tangent to *span* $\{U_1, \ldots, U_d\}$, and the induced topology coincides with the original topology of $g^2(\mathbb{R}^d)$; the *Carnot–Caratheodory norm* is defined as $\|x\| = d(0,x)$. See [22,37] or [10]. From [4, Lemma 5.29], this distance coincides with the *intrinsic metric* of $\mathcal{E}$,

$$d(x,y) = \sup\{f(x) - f(y) : f \in \mathcal{F}_{\text{loc}} \text{ and } f \text{ continuous}, \Gamma(f,f) \le 1\}.$$

Proposition 2 (I) *Completeness Property*: *In the metric space $\left(g^2(\mathbb{R}^d), d\right)$, every closed ball $\bar{B}$*

$$\bar{B}(x,r) = \left\{ y \in g^2\left(\mathbb{R}^d\right) : d(x,y) \le r \right\}$$

is complete and compact.

(II) *Doubling Property*: *The volume-doubling property*

$$\forall r \ge 0 : m\left(B(x,2r)\right) \le 2^N m\left(B(x,r)\right).$$

holds with $N = d^2$.

(III) *Poincaré Inequality*: *For all $r \ge 0$ and $f \in D(\mathcal{E})$*

$$\int_{B(x,r)} \left| f - \bar{f}_r \right|^2 \mathrm{d}m \le C_2 r^2 \int_{B(x,r)} \Gamma(f,f)\, \mathrm{d}m$$

where

$$\bar{f}_r = m\left(B(x,r)\right)^{-1} \int_{B(x,r)} f\,\mathrm{d}m.$$

(IV) *Nash Inequality*: *For all $f \in D(\mathcal{E}) \cap L^1$ we have*

$$\|f\|_{L^2}^{2+4/d^2} \le C_2' \mathcal{E}(f,f)\, \|f\|_{L^1}^{4/d^2}.$$

Proof Property (I) is a simple consequence of $\left(g^2(\mathbb{R}^d), d\right)$ being complete, property (II) follows from left then, every closed subset is complete. (II) follows readily from invariance of m under translation, $B(0,r) = \delta_r B(0,1)$ and the Jacobian of δ_λ (as map from $g^2(\mathbb{R}^d) = \mathbb{R}^d \oplus so(d)$ into itself) being equal to $\lambda^d.\left(\lambda^2\right)^{\frac{d(d-1)}{2}} = \lambda^{d^2}$. Property (III) appears explicitly in an appropriate Lie group setting in [14]. At last, Property (IV) follows from [4,26] or [37]. □

3 Uniformly subelliptic dirichlet forms

For $\Lambda \geq 1$ we call $\Xi(\Lambda)$ the set of all measurable maps a from $g^2\left(\mathbb{R}^d\right)$ into the space of symmetric matrics such that

$$\forall \xi \in \mathbb{R}^d : \frac{1}{\Lambda}|\xi|^2 \leq \xi \cdot a\xi \leq \Lambda |\xi|^2 .$$

A symmetric Dirichlet form on $L^2\left(g^2\left(\mathbb{R}^d\right), \mathrm{d}x\right)$ is defined by

$$\begin{aligned}\mathcal{E}^a(f, g) &= \int_{g^2(\mathbb{R}^d)} \nabla f(x) \cdot a(x) \nabla g(x)\, \mathrm{d}m \\ &= \sum_{i,j=1}^{d} \int_{g^2(\mathbb{R}^d)} a^{ij}(x) U_i f(x) U_j g(x)\, \mathrm{d}x.\end{aligned}$$

The associated carré du champ operator and energy measure are given by

$$\Gamma^a(f, g) = \nabla f(x) \cdot a(x) \nabla g(x), \quad \mathrm{d}\Gamma^a(f, g) = \Gamma^a(f, g)\, \mathrm{d}m,$$

respectively. The forms $\mathcal{E}^a$ and $\mathcal{E}$ are *quasi-isometric* in the sense that $D(\mathcal{E}) = D(\mathcal{E}^a)$ and for all f in the common domain,

$$\frac{1}{\Lambda}\mathcal{E}(f, f) \leq \mathcal{E}^a(f, f) \leq \Lambda \mathcal{E}(f, f). \tag{4}$$

The intrinsic metric associated to $\mathcal{E}^a(f, f)$,

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

is obviously Lipschitz equivalent to $d(x, y)$ and hence a metric on $g^2\left(\left(\mathbb{R}^d\right)\right)$ which induces the original topology so that, in particular, $d^a(\cdot, \cdot)$ is continuous. Moreover, $\left(g^2\left(\left(\mathbb{R}^d\right)\right), d^a\right)$ is complete since $\left(g^2\left(\left(\mathbb{R}^d\right)\right), d\right)$ is and closed balls are easily seen to be compact, see property (I) above and in Propositions 2 and 4. The following proposition is a special case of a result in [32].

Proposition 3 *For all $a \in \Xi(\Lambda)$, the space $\left(g^2\left(\left(\mathbb{R}^d\right)\right), d^a\right)$ is a geodesic space in the sense that for all x, y there exists a continuous map $\gamma : [0, 1] \to g^2\left(\left(\mathbb{R}^d\right)\right)$ with $\gamma_0 = x,\ \gamma_1 = y$ and*

$$d^a(\gamma_r, \gamma_t) = d^a(\gamma_r, \gamma_s) + d^a(\gamma_s, \gamma_t) \quad \text{for all } 0 \leq r < s < t \leq 1.$$

Proposition 4 *Let $a \in \Xi(\Lambda)$. Properties* (I),(II),(III),(IV) *in proposition 2 remain valid when we replace $\mathcal{E}$ by $\mathcal{E}^a$ and d by d^a.*

Proof Such properties are invariant under quasi-isometry, i.e., whenever we have (4). This is easy to see for properties (I), (II), (IV). Invariance of the Poincaré inequality (III), discussed in detail in [30], is seen by first proving that the Poincaré inequality is equivalent to a *weak Poincaré inequality* for which quasi-isometry is obvious.) □

Standard semigroup theory [5,12] allows us to associate a non-positive self-ajoint operator L^a to $\mathcal{E}^a$. We then have[3]

Proposition 5 (V) *Parabolic Harnack Inequality*: *Let* $a \in \Xi(\Lambda)$. *There exists a constant* $C_5 = C_5(\Lambda)$ *such that*

$$\sup_{(s,y)\in Q^-} u(s,y) \leq C_5 \inf_{(s,y)\in Q^+} u(s,y),$$

whenever u is a nonnegative weak solution of the parabolic partial differential equation $\partial_t u = L^a u$ *on some cylinder* $Q = \left(t-4r^2, t\right) \times B(x,2r)$ *for some reals* $t, r > 0$. *Here,* $Q^- = \left(t-3r^2, t-2r^2\right) \times B(x,r)$ *and* $Q^+ = \left(t-r^2, t\right) \times B(x,r)$ *are lower and upper sub-cylinders of Q separated by a lapse of time. The statement remains valid for balls with respect to* d^a.

Proof Based on the classical ideas by Moser [23,24], Grigor'yan, Saloff–Coste, it is shown in [30] that if (I) holds then (II)+ (III) $\Leftrightarrow$ (V). For a more direct proof along ideas of Nash, see [27,28]. □

Following [8,28,30] (these paper building on the seminal works of De Giorgi–Moser–Nash) we have also Hölder regularity of such weak solution (and in particular of the heat kernels discussed below). We will refer to this simply as *De Giorgi–Moser–Nash regularity*:

Proposition 6 *Let* $a \in \Xi(\Lambda)$. *Then there exist constants* $\eta \in (0,1)$ *and* C_6, *only depending on* Λ, *such that*

$$\sup_{(s,y),(s',y')\in Q_1} \left|u(s,y) - u\left(s',y'\right)\right| \leq C_6 \sup_{u\in Q_2} |u| \,. \left(\frac{\left|s-s'\right|^{1/2} + d\left(y,y'\right)}{r}\right)^{\eta}.$$

whenever u is a nonnegative weak solution of the parabolic partial differential equation $\partial_s u = L^a u$ *on some cylinder* $Q_2 \equiv \left(t-4r^2, t\right) \times B(x,2r)$ *for some reals* $t, r > 0$. *Here* $Q_1 \equiv \left(t-r^2, t-2r^2\right) \times B(x,r)$ *is a subcylinder of* Q_2.

3.1 Upper and lower heat kernel bounds

Heat kernel existence is not an issue here. (For instance, [5,8,27,28], Nash's inequality (IV) implies an estimate on $\left\| P_t^a \right\|_{L^1\to L^2}$ and then via duality on $\left\| P_t^a \right\|_{L^1\to L^\infty}$ which implies existence of the heat kernel $p^a = p^a(t,x,y)$.) We now turn to Aronson-type

[3] In view of De Giorgi–Moser–Nash regularity, see below, we may indeed write inf, sup rather than ess-inf, ess-sup.

[1] heat-kernel estimates. As a well-known consequence of our proposition 4 (see [30, Corollary 4.2], also [5,8,28]) we get

Theorem 7 *Let $a \in \Xi(\Lambda)$. The heat kernel p^a satisfies, for $\varepsilon > 0$ fixed,*

$$p^a(t,x,y) \le \frac{C_7}{t^{d^2/2}} \exp\left(-\frac{d^a(x,y)^2}{(4+\varepsilon)\,t}\right)$$

for some constant $C_7 = C_7(\varepsilon,\Lambda)$.

Theorem 8 *Let $a \in \Xi(\Lambda)$. The heat kernel p^a satisfies*

$$p^a(t,x,y) \ge \frac{1}{C_8}\frac{1}{t^{d^2/2}} \exp\left(-\frac{C_8 d^a(x,y)^2}{t}\right)$$

for some constant $C_8 = C_8(\Lambda)$.

Let $a \in \Xi(\Lambda)$. Let C_7, C_8 denote the constants of the previous two theorems. Then

$$\frac{1}{C_8}\frac{1}{t^{d^2/2}} \exp\left(-\frac{C_8 \Lambda d(x,y)^2}{t}\right) \le p^a(t,x,y) \le \frac{C_7}{t^{d^2/2}} \exp\left(-\frac{\mathrm{d}(x,y)^2}{\Lambda\,(4+\varepsilon)\,t}\right).$$

Proof Lipschitz-equivalence of $d(x,y)$ and $d^a(x,y)$. □

3.2 The associated Markov process

Following a standard construction, the heat kernel p^a gives rise to a consistent family of finite-dimensional distributions and determines a $g^2(\mathbb{R}^d)$-valued (strong) Markov process $\left(\mathbf{X}_t^{a,x} : t \ge 0\right)$ where $a \in \Xi(\Lambda)$ and $\mathbf{X}_0^{a,x} = x \in g^2(\mathbb{R}^d)$. The natural time horizon is $[0,\infty)$ but our focus will be on finite time horizon and by scaling (cf. next section) there is no loss of generality to work on $[0,1]$. The heat kernel estimates are more than enough, via Kolmogorov's criterion, to guarantee that any such process can be taken with continuous sample paths; the law of $\mathbf{X}^{a,x}$ is then denoted by $\mathbb{P}^{a,x}$, a Borel measure on $C\left([0,1], g^2(\mathbb{R}^d)\right)$, under which we can think of $\mathbf{X} = \mathbf{X}^{a,x}$ simply as coordinate process $\mathbf{X}_t(\omega) = \omega_t$. By construction, the density of $\mathbf{X}_t$ under $\mathbb{P}^{a,x}$, or equivalently, the density of $\mathbf{X}_t^{a,x}$, with respect to m is given by $p^a(t,x,\cdot)$.

3.3 Scaling

We will refer to the following simple proposition as *scaling*. Recall that the dilation operator δ extends scalar multiplication to $g^2(\mathbb{R}^d)$.

Proposition 9 *For any $a \in \Xi(\Lambda)$, $r \ne 0$ set $a^r(x) := a\left(\delta_{1/r}x\right) \in \Xi(\Lambda)$. Then*

$$\left(\mathbf{X}_t^{a^r,x} : t \ge 0\right) \overset{\mathcal{D}}{=} \left(\delta_r \mathbf{X}_{t/r^2}^{a,\delta_{1/r}(x)} : t \ge 0\right).$$

3.4 Short time asymptotics

When $a = I$, the identity matrix, an essentially sharp lower bound with $1/C_8 = 4(1-\varepsilon)$ is known, see [36]. This implies Varadhan's formula

$$4t \log p^I(t,x,y) \to -d^I(x,y)^2 \quad \text{as } t \to 0.$$

The generalization to arbitrary $a \in \Xi(\Lambda)$ follows from the recent work of Ramírez [25] and will be central to our discussion of large deviations.

Theorem 10 *The heat kernel associated to L^a satisfies, for all $x, y \in g^2(\mathbb{R}^d)$*

$$4t \log p^a(t,x,y) \to -d^a(x,y)^2 \quad \text{as } t \to 0.$$

3.5 A lower bound for the killed process

Theorem 11 *Let $a \in \Xi(\Lambda)$. For $x_0 \in g^2(\mathbb{R}^d)$ and $r > 0$, define*

$$\xi^{a;x}_{B(x_0,r)} = \inf\left\{t \geq 0 : \mathbf{X}^{a;x}_t \notin B(x_0,r)\right\},$$
$$\mathbb{P}^{a;x}_{B(x_0,r)}(t,\cdot) = \mathbb{P}\left(\mathbf{X}^{a,x}_t \in \cdot\,, \xi^{a;x}_{B(x_0,r)} > t\right).$$

Then $\mathbb{P}^{a;x}_{B(x_0,r)}(t,dy) = p^a_{B(x_0,r)}(t,x,y)\mathrm{d}y$. $> >>$ CHECK B versus B^a.$<<<$ Moreover, if x, y are two elements of $B^a(x_0,r)$ joined by a curve γ which is at a d^a-distance $R > 0$ of $g^2(\mathbb{R}^d)/B^a(x_0,r)$ there exists constant C_{11} depending only on Λ,

$$p^a_{B(x_0,r)}(t,x,y) \geq \frac{1}{C_{11}\delta^{d^2/2}} \exp\left(-C_{11}\frac{d^a(x,y)^2}{t}\right)\exp\left(-\frac{C_{11}t}{R^2}\right)$$

where $\delta = \min\left\{t, R^2\right\}$.

Proof See [30] or [27], the ideas are adapted from [8,28]. □

One should observe that d^a can be replaced by d, at the price of changing the constants.

4 Construction of associated rough paths

In conjunction with the ever useful Garsia–Rodemich–Rumsey's lemma, the upper heat bounds leads to Hölder regularity of the sample paths $t \mapsto \mathbf{X}^{a;x}_t(\omega)$. Moreover, a *Fernique estimate* holds by which we mean that the homogenous Hölder norm of the $g^2(\mathbb{R}^d)$-valued process $\mathbf{X}^{a;x}$ has a Gauss tail.

Lemma 12 *For all* $\eta < \frac{1}{4\Lambda}$ *we have*

$$\sup_{a\in\Xi(\Lambda)}\sup_{x\in g^2(\mathbb{R}^d)}\sup_{0\le s<t\le 1}\mathbb{E}^{a,x}\left(\exp\left(\eta\frac{\mathrm{d}\left(\mathbf{X}_t,\mathbf{X}_s\right)^2}{t-s}\right)\right)<\infty.$$

Proof By scaling and the Markov property, for any $a\in\Xi(\Lambda)$,

$$\begin{aligned}&\sup_{x\in g^2(\mathbb{R}^d)}\sup_{0\le s<t\le 1}\mathbb{E}^{a,x}\left(\exp\left(\eta\frac{\mathrm{d}\left(\mathbf{X}_t,\mathbf{X}_s\right)^2}{t-s}\right)\right)\\&\quad\le\sup_{x\in g^2(\mathbb{R}^d)}\sup_{a\in\Xi(\Lambda)}\mathbb{E}^{a,x}\left(\exp\left(\eta\left\|\mathbf{X}_{0,1}\right\|^2\right)\right).\end{aligned}$$

(Recall that $d\left(\mathbf{X}_t,\mathbf{X}_s\right)=d\left(0,\mathbf{X}_s^{-1}*\mathbf{X}_t\right)=\left\|\mathbf{X}_{s,t}\right\|$where $\|\cdot\|=d\left(0,\cdot\right)$ denotes the Carnot–Caratheodory norm.) Fix $\eta<\frac{1}{4\Lambda}$, and $\varepsilon>0$ such that $\eta<\frac{1}{4(1+\varepsilon)\Lambda}$. Then, from the heat kernel upper-bound, we obtain

$$\begin{aligned}\mathbb{E}^{a,x}\left(\exp\left(\eta\left\|\mathbf{X}_{0,1}\right\|^2\right)\right)&=\int\exp\left(\eta d(x,y)^2\right)p^a(1,x,y)\mathrm{d}y\\&\le C_7\int\exp\left(-\left(\frac{1}{4(1+\varepsilon)\Lambda}-\eta\right)d(x,y)^2\right)\mathrm{d}y\end{aligned}$$

From $m\left(B(x,r)\right)=m\left(B\left(0,1\right)\right)r^{d^2}$ we have $\mathrm{d}m\left(B(x,r)\right)/\mathrm{d}r=m\left(B\left(0,1\right)\right)$ $d^2r^{d^2-1}$ so that

$$\begin{aligned}&\mathbb{E}^{a,x}\left(\exp\left(\eta\left\|\mathbf{X}_{0,1}\right\|^2\right)\right)\\&\quad\le C_7m\left(B\left(0,1\right)\right)d^2\int_{r=0}^{\infty}\exp\left(-\left(\frac{1}{4(1+\varepsilon)\Lambda}-\eta\right)r^2\right)r^{d^2-1}\mathrm{d}r\end{aligned}$$

and by our choice of η,ε the right hand side is finite, uniformly in x and $a\in\Xi(\Lambda)$ as required. □

The previous lemma combined with a standard application of the Garsia–Rodemich–Rumsey lemma leads immediately to Fernique estimate for homogenous α-Hölder norm

$$\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]}=\sup_{0\le s<t\le 1}\frac{\mathrm{d}\left(\mathbf{X}_t,\mathbf{X}_s\right)}{|t-s|^\alpha}.$$

More precisely, we have

Theorem 13 *Let* $0 \leq \alpha < 1/2$*.There exists a constant* $C_{13} = C_{13}(\Lambda, \alpha) > 0$ *such that*

$$\sup_{a \in \Xi(\Lambda)} \sup_{x \in g^2(\mathbb{R}^d)} \mathbb{E}^{a,x} \left[\exp \left(C_{13} \|\mathbf{X}\|^2_{\alpha-\text{Höl};[0,1]} \right) \right] < \infty.$$

In particular, for $\alpha \in (1/3, 1/2)$ *almost every sample path* $t \mapsto \mathbf{X}_t^{a;x}(\omega)$ *is an* α*-Hölder geometric rough path.*

For later use—namely our discussion of Wong–Zakai approximations—we record the following estimate.

Corollary 14 *Let*

$$M_\eta := \sup_{a \in \Xi(\Lambda)} \sup_{x \in g^2(\mathbb{R}^d)} \sup_{0 \leq s < t \leq 1} \mathbb{E}^{a,x} \left(\exp \left(\eta \frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^2}{t-s} \right) \right). \tag{5}$$

Then there exists $C_{14} = C_{14}(\Lambda)$ *such that* $M_\eta \leq \exp(C_{14}\eta)$ *for all* $\eta \in \left[0, \frac{1}{16\Lambda}\right)$.

Proof It suffices to show $M_\eta \leq 1 + C_{14}\eta$. From the inequality $\exp(x) \leq 1 + x\exp(x)$ for $x > 0$ we obtain

$$M_\eta \leq 1 + \eta \sup_{x \in \mathbb{R}^d} \sup_{s < t \in [0,1]} \mathbb{E}^{a,x} \left(\frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^2}{t-s} \exp \left(\eta \frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^2}{t-s} \right) \right).$$

Define

$$Q_4 := \sup_{a \in \Xi(\Lambda)} \sup_{x \in g^2(\mathbb{R}^d)} \sup_{s < t \in [0,1]} \mathbb{E}^{a,x} \left(\frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^4}{|t-s|^2} \right) < \infty.$$

The proof is now finished by Cauchy–Schwarz,

$$\begin{aligned} M_\eta &\leq 1 + \eta Q_4^{1/2} \sqrt{\sup_{x \in \mathbb{R}^d} \sup_{s < t \in [0,1]} \mathbb{E}^{a,x} \left(\exp \left(2\eta \frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^2}{t-s} \right) \right)} \\ &\leq 1 + \eta Q_4^{1/2} \sqrt{\sup_{x \in \mathbb{R}^d} \sup_{s < t \in [0,1]} \mathbb{E}^{a,x} \left(\exp \left(\frac{1}{8\Lambda} \frac{\mathrm{d}(\mathbf{X}_t, \mathbf{X}_s)^2}{t-s} \right) \right)} \end{aligned}$$

and Lemma 12. □

5 Approximations

5.1 Weak convergence

Theorem 15 *Let* (a_n) *be a sequence of* (*smooth*) *functions in* $\Xi(\Lambda)$ *such that* a_n *converges almost everywhere to* $a \in \Xi(\Lambda)$. *Then we have*

(i) *uniformly on compacts in* $(0,\infty) \times g^2(\mathbb{R}^d) \times g^2(\mathbb{R}^d)$,

$$p^{a_n}(t,x,y) \to p^a(t,x,y) \quad \text{as } n \to \infty;$$

(ii) *convergence in distribution* $\mathbf{X}^{a_n,x} \xrightarrow{\mathcal{D}} \mathbf{X}^{a,x}$ *with respect to uniform topology on* $\{\omega : C([0,1], g^2(\mathbb{R}^d)) : \omega(0) = x\}$, *with fixed* $x \in g^2(\mathbb{R}^d)$;

(iii) *the convergence in distribution remains valid with respect to homogenous* α*-Hölder topology of exponent for* $\alpha \in [0,1/2)$.

Proof The proof of (i) is identical to the proof of [28, Theorem II.3.1] and implies convergence of the finite-dimensional distributions. A standard tightness argument leads to (ii) and (iii). □

Remark 16 [16] discusses the case when $a(x)$ depends only on the projection $\pi_1(x) \in \mathbb{R}^d$.

5.2 Strong convergence

5.2.1 Geodesic approximations

Recall that $g^2(\mathbb{R}^d)$ equipped with Carnot–Caratheodory distance is a geodesic space. Given a dissection D of $[0,1]$ and a deterministic path $\mathbf{x} \in C^{\alpha-\text{Hölder}}([0,1], g^2(\mathbb{R}^d))$ we can approximate $\mathbf{x}$ by a path $\mathbf{x}^D \in C^{\text{Lip}}([0,1], g^2(\mathbb{R}^d))$ obtained by connecting the points $(\mathbf{x}_{t_i} : t_i \in D)$ with geodesics run at unit speed. If there are several geodesics between two points $\mathbf{x}_{t_i}$ and $\mathbf{x}_{t_{i+1}}$ it is immaterial which one is chosen. It is not hard to show that

$$\left\| \mathbf{x}^D \right\|_{\alpha-\text{Hölder}} \leq 3 \left\| \mathbf{x} \right\|_{\alpha-\text{Hölder}}. \tag{6}$$

Clearly, $\mathbf{x}^D \to \mathbf{x}$ pointwise as $|D| \to 0$ and, in fact, this convergence is uniform in view of the uniform bound (6). A simple interpolation argument then gives α'-Hölder convergence, $\alpha' \in (0,\alpha)$. All this results are purely deterministic and discussed in detail in [10]. By Theorem **13** these approximation results apply to, a.e., sample path of $\mathbf{X}^{a,x}$. We emphasize that these approximations required apriori knowledge of the area $\pi_2(\mathbf{X}^{a,x})$. In fact, $\pi_1(\mathbf{x}^D)$ is simply the concatenation of path segments designed to wipe out prescribed areas.

5.2.2 Piecewise linear approximations: Wong–Zakai

In contrast to geodesic approximation, convergence of piecewise linear approximations, based on the $\mathbb{R}^d$-valued path $\pi_1(\mathbf{X}^{a,x})$ alone and without apriori knowledge of the area $\pi_2(\mathbf{X}^{a,x})$, is a genuine probabilistic statement and relies on subtle cancellations. (An example by McShane, see [13], shows what can go wrong if one replaces linear cords by general interpolation functions.)

The Idea Fix a dissection $D = \{t_i : i\}$ of $[0, 1]$ and $a \in \Xi(\Lambda)$. Let us project $\mathbf{X} = \mathbf{X}^a$ to the $\mathbb{R}^d$-valued process $X = X^a$ and consider piecewise-linear approximations to X based on D, denoted by X^D. Of course, X^D has a canonically defined area given by the usual iterated integrals and thus gives rise to an $g^2(\mathbb{R}^d)$-valued path which we denote by $S(X^D)$. For $0 \le \alpha < 1/2$ as usual, the convergence

$$d_{\alpha-\text{Hölder}}\left(S\left(X^D\right), \mathbf{X}\right) \to 0 \text{ in probability} \tag{7}$$

as $|D| \to 0$ is a subtle problem and the difficulty is already present in the pointwise convergence statement $S\left(X^D\right)_{0,t} \to \mathbf{X}_{0,t}$ as $|D| \to 0$. Our idea is simple. Noting that straight line segments do not produce area, it is an elementary application of the Campbell–Baker–Hausdorff formula to see that for $t \in D = \{t_i\}$

$$\left(S\left(X^D\right)_{0,t}\right)^{-1} * \mathbf{X}_{0,t} = \sum_i A_{t_i,t_{i+1}}, \tag{8}$$

where A is the area of $\mathbf{X}$ and $\cup_i \left[t_i, t_{i+1}\right] = [0, t]$. On the other hand, it is relatively straight-forward to show that the L^p norm of $\left\|S\left(X^D\right)\right\|_{\alpha-\text{Höl};[0,1]}$ is finite uniformly over all D. In essence, this reduces (7) to the pointwise convergence statement which we can rephrase as $\sum_i A_{t_i,t_{i+1}} \to 0$. It is natural to show this in L^2 since this allows to write[4]

$$\mathbb{E}\left[\left|\sum_i A_{t_i,t_{i+1}}\right|^2\right] = \sum_i \mathbb{E}\left(\left|A_{t_i,t_{i+1}}\right|^2\right) + 2\sum_{i<j} \mathbb{E}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right).$$

For simplicity only, assume $t_{i+1} - t_i \equiv \delta$ for all i. As a sanity check, if X were a Brownian motion and A the usual Lévy area, all off-diagonal terms are zero and

$$\sum_i \mathbb{E}\left(\left|A_{t_i,t_{i+1}}\right|^2\right) \sim \sum_i \delta^2 \sim \frac{1}{\delta}\delta^2 \to 0 \quad \text{with } |D| = \delta \to 0$$

which is what we want. Back to the general case of $\mathbf{X} = \mathbf{X}^a$, the plan must be to cope with the off-diagonal sum. Since there are $\sim \delta^2/2$ terms what we need is

[4] We equip $so(d) \subset \mathbb{R}^d \otimes \mathbb{R}^d$ with the Euclidean structure $A \cdot \tilde{A} = \sum_{k,l=1}^d A^{k,l}\tilde{A}^{k,l}$ and $|A|^2 = A \cdot A$. It may be instructive to consider $d = 2$ in which case A can be viewed as scalar.

$\mathbb{E}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right) = o\left(\delta^2\right)$.To this end, let us momentarily assume that

$$\sup_x \mathbb{E}^{a,x}\left(A_{0,\delta}\right) = o\left(\delta\right). \tag{9}$$

holds. Then, using the Markov property,

$$\left|\mathbb{E}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right)\right| \le \mathbb{E}\left(\left|A_{t_i,t_{i+1}}\right| \times \left|\mathbb{E}^{\mathbf{X}_{t_j}} A_{0,\delta}\right|\right) = \mathbb{E}\left(\left|A_{t_i,t_{i+1}}\right|\right) \times o\left(\delta\right)$$

and since $\mathbb{E}\left(\left|A_{t_i,t_{i+1}}\right|\right) \sim \delta$, by a soft scaling argument, we are done. Unfortunately, (9) seems to be too strong to be true but we are able to establish a weak version of (9) which is good enough to successfully implement what we just outlined. The key to all this (cf. the proof of the forthcoming Proposition 18) is a semi-group argument which leads to the desired cancellations.

Uniform Hölder Bound Let X^D denote the piecewise linear approximation to $X = X(\omega)$. We now show $L^q\left(\mathbb{P}^{a,x}\right)$-bounds, uniformly over all dissections D, of the homogenous α-Hölder norm of the path X^D and its area.

Theorem 17 *There exists* $\eta = \eta\left(\Lambda\right) > 0$ *such that*

$$\sup_{a\in\Xi(\Lambda), x\in g^2(\mathbb{R}^d)} \sup_D \sup_{0\le s<t\le 1} \mathbb{E}^{a,x}\left(\exp\left(\eta \frac{\left\|S\left(X^D\right)_{s,t}\right\|^2}{t-s}\right)\right) < \infty.$$

As a consequence, for any $\alpha \in [0, 1/2)$ *there exists* $C_{17} = C_{17}\left(\alpha, \Lambda\right) > 0$ *so that*

$$\sup_{a\in\Xi(\Lambda), x\in g^2(\mathbb{R}^d)} \sup_D \mathbb{E}^{a,x}\left(\exp\left(C_{17}\left\|S\left(X^D\right)\right\|^2_{\alpha-\text{Höl};[0,1]}\right)\right) < \infty.$$

Proof The consequence is an immediate application of the Garsia–Rodemich–Rumsey lemma and we only have to discuss the first estimate. We remind the reader that from Lemma 12 for $\eta \in [0, \frac{1}{4\Lambda})$,

$$M_\eta \equiv \sup_{a\in\Xi(\Lambda), x\in g^2(\mathbb{R}^d)} \sup_{0\le s<t\le 1} \mathbb{E}^{a,x}\left(\exp\left(\eta \frac{\left\|\mathbf{X}_{s,t}\right\|^2}{t-s}\right)\right) < \infty.$$

By the triangle inequality (recall t_D, t^D were defined at the end of the Introduction)

$$\begin{aligned}\frac{\left\| S\left(X^D\right)_{s,t}\right\|}{\sqrt{t-s}} &\leq \frac{\left\| S\left(X^D\right)_{s,s^D}\right\|}{\sqrt{s^D-s}} + \frac{\left\| S\left(X^D\right)_{s^D,t_D}\right\|}{\sqrt{t_D-s^D}} + \frac{\left\| S\left(X^D\right)_{t_D,t}\right\|}{\sqrt{t-t_D}} \\ &\leq \frac{\left| X^D_{s,s^D}\right|}{\sqrt{s^D-s}} + \frac{\left\| S\left(X^D\right)_{s^D,t_D}\right\|}{\sqrt{t_D-s^D}} + \frac{\left| X^D_{t_D,t}\right|}{\sqrt{t-t_D}} \\ &\leq \frac{\left\| \mathbf{X}_{s,s^D}\right\|}{\sqrt{s^D-s}} + \frac{\left\| S\left(X^D\right)_{s^D,t_D}\right\|}{\sqrt{t_D-s^D}} + \frac{\left\| \mathbf{X}_{t_D,t}\right\|}{\sqrt{t-t_D}} \\ &\leq \left(\frac{3\left\| \mathbf{X}_{s,s^D}\right\|^2}{s^D-s} + \frac{3\left\| S\left(X^D\right)_{s^D,t_D}\right\|^2}{t_D-s^D} + \frac{3\left\| \mathbf{X}^x_{t_D,t}\right\|^2}{t-t_D}\right)^{1/2}.\end{aligned}$$

Hence

$$\begin{aligned}&\mathbb{E}^{a,x}\left(\exp\left(\eta\frac{\left\| S\left(X^D\right)_{s,t}\right\|^2}{t-s}\right)\right) \\ &\leq \mathbb{E}^{a,x}\left\{\exp\left[\eta\left(\frac{3\left\| \mathbf{X}_{s,s^D}\right\|^2}{s^D-s} + \frac{3\left\| S\left(X^D\right)_{s^D,t_D}\right\|^2}{t_D-s^D} + \frac{3\left\| \mathbf{X}^x_{t_D,t}\right\|^2}{t-t_D}\right)\right]\right\} \\ &\leq M^2_{6\eta}\mathbb{E}^{a,x}\left(\exp\left(6\eta\frac{\left\| S\left(X^D\right)_{s^D,t_D}\right\|^2}{t_D-s^D}\right)\right)\end{aligned}$$

and the proof is reduced to show that for some $\eta > 0$ small enough

$$\sup_{a\in\Xi(\Lambda),x\in g^2(\mathbb{R}^d)} \sup_D \sup_{s<t\in D} \mathbb{E}^{a,x}\left(\exp\left(6\eta\frac{\left\| S\left(X^D\right)_{s,t}\right\|^2}{t-s}\right)\right) < \infty.$$

By the triangle inequality for the Carnot–Caratheodory distance, for $t_i, t_j \in D$,

$$\left\| S\left(X^D\right)_{t_i,t_j}\right\| \leq \left\| \mathbf{X}_{t_i,t_j}\right\| + d\left(\mathbf{X}_{t_i,t_j}, S\left(X^D\right)_{t_i,t_j}\right).$$

To proceed we note that, similar to equation (8),

$$\left(S\left(X^D\right)_{t_i,t_j}\right)^{-1} * \mathbf{X}_{t_i,t_j} = \sum_{k=i}^{j-1} A_{t_k,t_{k+1}}.$$

By left-invariance of the Carnot–Caratheodory distance d and equivalence of continuous homogenous norms (so that, in particular, $\|(x, A)\| \sim |x| + |A|^{1/2}$ where $|\cdot|$ denotes Euclidean norm on $\mathbb{R}^d$ resp. $\mathbb{R}^d \otimes \mathbb{R}^d$) there exists C such that

$$\begin{aligned} d\left(\mathbf{X}_{t_i,t_j}, S\left(X^D\right)_{t_i,t_j}\right) &= \left\|\left(0, \sum_{k=i}^{j-1} A_{t_k,t_{k+1}}\right)\right\| \\ &\leq C\left|\sum_{k=i}^{j-1} A_{t_k,t_{k+1}}\right|^{1/2} \leq C\sqrt{\sum_{k=i}^{j-1}\left|A_{t_k,t_{k+1}}\right|} \\ &\leq C\sqrt{\sum_{k=i}^{j-1}\left\|\mathbf{X}_{t_k,t_{k+1}}\right\|^2}. \end{aligned}$$

By Cauchy–Schwartz,

$$\begin{aligned} &\mathbb{E}^{a,x}\left(\exp\left(6\eta\frac{\left\|S\left(X^D\right)_{t_i,t_j}\right\|^2}{t_j - t_i}\right)\right) \\ &\leq \mathbb{E}^{a,x}\left(\exp\left(12\eta\frac{\left\|\mathbf{X}_{t_i,t_j}\right\|^2}{t_j - t_i}\right)\exp\left(12C\eta\frac{\sum_{k=i}^{j-1}\left\|\mathbf{X}_{t_k,t_{k+1}}\right\|^2}{t_j - t_i}\right)\right) \\ &\leq M_{24\eta}\mathbb{E}^{a,x}\left(\prod_{k=i}^{j-1}\exp\left(24C\eta\frac{\left\|\mathbf{X}_{t_k,t_{k+1}}\right\|^2}{t_j - t_i}\right)\right). \end{aligned}$$

and the $\mathbb{E}^{a,x}(\cdots)$ term in the last line is estimated using the Markov property as follows.

$$\begin{aligned} &\mathbb{E}^{a,x}\left(\prod_{k=i}^{j-1}\exp\left(24C\eta\frac{\left\|\mathbf{X}_{t_k,t_{k+1}}\right\|^2}{t_j - t_i}\right)\right) \\ &\leq \prod_{k=i}^{j-1}\sup_{x\in\mathbb{R}^d}\mathbb{E}\left(\exp\left(24C\eta\frac{t_{k+1}-t_k}{t_j-t_i}\frac{\left\|\mathbf{X}^x_{0,t_{k+1}-t_k}\right\|^2}{t_{k+1}-t_k}\right)\right) \\ &\leq \prod_{k=i}^{j-1} M_{24C\eta\frac{t_{k+1}-t_k}{t_j-t_i}} \\ &\leq \prod_{k=i}^{j-1}\exp\left(C_{14}\times 24C\eta\frac{t_{k+1}-t_k}{t_j-t_i}\right) \quad \text{for } \eta \text{ small enough} \\ &= \exp\left(24C_{14}C\eta\right) < \infty. \end{aligned}$$

where we used Corollary 14, valid for η small enough. The proof is finished. □

The Subtle Cancellation Let us define

$$r_\delta(t,x) = \frac{1}{\delta}\mathbb{E}^{a,x}\left(A_{t,t+\delta}\right) \in so(d) \quad \text{and} \quad r_\delta(x) = r_\delta(0,x).$$

For instance, (9) is now expressed as $\lim_{\delta\to 0} r_\delta(x) \to 0$ uniformly in x. Our goal here is to establish a weak version of this. We also recall that

$$A_{t,t+\delta} = \pi_2\left(\mathbf{X}_{t,t+\delta}\right) = \pi_2\left(\mathbf{X}_t^{-1} * \mathbf{X}_{t+\delta}\right).$$

Proposition 18 (i) *We have uniform boundedness of* $r_{\delta;t}(x)$,

$$\sup_{x\in g^2(\mathbb{R}^d)} \sup_{\delta\in[0,1]} \sup_{t\in[0,1-\delta]} r_\delta(t,x) < \infty.$$

(ii) *For all* $h \in L^1\left(g^2\left(\mathbb{R}^d\right), \mathrm{d}x\right)$,

$$\lim_{\delta\to 0} \int_{g^2(\mathbb{R}^d)} \mathrm{d}x h(x) r_\delta(x) \equiv 0.$$

Proof (i) follows from Lemma 12. For (ii) we may consider h smooth and compactly supported. Now the problem is local and we can assume that smooth locally bounded functions such as the coordinate projections $\pi_{1;j}$ and $\pi_{2;k,l}$ are in $D(\mathcal{E}^a)$. (More formally, we could smoothly truncate outside the support of h and work on a big torus). Clearly, it is enough to show the componentwise statement

$$\lim_{\delta\to 0} \int_{g^2(\mathbb{R}^d)} \mathrm{d}x h(x) \pi_{2;k,l}\left(r_\delta(x)\right) \equiv 0$$

for $k < l$ fixed in $\{1,\ldots,d\}$. To keep notation short we set $f \equiv \pi_{2;k,l}(\cdot)$ and abuse notation by writing A instead of $A^{k,l}$. We can then write

$$\mathbb{E}^{a,\cdot}(A_t) \equiv \mathbb{E}^{a,\cdot}(f(\mathbf{X}_t)) =: P_t^a f(.)$$

and note that $P_0^a f(x) = A$ when $x = \left(x^1, A\right) \in g^2\left(\mathbb{R}^d\right)$. Writing $\langle\cdot,\cdot\rangle$ for the usual inner product on $L^2\left(g^2\left(\mathbb{R}^d\right), \mathrm{d}x\right)$ we have

$$\begin{aligned}\left\langle h, \mathbb{E}^{a,\cdot} A_{0,t}\right\rangle &= \left\langle h, \mathbb{E}^{a,\cdot} f\left(\mathbf{X}_t\right) - A - \frac{1}{2}\mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]\right)\right\rangle \\ &= \left\langle h, P_t^a f - P_0^a f\right\rangle - \left\langle h, \frac{1}{2}\mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]\right)\right\rangle \\ &= \int_0^t \mathcal{E}^a\left(h, P_s^a f\right) - \left\langle h, \frac{1}{2}\mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]\right)\right\rangle \\ &= \mathcal{E}^a\left(h, f\right) \times t - \left\langle h, \frac{1}{2}\mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]\right)\right\rangle + o\left(t\right).\end{aligned}$$

Here, again, we abused notation by writing $[\cdot,\cdot]$ instead of picking out the (k,l) component and using the cumbersome notation $[\cdot,\cdot]^{k,l}$. Note that in general $\mathcal{E}^a\left(h, f\right) \times t \neq o\left(t\right)$ and our only hope is cancellation of $2\mathcal{E}^a\left(h, f\right)$ with the bracket term

$$\left\langle h, \mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]\right)\right\rangle \equiv \left\langle h, \mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]^{k,l}\right)\right\rangle.$$

To see this cancellation, we compute the bracket term,

$$\begin{aligned}\left\langle h, \mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]^{k,l}\right)\right\rangle &= \int \mathrm{d}x\, h(x)\mathbb{E}^{a,x}\left(x^{1;k}\mathbf{X}_t^{1;l} - x^{1;l}\mathbf{X}_t^{1;k}\right) \\ &= \int \mathrm{d}x\, h(x)\left(\left(x^{1;k}\left[P_t^a \pi_{1;l}\right](x) - x^{1;l}\left[P_t^a \pi_{1;k}\right](x)\right)\right),\end{aligned}$$

and by adding and subtracting $x^{1;k}x^{1;l}$ inside the integral this rewrites as

$$\int \mathrm{d}x\, h(x)x^{1;k}\left\{\left[P_t^a \pi_{1;l}\right](x) - \pi_{1;l}(x)\right\} - \int \mathrm{d}x\, h(x)x^{1;l}\left\{\left[P_t^a \pi_{1;k}\right](x) - \pi_{1;k}(x)\right\}.$$

It now follows as earlier that

$$\left\langle h, \mathbb{E}^{a,\cdot}\left([\cdot, \mathbf{X}_t^1]^{k,l}\right)\right\rangle = \left[\mathcal{E}^a\left(h\pi_{1;k}, \pi_{1;l}\right) - \mathcal{E}^a\left(h\pi_{1;l}, \pi_{1;k}\right)\right] \times t + o\left(t\right)$$

and we see that the required cancellation takes place if, for all h smooth and compactly supported,

$$\left[\mathcal{E}^a\left(h\pi_{1;k}, \pi_{1;l}\right) - \mathcal{E}^a\left(h\pi_{1;l}, \pi_{1;k}\right)\right] \equiv 2\mathcal{E}^a\left(h, \pi_{2;k,l}\right).$$

We will check this with a direct computation. First note that

$$\mathcal{E}^a\left(h\pi_{1;k}, \pi_{1;l}\right) - \mathcal{E}^a\left(h\pi_{1;l}, \pi_{1;k}\right) = \int \pi_{1,k}\mathrm{d}\Gamma^a\left(h, \pi_{1,l}\right) - \int \pi_{1,l}\mathrm{d}\Gamma^a\left(h, \pi_{1,k}\right)$$

which is immediately seen via symmetry of $\mathrm{d}\Gamma^a(\cdot,\cdot)$, inherited from the symmetric of $\left(a^{ij}\right)$, and the Leibnitz formula

$$\mathcal{E}^a\left(gg',h\right)=\int g\mathrm{d}\Gamma^a\left(g',h\right)+\int g'\mathrm{d}\Gamma^a\left(g,h\right).$$

It is immediately checked from the definition of the vector fields U_i, see equation (3), that

$$U_i f\equiv U_i\pi_{2;k,l}=\begin{cases}-(1/2)\,\pi_{1;l} & \text{if } i=k\\ (1/2)\,\pi_{1;k} & \text{if } i=l\\ 0 & \text{otherwise}\end{cases}$$

so that

$$\int\pi_{1,k}\mathrm{d}\Gamma^a\left(h,\pi_{1,l}\right)=\sum_{i,j}\int\pi_{1,k}a^{ij}U_ihU_j\pi_{1,l}=2\sum_i\int\left(U_lf\right)a^{il}\left(U_ih\right)$$

and similarly

$$-\int\pi_{1,l}\mathrm{d}\Gamma^a\left(h,\pi_{1,k}\right)=\sum_{i,j}\int\left(-\pi_{1,l}\right)a^{ij}U_ihU_j\pi_{1,k}=2\sum_i\int\left(U_kf\right)a^{ik}\left(U_ih\right).$$

Therefore, using $U_jf=0$ for $j\neq\{k,l\}$ in the second equality,

$$\begin{aligned}\mathcal{E}^a\left(h\pi_{1;k},\pi_{1;l}\right)-\mathcal{E}^a\left(h\pi_{1;l},\pi_{1;k}\right)&=2\sum_{j=k,l}\sum_i\int\left(U_jf\right)a^{ij}\left(U_ih\right)\\&=2\sum_{i,j}\int\left(U_jf\right)a^{ij}\left(U_ih\right)\end{aligned}$$

and this equals precisely $2\mathcal{E}^a(h,f)$ as required. □

Corollary 19 *For all $t\in[0,1)$ and all $h\in L^1\left(g^2\left(\mathbb{R}^d\right),\mathrm{d}x\right)$,*

$$\lim_{\delta\to 0}\int_{g^2(\mathbb{R}^d)}\mathrm{d}x h(x)\mathbb{E}^{a,x}\left(\frac{A_{t,t+\delta}}{\delta}\right)\equiv 0.$$

Proof We first write

$$\begin{aligned}\int\mathrm{d}x h(x)\mathbb{E}^{a,x}\left(\frac{A_{t,t+\delta}}{\delta}\right)&=\int\int h(x)p^a(t,x,y)r_\delta(y)\mathrm{d}x\,\mathrm{d}y\\&=\int\left(\int h(x)p^a(t,x,y)\mathrm{d}x\right)r_\delta(y)\mathrm{d}y.\end{aligned}$$

Then, noting that $y \mapsto \int h(x) p_t(x, y) \mathrm{d}x$ is in $L^1\left(g^2\left(\mathbb{R}^d\right), \mathrm{d}x\right)$, the proof is finished by applying the previous proposition. □

Theorem 20 *For all bounded sets* $K \subset g^2\left(\mathbb{R}^d\right)$ *and all* $\sigma \in (0, 1]$,

$$\lim_{\delta \to 0} \sup_{t \in [\sigma, 1]} \sup_{y \in K} \left| \mathbb{E}^{a,y}\left(\frac{A_{t,t+\delta}}{\delta}\right)\right| = 0.$$

Proof It suffices to prove this for a compact ball $K = \bar{B}(0, R) \subset g^2\left(\mathbb{R}^d\right)$ of arbitrary radius $R > 0$. We fix $\sigma \in (0, 1]$ and think of $r_\delta = r_\delta(t, y)$ as a family of maps, indexed by $\delta > 0$, defined on the cylinder $[\sigma, 1] \times K$, that is

$$(t, y) \in [\sigma, 1] \times K \mapsto r_\delta(t, y) \in so\,(d).$$

By Proposition 18, (i) we know that $\sup_{\delta > 0} |r_\delta|_\infty < \infty$. We now show equicontinuity of $\{r_\delta : \delta > 0\}$. By the Markov property, $r_\delta(t, y)$ equals

$$\mathbb{E}^{a,y}\left(\frac{A_{t,t+\delta}}{\delta}\right) = \left\langle p^a\,(t, y, \cdot), \frac{\mathbb{E}^{a,\cdot}\left(A_{0,\delta}\right)}{\delta}\right\rangle = \left\langle p^a\,(t, y, \cdot), r_\delta\,(0, \cdot)\right\rangle,$$

so that, for all $(s, x), (t, y) \in [\sigma, 1] \times K$,

$$\begin{aligned} |r_\delta\,(s, x) - r_\delta(t, y)| &= \left|\left\langle p^a\,(s, x, \cdot) - p^a\,(t, y, \cdot), r_\delta\,(.)\right\rangle\right| \\ &\le \left(\sup_{\delta \in (0,1]} |r_\delta|_\infty\right) \left|p^a\,(s, x, \cdot) - p^a\,(t, y, \cdot)\right|_{L^1}. \end{aligned}$$

From Proposition 6, $(t, y) \in [\sigma, 1] \times K \mapsto p^a\,(t, y, z)$ is continuous for all z ; the dominated convergence theorem then gives easily continuity of $(t, y) \mapsto p^a\,(t, y, \cdot) \in L^1$. In fact, this map is uniformly continuous when restricted to the compact $[\sigma, 1] \times K$ and it follows that $\{r_\delta : \delta > 0\}$ is equicontinuous as claimed. By Arzela–Ascoli, there exists a subsequence (δ^n) such that r_{δ^n} converges uniformly on $[\sigma, 1] \times K$ to some (continuous) function r. On the other hand, Proposition 18, (ii), applied to $h = p^a\,(t, y, \cdot)$, shows that $r_\delta(t, y) \to 0$ as $\delta \to 0$ for all fixed $y, t > 0$. This shows that $r \equiv 0$ is the only limit point and hence

$$\lim_{\delta \to 0} \sup_{t \in [\sigma, 1]} \sup_{y \in K} \left| \mathbb{E}^{a,y}\left(\frac{A_{t,t+\delta}}{\delta}\right)\right| = 0.$$

□

Convergence of the Sum of the Small Areas For fixed $a \in \Xi(\Lambda)$ and $x \in g^2(\mathbb{R}^d)$ let us define the real-valued quantity

$$K_{\sigma,\delta} := \sup_{\substack{0 \le u_1 < u_2 < v_1 < v_2 \le 1: \\ v_1 - u_2 \ge \sigma, \\ |u_2 - u_1|, |v_2 - v_1| \le \delta}} \frac{\left|\mathbb{E}^{a,x}\left(A_{u_1,u_2} \cdot A_{v_1,v_2}\right)\right|}{(u_2 - u_1)(v_2 - v_1)}$$

where $\delta, \sigma \in (0, 1)$. As above $\cdot$ denotes the scalar product in $so(d)$.

Proposition 21 *For fixed* $\sigma \in (0, 1)$, $k, l \in \{1, \ldots, d\}$ *we have* $\lim_{\delta \to 0} K_{\sigma,\delta} = 0$.

Proof By the Markov property,

$$\begin{aligned}
\frac{\left|\mathbb{E}^{a,x}\left(A_{u_1,u_2} \cdot A_{v_1,v_2}\right)\right|}{(u_2 - u_1)(v_2 - v_1)} &= \frac{\left|\mathbb{E}^{a,x}\left(A_{u_1,u_2} \cdot \mathbb{E}^{a,\mathbf{X}_{u_2}}\left(A_{v_1-u_2,v_2-u_2}\right)\right)\right|}{(u_2 - u_1)(v_2 - v_1)} \\
&\le \frac{\left|\mathbb{E}^{a,x}\left(A_{u_1,u_2} \cdot \mathbb{E}^{a,\mathbf{X}_{u_2}}\left(A_{v_1-u_2,v_2-u_2}; \left\|\mathbf{X}_{u_2}\right\| \le R\right)\right)\right|}{(u_2 - u_1)(v_2 - v_1)} \\
&\quad + \frac{\left|\mathbb{E}^{a,x}\left(A_{u_1,u_2} \cdot \mathbb{E}^{a,\mathbf{X}_{u_2}}\left(A_{v_1-u_2,v_2-u_2}; \left\|\mathbf{X}_{u_2}\right\| > R\right)\right)\right|}{(u_2 - u_1)(v_2 - v_1)} \\
&\le \frac{\mathbb{E}^{a,x}\left(\left|A_{u_1,u_2}\right|; \left\|\mathbf{X}_{u_2}\right\| \le R\right)}{(u_2 - u_1)} \sup_{\delta' \le \delta} \sup_{\substack{\|y\| \le R \\ u \in [\sigma,1]}} \frac{\left|\mathbb{E}^{a,y}\left(A_{u,u+\delta'}\right)\right|}{\delta'} \\
&\quad + \mathbb{E}^{a,x}\left(\frac{\left|A_{u_1,u_2}\right|}{u_2 - u_1}; \left\|\mathbf{X}_{u_2}\right\| > R\right) \sup_{\delta',u,x} \frac{\mathbb{E}^{a,x}\left(\left|A_{u,u+\delta'}\right|\right)}{\delta'}. \\
&\le \frac{\mathbb{E}^{a,x}\left(\left|A_{u_1,u_2}\right|\right)}{(u_2 - u_1)} \sup_{\delta' \le \delta} \sup_{\substack{\|y\| \le R \\ u \in [\sigma,1]}} \frac{\left|\mathbb{E}^{a,y}\left(A_{u,u+\delta'}\right)\right|}{\delta'} \\
&\quad + \sqrt{\mathbb{P}^{a,x}\left(\left\|\mathbf{X}_{u_2}\right\| > R\right)} \sqrt{\mathbb{E}^{a,x}\left(\left|\frac{A_{u_1,u_2}}{u_2 - u_1}\right|^2\right)} \\
&\quad \times \sup_{\delta',u,x} \frac{\mathbb{E}^{a,x}\left(\left|A_{u,u+\delta'}\right|\right)}{\delta'} \\
&\le C \sup_{\delta' \le \delta} \sup_{\substack{|y| \le R \\ u \in [\sigma,1]}} \frac{\left|\mathbb{E}^{a,y}\left(A_{u,u+\delta'}\right)\right|}{\delta'} + C\sqrt{\mathbb{P}^{a,x}\left(\left\|\mathbf{X}_{u_2}\right\| > R\right)}
\end{aligned}$$

for some constant $C = C(\|x\|, \sigma, \Lambda)$ using Lemma 12 and Proposition 18, (i). We then fix $\varepsilon > 0$ and choose $R = R(\epsilon)$ large enough so that

$$C \sup_{u_2 \in [0,1]} \sqrt{\mathbb{E}^{a,x}\left(\left|\mathbf{X}_{u_2}\right| > R\right)} \le \varepsilon/2.$$

On the other hand, Theorem 20 shows that

$$C \sup_{\delta' \leq \delta} \sup_{\substack{|y| \leq R \\ u \in [\sigma,1]}} \frac{\left|\mathbb{E}^{a,y}\left(A_{u,u+\delta'}\right)\right|}{\delta'} \leq \frac{\varepsilon}{2}$$

for all δ small enough and the proof is finished. □

Corollary 22 *There exists* $C_{22} = C_{22}(\Lambda)$ *such that for all subdivisions* D *of* $[0, 1]$, $s, t \in D$, *for any* $\sigma \in (0, 1)$,

$$\mathbb{E}^{a,x}\left(\left|d\left(S\left(X^D\right)_{s,t}, \mathbf{X}_{s,t}\right)\right|^4\right) \leq C_{22}\left[(t-s)^2 K_{\sigma,|D|} + (t-s)\,\sigma\right].$$

Proof Recalling the discussion around (8), equivalence of homogenous norms leads to

$$\mathbb{E}^{a,x}\left(\left|d\left(S\left(X^D\right)_{s,t}, \mathbf{X}_{s,t}\right)\right|^4\right) \leq C\mathbb{E}^{a,x}\left(\left|\sum_{i:t_i \in D\cap[s,t)} A_{t_i,t_{i+1}}\right|^2\right).$$

Let us abbreviate $\sum_{i:t_i\in D\cap[s,t)}$ to $\sum_i$ in what follows. Clearly, $\mathbb{E}^{a,x}(|\sum_i A_{t_i,t_{i+1}}|^2)$ is estimated by 2 times

$$\begin{aligned}
&\sum_{i\leq j} \mathbb{E}^{a,x}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right)\\
&\quad\leq \sum_{\substack{i \leq j \\ t_j - t_{i+1} \geq \sigma}} \mathbb{E}^{a,x}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right) + \sum_{\substack{i \leq j \\ t_j - t_{i+1} < \sigma}} \mathbb{E}^{a,x}\left(A_{t_i,t_{i+1}} \cdot A_{t_j,t_{j+1}}\right)\\
&\quad\leq K_{\sigma,|D|} \sum_{\substack{i \leq j \\ t_j - t_{i+1} \geq \sigma}} (t_{i+1} - t_i)\left(t_{j+1} - t_j\right)\\
&\qquad + \sum_{\substack{i \leq j \\ t_j - t_{i+1} < \sigma}} \sqrt{\mathbb{E}^{a,x}\left(\left|A_{t_i,t_{i+1}}\right|^2\right)\mathbb{E}^{a,x}\left(\left|A_{t_j,t_{j+1}}\right|^2\right)}\\
&\quad\leq K_{\sigma,|D|}(t-s)^2 + C \sum_{\substack{i,j \\ t_j - t_{i+1} < \sigma}} (t_{i+1} - t_i)\left(t_{j+1} - t_j\right)
\end{aligned}$$

and the very last sum is estimated as follows,

$$\left|\sum_i (t_{i+1} - t_i) \sum_{\substack{j \\ t_j - t_{i+1} < \sigma}} \left(t_{j+1} - t_j\right)\right| \leq \sigma \sum_i (t_{i+1} - t_i) = \sigma\,(t-s).$$

The proof is finished. □

Putting Things Together

Theorem 23 *Let D be a dissection of $[0,1]$ with mesh $|D|$. Then, for all $1 \leq q < \infty$ and $0 \leq \alpha < 1/2$,*

$$d_{\alpha-\text{Hölder}}\left(S\left(X^D\right),\mathbf{X}\right) \rightarrow 0 \text{ in } L^q\left(\mathbb{P}^{a,x}\right) \text{ as } |D| \rightarrow 0.$$

Proof We first show pointwise convergence. We fix $\varepsilon > 0$ and apply Corollary 22 with $\sigma = \varepsilon/2C$. Then,

$$\sup_{s,t\in D:s<t} \mathbb{E}^{a,x}\left(\left|d\left(S\left(X^D\right)_{s,t},\mathbf{X}_{s,t}\right)\right|^4\right) \leq CK_{\sigma,|D|} + \frac{\varepsilon}{2}$$

By Proposition 21 it then follows that, for $|D|$ small enough,

$$\sup_{s,t\in D:s<t} \left\|d\left(S\left(X^D\right)_{s,t},\mathbf{X}_{s,t}\right)\right\|^4_{L^4(\mathbb{P}^{a,x})} \leq \varepsilon.$$

By Theorem 17 we have for all $q \in [1,\infty)$,

$$\sup_D \left\|\left\|S\left(X^D\right)\right\|_{\alpha-\text{Hölder}}\right\|_{L^q(\mathbb{P}^{a,x})} + \left\|\|\mathbf{X}\|_{\alpha-\text{Hölder}}\right\|_{L^q(\mathbb{P}^{a,x})} < \infty \tag{10}$$

and both results combined yield

$$\lim_{|D|\rightarrow 0} \sup_{0\leq s<t\leq 1} \left\|d\left(S\left(X^D\right)_{s,t},\mathbf{X}_{s,t}\right)\right\|_{L^4(\mathbb{P}^{a,x})} = 0$$

and by Hölder's inequality the last statement remains valid even when we replace L^4 by L^q for any $q \in [1,\infty)$. Now, for every $m > 0$,

$$\begin{aligned}
&\mathbb{E}^{a,x}\left(d_\infty\left(S\left(X^D\right),\mathbf{X}\right)^q\right) \\
&\leq c_q \mathbb{E}^{a,x}\left(\sup_{1\leq i\leq m} d\left(S\left(X^D\right)_{\frac{i}{m}},\mathbf{X}_{\frac{i}{m}}\right)^q\right) \\
&\quad + c_q \mathbb{E}^{a,x}\left(\sup_{|t-s|<\frac{1}{m}}\left(\left\|S\left(X^D\right)_{s,t}\right\|^q + \|\mathbf{X}_{s,t}\|^q\right)\right) \\
&\leq c_q m \sup_{0\leq t\leq 1}\left\|d\left(S\left(X^D\right)_t,\mathbf{X}_t\right)\right\|^q_{L^q(\mathbb{P}^{a,x})} \\
&\quad + c_q\left(\frac{1}{m}\right)^{\alpha q}\mathbb{E}^{a,x}\left(\left(\left\|S\left(X^D\right)\right\|^q_{\alpha-\text{Hölder}} + \|\mathbf{X}\|^q_{\alpha-\text{Hölder}}\right)\right) \\
&\leq c_q m \sup_{0\leq t\leq 1}\left\|d\left(S\left(X^D\right)_t,\mathbf{X}_t\right)\right\|^q_{L^q(\mathbb{P}^{a,x})} + C\left(\frac{1}{m}\right)^{\alpha q}.
\end{aligned}$$

By choosing first m large enough and then D with $|D|$ small enough we see that $d_\infty\left(S\left(X^D\right),\mathbf{X}\right)\to 0$ in L^q as $|D|\to 0$, for all $q<\infty$. An easy application of the Campell–Hausdorff formula gives a d_0/d_∞-estimate,

$$\forall \mathbf{x},\mathbf{y}\in C\left([0,1],g^2\left(\mathbb{R}^d\right)\right):d_0(\mathbf{x},\mathbf{y})\le d_\infty(\mathbf{x},\mathbf{y})+C\sqrt{\|\mathbf{y}\|_\infty d_\infty(\mathbf{x},\mathbf{y})}.$$

With Cauchy–Schwarz and a standard Hölder interpolation argument, using (10) with $\alpha'\in(\alpha,1/2)$, we then see that

$$d_{\alpha-\text{Hölder}}\left(S\left(X^D\right),\mathbf{X}\right)\to 0 \text{ in } L^q\left(\mathbb{P}^{a,x}\right)\quad \text{as } |D|\to 0.$$

□

Remark 24 This convergence result implies that $\sigma\left(A_{s,t}:u\le s\le t\le v\right)\subset\mathcal{F}_{s,t}=\sigma\left(X_{s,r}:s\le r\le t\right)$ where $X=\pi_1(\mathbf{X})$ and $X_{s,r}=X_r-X_s\in\mathbb{R}^d$.

Corollary 25 *Let $Y=\pi(0,y_0;\mathbf{X})\equiv\pi(\mathbf{X})$ denote the $\mathbb{R}^e$-valued (random) RDE solution driven by $\mathbf{X}^{a,x}$ along fixed* Lip$^\gamma$ *vector fields $V_1,\dots,V_d$ on $\mathbb{R}^e$, with $\gamma>2$, and started at time* 0 *from y_0 fixed. Let $Y^D=\pi\left(0,y_0,X^D\right)$ be the piecewise smooth solution to corresponding control ODE*

$$\mathrm{d}Y^D=\sum_{i=1}^d V_i\left(Y^D\right)\mathrm{d}X^{D;i}.$$

Then for any $\alpha\in[0,1/2)$ we have $\left|Y-Y^D\right|_{\alpha-\text{Höl};[0,1]}\to 0$ in $L^q\left(\mathbb{P}^{a,x}\right)$, for all $q<\infty$.

Proof The universal limit theorem [17,19] shows immediately that

$$\left|Y-Y^D\right|_{\alpha-\text{Höl};[0,1]}\to 0 \text{ in probability.}$$

It then suffices to remark that the estimates on the Itô-Lyons map in [11] combined with Theorem 17 show that for all $q<\infty$,

$$\sup_D\mathbb{E}\left|Y^D\right|^q_{\alpha-\text{Höl};[0,1]},\quad |Y|^q_{\alpha-\text{Höl};[0,1]}<\infty.$$

□

6 RDE solutions as Markov processes

The following is an immediate consequence of the stochastic Taylor formula for random RDEs [11].

Lemma 26 *Let $\alpha \in (1/3, 1/2)$, $N = 2$. Assume the random rough path $\mathbf{X}$ is such that $\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} < \infty$ has a Gauss tail and let Z denote the random RDE solution driven by $\mathbf{X}$ along fixed Lip^γ vector fields $V_1, \ldots, V_d$, with $\gamma > 2$, and started from z. Then for all $f \in C_b^\infty$ we have*

$$\begin{aligned}
\mathbb{E}[f(Z_t)] = f(z) &+ \sum_{i=1}^d V_i f(z)\,\mathbb{E}\left[\pi_{1,i}\left(\mathbf{X}_{0,t}\right)\right] \\
&+ \frac{1}{2}\sum_{i,j=1}^d V_i V_j f(z)\mathbb{E}\left[\pi_{1,i}\left(\mathbf{X}_{0,t}\right)\pi_{1,j}\left(\mathbf{X}_{0,t}\right)\right] \\
&+ \frac{1}{2}\sum_{i,j=1}^d \left[V_i, V_j\right] f(z)\mathbb{E}\left[\pi_{2,i,j}\left(\mathbf{X}_{0,t}\right)\right] + \mathbb{E}\left[R_2(t, f)\right].
\end{aligned}$$

with remainder term,

$$\mathbb{E}\left[\left|R_2(t, f)\right|\right] = o(t) \quad \text{as } t \to 0.$$

As earlier, $\langle \cdot, \cdot \rangle$ denotes the inner product on $L^2\left(g^2\left(\mathbb{R}^d\right), m\right)$.

Lemma 27 *Let g be a compactly supported smooth function. Then, for all $k, l \in \{1, \ldots, d\}$,*

$$\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle = \sum_{j=1}^d \int_{g^2(\mathbb{R}^d)} a^{kj}(y) U_j g(y)\mathrm{d}y,$$

$$\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\pi_{1,l}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle = -2\int_{g^2(\mathbb{R}^d)} a^{kl}(y) g(y)\mathrm{d}y,$$

$$\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{2,i,j}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle = 0.$$

Proof Third equality was shown in Proposition 18. For the first statement, almost by definition of $\mathcal{E}^a$,

$$\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle = \mathcal{E}^a\left(\pi_{1,k}, g\right) = \sum_{j=1}^d \int_{g^2(\mathbb{R}^d)} a^{kj}(y) U_j g(y)\mathrm{d}y.$$

Let us now consider the second equality. First rewrite $\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\pi_{1,l}\left(\mathbf{X}_{0,t}\right)$ as

$$\begin{aligned}
&\pi_{1,k}(\mathbf{X}_t)\,\pi_{1,l}(\mathbf{X}_t) - \pi_{1,k}(\mathbf{X}_0)\,\pi_{1,l}(\mathbf{X}_0) - \pi_{1,k}(\mathbf{X}_0)\,\pi_{1,l}\left(\mathbf{X}_{0,t}\right) \\
&\quad -\pi_{1,l}(\mathbf{X}_0)\,\pi_{1,k}\left(\mathbf{X}_{0,t}\right).
\end{aligned}$$

Then, by a similar argument as above,

$$\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\pi_{1,l}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle = \mathcal{E}^a\left(\pi_{1,k}\pi_{1,l}, g\right) - \mathcal{E}^a\left(\pi_{1,k}, \pi_{1,l}g\right) - \mathcal{E}^a\left(\pi_{1,l}, \pi_{1,k}g\right).$$

By the Leibniz formula, recalling that $\mathrm{d}\Gamma^a\left(f, f'\right) \equiv \left(\sum_{i,j} a^{ij} U_i f U_j f'\right)\mathrm{d}m$ is the energy measure of $\mathcal{E}^a$, we have

$$\begin{aligned}
\mathcal{E}^a\left(\pi_{1,k}\pi_{1,l}, g\right) &= \int \pi_{1,l}\mathrm{d}\Gamma^a\left(\pi_{1,k}, g\right) + \int \pi_{1,k}\mathrm{d}\Gamma^a\left(\pi_{1,l}, g\right)\\
\mathcal{E}^a\left(\pi_{1,k}, \pi_{1,l}g\right) &= \int \pi_{1,l}\mathrm{d}\Gamma^a\left(\pi_{1,k}, g\right) + \int g\mathrm{d}\Gamma^a\left(\pi_{1,k}, \pi_{1,l}\right)\\
\mathcal{E}^a\left(\pi_{1,l}, \pi_{1,k}g\right) &= \int \pi_{1,k}\mathrm{d}\Gamma^a\left(\pi_{1,l}, g\right) + \int g\mathrm{d}\Gamma^a\left(\pi_{1,l}, \pi_{1,k}\right).
\end{aligned}$$

and using the symmetry of a we see that

$$\begin{aligned}
&\lim_{t\to 0}\left\langle g, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,k}\left(\mathbf{X}_{0,t}\right)\pi_{1,l}\left(\mathbf{X}_{0,t}\right)\right]}{t}\right\rangle\\
&= -2\sum_{i,j=1}^{d}\int_{g^2(\mathbb{R}^d)} a^{ij}(y)g(y)U_i\pi_{1,k}(y)U_j\pi_{1,l}(y)\mathrm{d}y\\
&= -2\int_{g^2(\mathbb{R}^d)} a^{kl}(y)g(y)\mathrm{d}y.
\end{aligned}$$

□

Let us fix a collection $V = (V_1, \ldots, V_d)$ of Lip^3 vector fields on $\mathbb{R}^e$ on let us consider the RDE[5]

$$\begin{cases} \mathrm{d}Y = V(Y)\mathrm{d}\mathbf{X}^{a,x}\\ Y_0 = y.\end{cases}$$

where Y is the $\mathbb{R}^e$-valued solution path.[6] In general, Y is not Markov, but it is easy to see that $Z^z = (\mathbf{X}^{a,x}, Y) \in g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e$ started at $z = (x, y)$ is Markov and (unique)

[5] Regularity of the vector fields could be improved to Lip$^{2+\epsilon}$. Also, one can easily add a drift term $V_0(Y)\,\mathrm{d}t$ by considering the canonical space-time rough path $\left(\mathbf{X}^{a,x}, t\right)$.

[6] We could construct the solution as (random) geometric rough path with values in $g^2\left(\mathbb{R}^e\right)$ and the arguments which follow extend to this case.

solution of the RDE

$$\begin{cases} dZ^z = W(Z^z)\, d\mathbf{X}^{a,x}, \\ Z_0^z = (x, y). \end{cases}$$

where $W = (W_1, \ldots, W_d)$ with vector fields W_i on $g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e$ given by

$$W_i(x, y) = (U_i(x), V_i(y)), \quad (x, y) \in g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e.$$

Recall that $U_i : g^2\left(\mathbb{R}^d\right) \to g^2\left(\mathbb{R}^d\right)$ are the vector fields defined in (3) and, by the usual identification with first order differential operators, the U_i extend canonically to first order differential operators (and hence vector fields) on $g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e$ which we denote for clarity with $\widetilde{U}_i$. We now describe the infinitesimal behaviour of the associated semigroup $t \mapsto \mathbb{E}^{a,x}\left(f\left(Z_t\right)\right)$.

Proposition 28 *Let $f, g \in C_c^\infty\left(g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e\right)$. Then*

$$\lim_{t\to 0}\left\langle \frac{\mathbb{E}^{a,\cdot}\left(f\left(Z_t^{\cdot}\right) - f(.)\right)}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)} = -\sum_{i,j=1}^{d} \int_{g^2(\mathbb{R}^d)\oplus\mathbb{R}^e} a^{ij}(x) W_i f(x, y) W_j^* g(x, y)\, dx\, dy =: \mathcal{E}^Z(f, g)$$

where W^* is the adjoint of W with respect to Lebesue measure on $g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e$.

Proof Let us fix f, g $C_c^\infty\left(g^2\left(\mathbb{R}^d\right) \oplus \mathbb{R}^e\right)$. We want to apply Lemma 26 with unbounded vector fields W (the unboundedness comes from the U_i) and we need to localize our problem. Let $R > 0$ such that f and g are 0 outside $B(0, R)$, and define compactly supported smooth vector fields U_i^R such that U_i^R and U_i agree on $B(0, 2R)$. Let Z^R denote the solution of the RDE driven by X along the vector fields $W_i^R = \left(U_i^R, V_i\right)$. Observe first that $W_i^R f = W_i f$ by construction. Applying Lemma 26, we obtain

$$\begin{aligned}
&\lim_{t\to 0}\left\langle \frac{\mathbb{E}\left(f\left(Z_t^{R,\cdot}\right) - f(.)\right)}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)} \\
&= \sum_{i=1}^{d} \lim_{t\to 0}\left\langle W_i f(.)\, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,i}\left(\mathbf{X}_{0,t}\right)\right]}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)} \\
&+ \frac{1}{2}\sum_{i,j=1}^{d} \lim_{t\to 0}\left\langle W_i W_j f(.)\, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{1,i}\left(\mathbf{X}_{0,t}\right)\pi_{1,j}\left(\mathbf{X}_{0,t}\right)\right]}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)} \\
&+ \frac{1}{2}\sum_{i,j=1}^{d} \lim_{t\to 0}\left\langle \left[W_i, W_j\right] f(.)\, \frac{\mathbb{E}^{a,\cdot}\left[\pi_{2,i,j}\left(\mathbf{X}_{0,t}\right)\right]}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)}.
\end{aligned}$$

As $Z_t^{R,\cdot}$ and Z differ only through the area of $\mathbf{X}^{a,\cdot}$, using that uniformly over $x \in B(0,R)$, the probability of $\mathbf{X}^{a,x}$ going outside $B(0,2R)$ is bounded above by $C\exp\left(-CR^2\right)$, we easily see that

$$\lim_{R\to\infty}\lim_{t\to 0}\left\langle \frac{\mathbb{E}\left(f\left(Z_t^{R,\cdot}\right)-f(.)\right)}{t}, g\right\rangle_{L^2} = \lim_{t\to 0}\left\langle \frac{\mathbb{E}\left(f\left(Z_t^{\cdot}\right)-f(.)\right)}{t}, g\right\rangle_{L^2}.$$

We then use lemma 27 to obtain

$$\begin{aligned}
&\lim_{t\to 0}\left\langle \frac{\mathbb{E}\left(f\left(Z_t^{\cdot}\right)-f(.)\right)}{t}, g\right\rangle_{L^2(g^2(\mathbb{R}^d)\oplus\mathbb{R}^e)}\\
&\quad= \sum_{i,j=1}^{d}\int_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} a^{ij}(x)\widetilde{U}_i\left[gW_jf\right](x,y)\mathrm{d}x\,\mathrm{d}y\\
&\qquad-\sum_{i,j=1}^{d}\int_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} a^{ij}(x)g(x,y)W_iW_jf(x,y)\mathrm{d}x\,\mathrm{d}y.
\end{aligned}$$

The proof is finished if we can show

$$\begin{aligned}
&\int_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} a^{ij}(x)\left(\widetilde{U}_i\left[gW_jf\right](x,y)-g(x,y)W_iW_jf(x,y)\right)\mathrm{d}x\,\mathrm{d}y\\
&\quad=\int_{g^2(\mathbb{R}^d)\oplus\mathbb{R}^e} a^{ij}(x)W_if(x,y)W_j^*g(x,y)\,\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

and to see this we may assume, by a simple limit argument, that $a\in\Xi(\Lambda)$ is smooth. We have $a^{ij}(x)\widetilde{U}_i\left[gW_jf\right](x,y)$ equal to

$$\widetilde{U}_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)gW_jf\right](x,y)-\widetilde{U}_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)\right]\left(gW_jf\right)(x,y),$$

and $a^{ij}(x)g(x,y)W_iW_jf(x,y)$ equal to

$$g(x,y)W_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)W_jf\right](x,y)-W_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)\right]\left(gW_jf\right)(x,y).$$

But by construction of W_i we have $W_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)\right]=\widetilde{U}_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)\right]$. Moreover, by integration by parts,

$$\int_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e}\widetilde{U}_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)gW_jf\right](x,y)\mathrm{d}x\,\mathrm{d}y=0,$$

and we see that

$$\begin{aligned}
&\int\limits_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} a^{ij}(x)\left(\widetilde{U}_i\left[gW_jf\right](x,y)-g(x,y)W_iW_jf(x,y)\right)\mathrm{d}x\,\mathrm{d}y\\
&=-\int\limits_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} g(x,y)W_i\left[a^{ij}\left(\pi_{g^2(\mathbb{R}^d)}(.)\right)W_jf\right](x,y)\mathrm{d}x\,\mathrm{d}y\\
&=-\int\limits_{(x,y)\in g^2(\mathbb{R}^d)\times\mathbb{R}^e} a^{ij}(x)W_i^*g(x,y)W_jf(x,y)\mathrm{d}x\,\mathrm{d}y,
\end{aligned}$$

by definition of W_i^*. □

Remark 29 The reader might want to check that when $a(x)$ is smooth and depends only on the projection of x onto $\mathbb{R}^d$, an application of Itô's lemma leads to the same result. In particular, when $a = I$ the process Z solves a Stratonovich equation along vector fields $W = (W_1, \ldots, W_d)$ with generator in Hörmander form

$$L^Z = \sum_{i=1}^{d} W_i^2$$

and the associated form $(f, g) \mapsto -\langle L^Z f, g\rangle = -\sum_{i=1}^{d}\int W_i f W_i^* g$ agrees with Proposition 28.

7 Large deviations

We fix $a \in \Xi(\Lambda)$. The law of $t \mapsto \mathbf{X}^{a;x}(\varepsilon t)$ where $\mathbf{X}^{a;x}$ is the $g^2(\mathbb{R}^d)$-valued process associated to the Dirichlet form $\mathcal{E}^a$, started at x, can be viewed as Borel measure on $C_x\left([0,1], g^2(\mathbb{R}^d)\right) \subset C\left([0,1], g^2(\mathbb{R}^d)\right)$, i.e., the space of continuous paths started at x, and is denoted by $\mathbb{P}_{\varepsilon}^{a;x}$. As usual, we write $\mathbf{X} = \mathbf{X}^{a;x}$ when no confusion is possible and in particular under $\mathbb{P}^{a;x}$ where $\mathbf{X}_t(\omega) = \omega(t) \equiv \omega_t$. We shall see that a sample path large deviation principle holds w.r.t. to uniform (and then homogenous Hölder!) topology on $C_x\left([0,T], g^2(\mathbb{R}^d)\right)$. Having properties (i)–(iii) of the of following propostion, the proof follows essentially Varadhan [34], see also [2], and we outline the key steps for the reader's convenience.

Proposition 30 (i) *$\left(g^2(\mathbb{R}^d), d^a\right)$ is a geodesic space.*
(ii) *The Varadhan-Ramírez short time formula holds,*

$$\lim_{\varepsilon\to 0} 4\varepsilon \log p^a(\varepsilon, x, y) = -d^a(x,y)^2. \tag{11}$$

(iii) *For* $\alpha \in (0, 1/2)$ *there exist a constant* $C_{30} = C_{30}(\alpha, \Lambda)$*such that*

$$\sup_{x \in g^2(\mathbb{R}^d)} \mathbb{P}^{a;x} \left(\sup_{0 \le s < t \le 1} \frac{d^a(\mathbf{X}_s, \mathbf{X}_t)}{|t-s|^\alpha} > r \right) \le C_{30} \exp\left(-\frac{r^2}{C_{30}} \right)$$

and the same estimate holds with d *instead of* d^a.

Proof (i) was shown in Proposition 3, (ii) was discussed in the section on short time asymptotics and (iii) follows from Theorem 13. □

On $C\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$, equipped with uniform topology, we define the *energy or action functional*

$$I^a(\omega) = \lim_{|D| \to 0} \sup \sum_{t_i \in D} \frac{d^a\left(\omega_{t_i}, \omega_{t_{i-1}}\right)^2}{t_i - t_{i-1}} \in [0, \infty]. \tag{12}$$

We shall see shortly that I^a is a good rate function in the sense that $\phi \mapsto I^a(\phi)$ is lower semicontinuous with compact level sets.

7.1 Upper bound

We first recall that d^a is a geodesic distance, i.e., that for all $x, y \in g^2\left(\mathbb{R}^d\right)$, there exists a continuous path joining x to y, of length $d^a(x, y)$.

Proposition 31 (i) *On* $C\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$ *we have*

$$\inf_{\omega:\, \omega(s)=y, \omega(t)=z} I^a(\omega) = \frac{d^a(y,z)^2}{t-s}$$

and the infimum is attained by a d^a*-geodesic path.*

(ii) *More generally,*

$$\inf_{\substack{\omega(t_i)=x_i \\ i=1,\ldots,m}} I^a(\omega) = I^a\left(\omega^D\right) = \sum_{i=1}^m \frac{d^a(x_i, x_{i-1})^2}{t_i - t_{i-1}}$$

where ω^D *is a piecewise* d^a*-geodesic path with* $\omega^D(t_i) = x_i$ *for all* $i = 1, \ldots, m$.

(iii) *In particular,*

$$d^a(\omega_s, \omega_t) \le I^a(\phi)^{1/2} (t-s)^{1/2}. \tag{13}$$

Proof Straight-forward, see [2,34] for instance. □

Lemma 32 (i) *The functional* I^a *is a good rate-function.*

(ii) *If C is closed and $C_\delta \supset C$ denotes the δ-neighbourhood of C (indifferently defined via d or d^a) then*

$$\lim_{\delta \to 0} \inf_{\omega \in C_\delta} I^a(\omega) = \inf_{\omega \in C} I^a(\omega).$$

Proof Using (13) and Arzela–Ascoli this is proved as in [35]. □

Lemma 33 *Let D be a dissection of $[0,1]$ with $\#D$ points and define the (continuous) evaluation map*

$$\Pi_D(\omega) := (\omega_t)_{t \in D} \in \left[g^2\left(\mathbb{R}^d\right)\right]^{\#D}.$$

Let C be a closed "cylindrical" set of form $\Pi_D^{-1} A$ with $A \in \left[g^2\left(\mathbb{R}^d\right)\right]^{\#D}$ closed. Then

$$\limsup_{\varepsilon \to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}(C) \le -\inf_{\omega \in C} I^a(\omega).$$

Proof Using the short time formula (11) and Lemma 31 this is proved in the same way as [34, Lemma 3.1]. □

Lemma 34 *For every $\delta > 0$,*

$$\limsup_{m \to \infty} \limsup_{\epsilon \to 0} \varepsilon \log \sup_{x \in g^2(\mathbb{R}^d)} \mathbb{P}_\varepsilon^{a;x}\left(\sup_{0 \le t \le 1} d^a\left(\mathbf{X}_t, \mathbf{X}_t^{D_m}\right) > \delta\right) = -\infty$$

where $\mathbf{X}^{D_m}$ is the d-geodesic approximation connecting the points $\{\mathbf{X}_t : t \in D^m\}$ with $D^m = \{j/m : j = 0, \ldots, m\}$.

Proof For a fixed t and $D = D_m$ let t_D be the closest point in D to the left of t. Noting that $\mathbf{X}_{t_D} = \mathbf{X}_{t_D}^D$ and using Lipschitz equivalence of d and d^a we have

$$\begin{aligned} d^a\left(\mathbf{X}_t, \mathbf{X}_t^D\right) &\le d^a\left(\mathbf{X}_t, \mathbf{X}_{t_D}\right) + d^a\left(\mathbf{X}_{t_D}, \mathbf{X}_{t_D}^D\right) + d^a\left(\mathbf{X}_t^D, \mathbf{X}_{t_D}^D\right) \\ &\le C_{34}^1\left(d\left(\mathbf{X}_t, \mathbf{X}_{t_D}\right) + d\left(\mathbf{X}_t^D, \mathbf{X}_{t_D}^D\right)\right) \end{aligned}$$

We know from the earlier section on strong geodesic approximation that

$$\sup_D \left\|\mathbf{X}^D\right\|_{\alpha-\text{Höl};[0,1]} \le 3\left\|\mathbf{X}\right\|_{\alpha-\text{Höl};[0,1]}$$

and it follows that

$$\sup_{0 \le t \le 1} d^a\left(\mathbf{X}_t, \mathbf{X}_t^D\right) \le 4\left\|\mathbf{X}\right\|_{\alpha-\text{Höl};[0,1]} \times |D|^\alpha$$

where $|D|$ denotes the mesh of D as usual. By a simple scaling argument (Sect. 3.3) and Proposition 30, (iii) we see that

$$\mathbb{P}_{\varepsilon}^{a;x}\left(4\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]}>\delta m^{\alpha}\right)\leq C_{30}\exp\left(-\frac{1}{C_{30}}\frac{\delta^{2}m^{2\alpha}}{\varepsilon}\right)$$

and, noting that C_{30} does not depend on x,

$$\begin{aligned}\sup_{x}\mathbb{P}_{\varepsilon}^{a;x}\left(\sup_{0\leq t\leq 1}d^{a}\left(\mathbf{X}_{t},\mathbf{X}_{t}^{D_{m}}\right)>\delta\right)&\leq\sup_{x}\mathbb{P}_{\varepsilon}^{a;x}\left(4\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]}>\delta m^{\alpha}\right)\\&\leq C_{30}\exp\left(-\frac{1}{C_{30}}\frac{\delta^{2}m^{2\alpha}}{\varepsilon}\right).\end{aligned}$$

It readily follows that

$$\lim_{m\to\infty}\sup\lim_{\varepsilon\to 0}\sup\varepsilon\log\sup_{x}\mathbb{P}_{\varepsilon}^{a;x}\left(\sup_{0\leq t\leq 1}d^{a}\left(\mathbf{X}_{t},\mathbf{X}_{t}^{D_{m}}\right)>\delta\right)=-\infty$$

as claimed. □

Theorem 35 *For any measurable $A\subset C_{x}\left([0,1],g^{2}\left(\mathbb{R}^{d}\right)\right)$*

$$\lim_{\varepsilon\to 0}\sup 4\varepsilon\log\mathbb{P}_{\varepsilon}^{a,x}(A)\leq-\inf_{\omega\in\bar{A}}I^{a}(\omega).$$

where $\bar{A}$ is the closure of A w.r.t. to the uniform topology on path space.

Proof It suffices to consider A closed. We write $A_{\delta}\supset A$ for the δ-neighbourhood of A (indifferently defined via d or d^{a}) and set

$$I^{\delta,a}(\omega):=\inf_{\tilde{\omega}:\sup_{t\in[0,1]}d^{a}(\omega_{t},\tilde{\omega}_{t})<\delta}I^{a}(\tilde{\omega})\quad\text{and}\quad T_{\delta}:=\inf_{\omega\in A_{\delta}}I^{a}(\omega).$$

If $\omega\in A$ then $I^{\delta,a}(\omega)\geq T_{\delta}$ and therefore, D^{m} being defined as above,

$$\begin{aligned}\mathbb{P}_{\varepsilon}^{a,x}(A)&\leq\mathbb{P}_{\varepsilon}^{a,x}\left(\omega:I^{\delta,a}(\omega)\geq T_{\delta}\right)\\&\leq\mathbb{P}_{\varepsilon}^{a,x}\left[\sup_{t}d^{a}\left(\omega_{t},\omega_{t}^{D_{m}}\right)\geq\delta\right]+\mathbb{P}_{\varepsilon}^{a,x}\left[I^{a}\left(\omega^{D_{m}}\right)\geq T_{\delta}\right].\end{aligned}$$

Noting that lemma 34 states precisely that

$$\lim_{m\to\infty}\sup\lim_{\varepsilon\to 0}\sup\varepsilon\log\sup_{x}\mathbb{P}_{\varepsilon}^{a,x}\left(\sup_{0\leq t\leq 1}d^{a}\left(\omega_{t},\omega_{t}^{D_{m}}\right)>\delta\right)=-\infty.$$

and that, by Proposition 31, (ii), the set $\left\{\omega : I^a\left(\omega^{D_m}\right) \geq T_\delta\right\}$ is equal to

$$C^m := \left\{\omega : \sum_{i=1}^{m} \frac{d^a\left(\omega_{t_i}, \omega_{t_{i-1}}\right)^2}{t_i - t_{i-1}} \geq T_\delta\right\}$$

we see from Lemma 33 that for any m,

$$\limsup_{\varepsilon\to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}\left[C^m\right] \leq -\inf_{\omega\in C^m} I^a(\omega) \leq -T_\delta.$$

By Lemma 32, $\lim_{\delta\to 0} T_\delta = \inf_{\omega\in A} I^a(\omega)$ and combining all these results yield the upper LDP bound. □

7.2 Lower bound

Lemma 36 *For every $\omega \in C_x\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$ and every $\delta > 0$,*

$$\liminf_{\varepsilon\to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}\left(B_\delta(\omega)\right) \geq -I^a(\omega)$$

where

$$B_\delta(\omega) = \left\{\tilde{\omega} \in C_x\left([0,1], g^2\left(\mathbb{R}^d\right)\right) : \sup_{t\in[0,1]} d^a\left(\omega_t, \tilde{\omega}_t\right) < \delta\right\}.$$

Proof Using the short time formula (11), Lemma 31 and the upper LDP this is proved as [34, Lemma 3.4]. □

Corollary 37 *For any measurable $A \subset C_x\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$*

$$-\inf_{\omega\in A^\circ} I^a(\omega) \leq \liminf_{\varepsilon\to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}(A)$$

where A° is the interior of A w.r.t. to the uniform topology on path space.

Proof W.l.o.g. assume that A is open. Take any $\omega \in A$ and $\delta > 0$ small enough such that $V = B_\delta(\omega) \subset A$. From the last lemma it then follows that

$$\liminf_{\varepsilon\to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}(A) \geq \liminf_{\varepsilon\to 0} 4\varepsilon \log \mathbb{P}_\varepsilon^{a,x}(V) \geq -I^a(\omega).$$

As this is true for all $f \in A$ we have the result. □

7.3 LDP in Hölder topology and Freidlin Wentzell

The above estimates are summarized in

Theorem 38 *Let $\mathbb{P}_\varepsilon^{a;x}$ be the law of $t \mapsto \mathbf{X}^{a;x}(\varepsilon t)$ where $\mathbf{X}^{a;x}$ is the $g^2(\mathbb{R}^d)$-valued process associated to the Dirichlet form $\mathcal{E}^a$. Then $(\mathbb{P}_\varepsilon^{a;x})_{\varepsilon>0}$ satisfies a large deviation principle in uniform topology on $C_x([0,T], g^2(\mathbb{R}^d))$ with good rate function I^a defined in Eq.* (12).

It would be easy to deduce from this result a functional form of Strassen's Law of Iterated Logarithm holds, see [7], but we shall not pursue this here.

Corollary 39 *Fix $\alpha \in [0, 1/2)$. Then $(\mathbb{P}_\varepsilon^{a;x})_{\varepsilon>0}$ satisfies a large deviation principle in α-Hölder topology on $C^{\alpha-\text{Hölder}}([0,T], g^2(\mathbb{R}^d))$ with good rate function I^a.*

Proof The random variable $\|\mathbf{X}^{a;x}\|_{\alpha-\text{Hölder}}$ has a Gaussian tail for all $\alpha < 1/2$. By the inverse contraction principle [6] we see that the large deviation principle in uniform topology can be strenghtened to α-Hölder topology. □

From the contraction principle and Lyons' universal limit theorem [17] we obtain

Corollary 40 (Freidlin–Wentzell) *Let $Y_\varepsilon = \pi(0, y_0; \mathbf{X}_\varepsilon^{a,x})$ denote the $\mathbb{R}^e$-valued (random) RDE solution driven by $\mathbf{X}_\varepsilon^{a,x} = \mathbf{X}^{a,x}(\varepsilon \cdot)$ along fixed* $\text{Lip}^{2+\varepsilon}$ *vector fields $V_1, \ldots, V_d$ on $\mathbb{R}^e$ and started at time* 0 *from y_0 fixed (i.e., π is the Itô map). Let $\mathbb{Q}_\varepsilon$ denote the law of Y_ε. Then $(\mathbb{Q}_\varepsilon : \varepsilon > 0)$ satisfies a large deviation principle in α-Hölder topology, $\alpha \in [0, 1/2)$, with good rate function*

$$J^a(y) = \inf\left\{I^a(\omega) : \omega \in C_x\left([0,1], g^2\left(\mathbb{R}^d\right)\right) \quad \text{and} \quad y = \pi(0, y_0; \omega)\right\}.$$

8 Support theorems

To prove an extension of the Stroock–Varadhan support theorem [13,29] ([3] and [21] for Hölder topology) to RDEs driven by the "Markovian" rough paths $\mathbf{X}^{a;x}$, it would be enough to show to for fixed $a \in \Xi(\Lambda)$ and some $\alpha \in (1/3, 1/2)$,

$$\text{supp}\left(\mathbb{P}^{a;x}\right) = x * C_0^{0,\alpha}\left([0,1], g^2\left(\mathbb{R}^d\right)\right).$$

The $\subset$ direction is obvious (from Sect. 5.2.2) but equality remains an open (and challenging) problem. Nonetheless, we are able to prove the desired extension of the Stroock–Varadhan support theorem. First, by shifting the argument of a we can and will assume $x = 0$. If we can show that for fixed $a \in \Xi(\Lambda)$, some $n \geq 2$ and $\alpha \in \left(\frac{1}{n+1}, \frac{1}{n}\right)$,

$$\text{supp}\left((S_n)_* \mathbb{P}^{a;0}\right) = C_0^{0,\alpha}\left([0,1], g^n\left(\mathbb{R}^d\right)\right)$$

where $S_n : C_0^{0,\gamma}([0,1], g^2(\mathbb{R}^d)) \to C_0^{0,\gamma}([0,1], g^n(\mathbb{R}^d))$ is the continuous Young-Lyons lift, $\gamma \in (1/3, 1/2)$, the extended Stroock–Varadhan support theorem (in Hölder topology of exponent less than $1/n$ and hence in uniform topology) is a consequence of basic consistency properties of RDE solutions and the fundamental continuity result of rough path theory. Validity of the Stroock–Varadhan support theorem for differential equations driven by $\mathbf{X}^{a,x}$ in the rough paths sense was conjectured, via conditional statements, by T. Lyons in [18].

8.1 Support in uniform topology

Let $h, x \in C_0^1([0,1], \mathbb{R}^d)$. Every such x can be lifted to $S(x) \in C_0^{1-\text{var}}([0,1], g^2(\mathbb{R}^d))$ via iterated integration. Similarly, one can lift $x+h$, the translation of x in direction h, to a $g^2(\mathbb{R}^d)$-valued path $S(x+h)$. Provided $\alpha \in (1/3, 1/2]$, this operation extends to a continuous translation operator T_h,

$$\mathbf{x} \in C^{\alpha-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right) \mapsto T_h\mathbf{x} \in C^{\alpha-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right).$$

We refer to [19] for details. We note that for $h \in C_0^1([0,1], \mathbb{R}^d)$ fixed and a sequence $(\mathbf{x}_k)$,

$$\|T_h\mathbf{x}_k\|_{\alpha-\text{Höl}} \to 0 \quad \text{as } k \to \infty \quad \text{iff } d_{\alpha-\text{Höl}}(S(h), \mathbf{x}_k) \to 0 \quad \text{as } k \to \infty.$$

Assuming that $a(x)$ only depends on $\pi_1(x)$, with abuse of notation $a = a(\pi_1(\cdot))$, we have that $X + h$ is Markov with (formal) generator

$$\sum_{i,j} \partial_i \left(a^{ij}(\cdot - h_t)\,\partial_j\right) + \sum_k \dot{h}_t^k \partial_k$$

and $T_h\mathbf{X}$ is a Markov with (formal) generator

$$\sum_{i,j} U_i \left(a^{ij}(\pi_1(\cdot) - h_t)\,U_j\right) + \sum_k \dot{h}_t^k U_k$$

where $U_1, \ldots, U_d$ are the generating left-invariant vector fields on $g^2(\mathbb{R}^d)$.

Proposition 41 *Let $h \in C_0^1([0,1], \mathbb{R}^d)$. There exists a constant C_{41} depending only on Λ and $|\dot{h}|_{\infty;[0,1]}$ such that for all $\varepsilon \in (0,1]$,*

$$\mathbb{P}^{a;0}\left(\|T_h(\mathbf{X})\|_{\infty,[0,1]} < \varepsilon\right) \geq \frac{1}{C_{41}} \exp\left(-\frac{C_{41}}{\varepsilon^2}\right).$$

As a consequence, the support of $\mathbb{P}^{a,0}$ equals the closure of $S\left(C_0^1([0,1], \mathbb{R}^d)\right)$, with respect to uniform topology on $C_0([0,1], g^2(\mathbb{R}^d))$.

Proof We first consider $h = 0$. Let n be the smallest integer such that $n^{-1/2} \leq \varepsilon/2$. Set $y_0 = 0 \in g^2\left(\mathbb{R}^d\right)$. Clearly, $\mathbb{P}^{a;y_0}\left(\|\mathbf{X}\|_{\infty,[0,1]} < \varepsilon\right)$ is greater or equal than

$$q_{\varepsilon,n} := \mathbb{P}^{a;y_0}\left(\forall i \in \{1,\dots,n\} : \left\|X_{\frac{i}{n}}\right\| < n^{-1/2} \quad \text{and} \quad X_t \in B(0,\varepsilon) \ \forall t \in \left[\frac{i}{n}, \frac{i+1}{n}\right]\right).$$

Hence, letting $p^{B(0,\varepsilon)}$denote the Dirichlet heat kernel for $\mathbf{X} = \mathbf{X}^{a;0}$, the Markov property implies

$$q_{\varepsilon,n} = \int_{B(0,n^{-1/2})} \cdots \int_{B(0,n^{-1/2})} p^a_{B(0,\varepsilon)}(1/n, y_0, y_1) \cdots p^a_{B(0,\varepsilon)}(1/n, y_{n-1}, y_n)\, \mathrm{d}y_1 \cdots \mathrm{d}y_n.$$

We join the points y_i and y_{i+1} by the curve γ_i, which is the concatenation of a geodesic curve joining y_i to 0 and a geodesic curve between 0 to y_{i+1}. In particular, the length of γ_i is bounded by $2n^{-1/2}$, and γ_i remains in the ball $B\left(0, n^{-1/2}\right) \subset B(0, \varepsilon/2)$. Hence

$$R_i \equiv d^a\left(\gamma_i, g^2\left(\mathbb{R}^d\right)/B(0,\varepsilon)\right) \geq \varepsilon/2 \geq n^{-1/2}$$

and we can apply the lower heat kernel bounds for the killed process with $t = 1/n$ and $\delta = \min\left(R_i^2, t\right) = 1/n$ to obtain

$$\begin{aligned} p^a_{B(0,\varepsilon)}(1/n, y_i, y_{i+1}) &\geq \frac{n^{d^2/2}}{C^1_{41}} \exp\left(-C^1_{41} n d(y_i, y_{i+1})^2\right) \exp\left(-\frac{C^1_{41}}{nR_i^2}\right) \\ &\geq \frac{n^{d^2/2}}{C^1_{41}} \exp\left(-5C^1_{41}\right) \end{aligned}$$

where we used $d(y_i, y_{i+1}) \leq 2n^{-1/2}$ and $R_i \geq n^{-1/2}$. Since $m(B_r(0)) \simeq r^N$ with doubling constant $N = d^2$ we find

$$\begin{aligned} q_{\varepsilon,n} &\geq \prod_{i=1}^{n} \left\{\frac{1}{C^1_{41}} \exp\left(-5C^1_{41}\right) \times \frac{m\left(B\left(0, n^{-1/2}\right)\right)}{\left(n^{-1/2}\right)^{d^2}}\right\} = \\ &\geq \left\{\exp\left(-C^2_{41}\right)\right\}^n \geq \exp\left(-\frac{C^3_{41}}{\varepsilon^2}\right). \end{aligned}$$

For $h \neq 0$ we note that the process $T_h(\mathbf{X})$ is described by a non-symmetric, time dependent Dirichlet form as in [31], for instance. More precisely, the $\mathbb{R}^d$-valued process $t \mapsto h(t) + \mathbf{X}^1(t)$ is desribed by the form

$$(f, g) \mapsto \int_{\mathbb{R}^d} \left[a^{ij}(\cdot, t)\, \partial_i f\, \partial_j g + g b^i(\cdot, t)\, \partial_i f\right] \mathrm{d}x$$

and the bilinear form for $T_h(\mathbf{X})$ its the natural lift obtained by replacing ∂_i by U_i for $i = 1, \ldots, d$,

$$(f, g) \mapsto \int_{g^2(\mathbb{R}^d)} \left[a^{ij} (\pi_1(\cdot), t)\, U_i f U_j g + g b^i (\cdot, t)\, U_i f \right] \mathrm{d}m$$

Such lower order perturbations and time-dependence have been discussed in [27,28, 33]. In particular, there are lower heat kernel bounds for the killed process which allow the above proof to go through. □

8.2 Support in Hölder topology: a conditional result

Motivated by [9] we first study the probability that $\mathbf{X}^{a;x}$ stays in bounded open domain $D \subset g^2(\mathbb{R}^d)$ for long times.

Proposition 42 *Let D be an open domain in $g^2(\mathbb{R}^d)$ with finite volume, no regularity assumptions are made about ∂D. Let $a \in \Xi(\Lambda)$ and $\mathbf{X}^a$ be the process associated to $\mathcal{E}^a$ started at $x \in g^2(\mathbb{R}^d)$ and assume $x \in D$. Then there exist positive constants $K_1 = K_1(x, D, \Lambda)$ and $K_2 = K_2(D, \Lambda)$ so that for all $t \geq 0$*

$$K_1 \mathrm{e}^{-\lambda t} \leq \mathbb{P}\left[\mathbf{X}_s^{a,x} \in D\, \forall s : 0 \leq s \leq t\right] \leq K_2 \mathrm{e}^{-\lambda t}$$

where $\lambda \equiv \lambda_1^a > 0$ is the simple and first Dirichlet eigenvalue of $-L^a$ on the domain D. Moreover,

$$\forall a \in \Xi(\Lambda) : 0 < \lambda_{\min} \leq \lambda_1^a \leq \lambda_{\max} < \infty$$

where $\lambda_{\min}, \lambda_{\max}$ depend only on Λ and D.

Remark 43 The proof will show that $K_1 \sim \psi_1^a(x)$. Noting that $\psi_1^a(x)\mathrm{e}^{-\lambda_1^a t}$ solves the same PDE as $u^a(t, x)$, the above can be regarded as a "partial" parabolic boundary Harnack statement.

Proof If p_D^a denotes the Dirichlet heat kernel for D we can write

$$u^a(t, x) := \mathbb{P}^x\left[\mathbf{X}_s^a \in D\, \forall s : 0 \leq s \leq t\right] = \int_D p_D^a(t, x, y)\mathrm{d}y.$$

Recall [12] that p_D^a is the kernel for a semigroup $P_D^a : L^2(D) \to L^2(D)$ which corresponds to the Dirichlet form $(\mathcal{E}^a, \mathcal{F}_D)$ whose domain $\mathcal{F}_D$ consists of all $f \in \mathcal{F} \equiv D(\mathcal{E}^a)$ with quasi continuous modifications equal to 0 q.e. on D^c. The infinitesimal generator of P_D^a, denoted by L_D^a, is a self-adjoint, densely defined operator with spectrum $\sigma\left(-L_D^a\right) \subset [0, \infty)$. We now use an ultracontractivity argument to show that $\sigma\left(-L_D^a\right)$ is discrete. To this end, we note that the upper bound on p^a plainly implies $\left|p_D^a(t, \cdot, \cdot)\right|_\infty = O(t^{-d^2/2})$. Since $|D| < \infty$ if follows that $\left\|P_D^a(t)\right\|_{L^1 \to L^\infty} < \infty$

which is, by definition, ultracontractivity of the semigroup P_D^a. It is now a standard consequence [5, Theorem 2.1.4] that $\sigma\left(-L_D^a\right) = \left\{\lambda_1^a, \lambda_2^a, \ldots\right\} \subset [0,\infty)$, listed in non-decreasing order. Moreover, it is clear that $\lambda_1^a \neq 0$; indeed the kernel estimates are plenty to see that $\left\| P_D^a(t)\right\|_{L^2\to L^2} \to 0$ as $t\to\infty$ which contradicts the existence of non-zero $f \in L^2(D)$ so that $P_D^a(t) f = f$ for all $t \geq 0$. Let us note that

$$\begin{aligned}
\lambda_1^a &= \inf \sigma(H)\\
&= \inf\left\{\mathcal{E}^a(f,f) : f\in\mathcal{F}_D \text{ with } |f|_{L^2(D)} = 1\right\} \quad \text{(by Rayleigh-Ritz)}\\
&= \inf\left\{\int_D \Gamma^a(f,f)\,\mathrm{d}m : f\in\mathcal{F}_D \text{ with } |f|_{L^2(D)} = 1\right\}
\end{aligned}$$

and since $\Gamma^a(f,f)/\Gamma^I(f,f) \in \left[\Lambda^{-1},\Lambda\right]$ for $f\neq 0$ it follows that $\lambda_1^a \in [\lambda_{\min},\lambda_{\max}]$ for all $a\in\Xi(\Lambda)$ where we set

$$\lambda_{\min} = \Lambda^{-1}\lambda_1^I, \qquad \lambda_{\max} = \Lambda\lambda_1^I. \tag{14}$$

From [5, Theorem 1.4.3] the lower heat kernel estimates for the killed process imply irreducibility of the semigroup P_D^a, hence simplicity of the first eigenvalue λ, and there is an a.s. strictly positive eigenfunction to $\lambda\equiv\lambda_1^a$, say $\psi\equiv\psi_1^a$, and by De Giorgi-Moser-Nash regularity we may assume that ψ is Hölder continuous and strictly positive away from the boundary (this follows also from Harnack's inequality). We also can (and will) assume that $\|\psi\|_{L^2(D)} = 1$.

Lower bound: Noting that $v(t,x) = \mathrm{e}^{-\lambda t}\psi(x)$ is a weak solution of $\partial_t v = L_D^a v$ with $v(0,\cdot) = \psi$ we have

$$v(t,x) = \int_D p_D^a(t,x,y)\psi(y)\mathrm{d}y,$$

at first for, a.e., x but by using a Hölder regular version of p_D^a the above holds for all $x\in D$. It follows that

$$\begin{aligned}
0 &< \psi(x)\\
&= \mathrm{e}^{\lambda t}\int_D p_D^a(t,x,y)\psi(y)\mathrm{d}y\\
&\leq \mathrm{e}^{\lambda(t+1)}\int_D p_D^a(t,x,y)\int_D p_D^a(1,y,z)\,\psi(z)\,\mathrm{d}z\,\mathrm{d}y\\
&\leq \mathrm{e}^{\lambda(t+1)}\int_D\left(p_D^a(t,x,y)\sqrt{\int_D\left[p_D^a(1,y,z)\right]^2\mathrm{d}z}\sqrt{\int\psi^2(z)\,\mathrm{d}z}\right)\mathrm{d}y\\
&\leq C(\Lambda,D)\,\mathrm{e}^{\lambda(t+1)}u^a(t,x)\\
&= \left[C(\Lambda,D)\,\mathrm{e}^{\lambda_{\max}}\right]\times\mathrm{e}^{\lambda t}u^a(t,x)
\end{aligned}$$

and this gives the lower bound with $K_1 = \psi(x)/\left[C(\Lambda, D)\, e^{\lambda_{\max}}\right]$. Clearly $\psi = \psi_1^a$ depends on a and a piori so does K_1. We now show that ψ (and hence K_1) depends on a only through Λ. From

$$p_D^a(t, y, y) = \sum_{i=1}^{\infty} e^{-\lambda_i^a t} \left|\psi_i^a(y)\right|^2$$

evaluated at $t = 1$ say we see that

$$|\psi(y)|^2 \leq e^{\lambda} p_D^a(1, y, y) \leq e^{\lambda_{\max}} p_D^a(1, y, y) \leq e^{\lambda_{\max}} p^a(1, y, y)$$

and by using our upper heat kernel estimates for p^a we see that there is a constant $M = M(\Lambda, D)$ such that $|\psi|_\infty \leq M$. Given x and M we can find a compact set $\mathfrak{K} \subset D$ so that $m(D \backslash \mathfrak{K}) \leq 1/(4M^2)$ and $x \in \mathfrak{K}$ (recall that m is Haar measure on $g^2\left(\mathbb{R}^d\right)$). By Harnack's inequality

$$\sup_{\mathfrak{K}} \psi \leq C\psi(x).$$

for $C = C(\mathfrak{K}, \Lambda) = C(x, D, \Lambda)$. We then have

$$1 = |\psi|_{L^2} \leq M\sqrt{m(D \backslash \mathfrak{K})} + C\psi(x)\sqrt{m(\mathfrak{K})} \leq 1/2 + C\psi(x)\sqrt{m(D)}$$

which gives the required lower bound on $\psi(x) \equiv \psi_1^a(x)$ which only depends on x, D and Λ.
Upper bound: Recall that $-\lambda \equiv -\lambda_1^a$ denotes the first eigenvalue of L_D^a with associated semigroup P_D^a. It follows that

$$\left|P_D^a(t) f\right|_{L^2} \leq e^{-\lambda t} |f|_{L^2}$$

which may be rewritten as

$$\left|\int_D p_D^a(t, \cdot, z) f(z)\, dz\right|_{L^2} \leq e^{-\lambda t} |f|_{L^2}.$$

Let $t > 1$. Using Chapman–Kolmogorov and symmetry of the kernel,

$$u(t, x) = \int_D p_D^a(t, x, z)\, dz = \int_D \int_D p_D^a(1, x, y) p_D^a(t-1, z, y) dy\, dz$$
$$= \sqrt{m(D)} \left(\int_D \left(\int_D p_D^a(t-1, z, y) p_D^a(1, x, y) dy\right)^2 dz\right)^{1/2}$$

$$
\begin{aligned}
&= \sqrt{m(D)} \left| \left(\int_D p_D^a(t-1,\cdot,y) p_D^a(1,x,y) \mathrm{d}y \right) \right|_{L^2(D)} \\
&= \sqrt{m(D)} \left| P_D^a(t-1)\ p_D^a(1,x,\cdot) \right|_{L^2(D)} \\
&\le \sqrt{m(D)} \mathrm{e}^{-\lambda(t-1)} \left| p_D^a(1,x,\cdot) \right|_{L^2(D)} \\
&\le \sqrt{m(D)} \mathrm{e}^{\lambda_{\max}} \mathrm{e}^{-\lambda t} \sqrt{p_D^a(2,x,x)} \\
&\le K_2 \mathrm{e}^{-\lambda t}.
\end{aligned}
$$

where we used upper heat kernel estimates in the last step to obtain $K_2 = K_2(D, \Lambda)$. □

Corollary 44 *Fix $a \in \Xi(\Lambda)$. There exists $K = K(\Lambda)$ and for all $\varepsilon > 0$ there exist $\lambda = \lambda^{(\varepsilon)}$ such that*

$$
K^{-1} \mathrm{e}^{-\lambda t \varepsilon^{-2}} \le \mathbb{P}^{a,0} \left[||\mathbf{X}||_{0,[0,t]} < \varepsilon \right] \tag{15}
$$

$$
\forall x\ :\ \mathbb{P}^{a,x} \left[||\mathbf{X}||_{0,[0,t]} < \varepsilon \right] \le K \mathrm{e}^{-\lambda t \varepsilon^{-2}}. \tag{16}
$$

Proof A straight-forward consequence of scaling and Proposition 42 applied to

$$
D = B(0,1) = \{ y : \|y\| < 1 \}
$$

where $\|\cdot\|$ is the standard CC norm on $g^2(\mathbb{R}^d)$. Then λ is the first eigenvalue corresponding to a scaled by factor ε. □

Proposition 45 *Let $\alpha \in [0, 1/2)$. There exists a constant C_{45} such that for all $\varepsilon \in (0, 1]$ and $R > 0$*

$$
\mathbb{P}^{a,0} \left(\sup_{|t-s|<\varepsilon^2} \frac{\left\| \mathbf{X}_{s,t} \right\|}{|t-s|^\alpha} > R \,\middle|\, \|\mathbf{X}\|_{0;[0,1]} < \varepsilon \right) \le C_{45} \exp\left(-\frac{1}{C_{45}} \frac{R^2}{\varepsilon^{2(1-2\alpha)}} \right).
$$

Proof There will be no confusion to write $\mathbb{P}^x \equiv \mathbb{P}^{a,x}$ and $\mathbb{P}_\varepsilon^x \equiv \mathbb{P}^x \left(\cdot | \|\mathbf{X}\|_{0;[0,1]} < \varepsilon \right)$. Suppose there exists a pair of times $s, t \in [0, 1]$ such that

$$
s < t,\ |t-s| < \varepsilon^2 \quad \text{and} \quad \frac{\left\| \mathbf{X}_{s,t} \right\|}{|t-s|^\alpha} > R.
$$

Then there exists a $k \in \{1, \dots, \lceil 1/\varepsilon^2 \rceil\}$ so that $[s,t] \subset \left[(k-1)\varepsilon^2, (k+1)\varepsilon^2 \right]$. In particular, the probability that such a pair of times exists is at most

$$
\sum_{k=1}^{\lceil 1/\varepsilon^2 \rceil} \mathbb{P}_\varepsilon^0 \left(\|\mathbf{X}\|_{\alpha, [(k-1)\varepsilon^2, (k+1)\varepsilon^2]} > R \right).
$$

Set $\left[(k-1)\,\varepsilon^2, (k+1)\,\varepsilon^2\right] =: [T_1, T_2]$. The rest of the proof is concerned with the existence of C such that

$$\mathbb{P}^0_\varepsilon\left(||\mathbf{X}||_{\alpha,[T_1,T_2]} > R\right) \le C \exp\left(-C^{-1}\frac{R^2}{\varepsilon^{2(1-2\alpha)}}\right)$$

since the factor $\lceil 1/\varepsilon^2 \rceil$ can be absorbed in the exponential factor be making C bigger. We estimate

$$\begin{aligned}
&\mathbb{P}^0\left(||\mathbf{X}||_{\alpha,[T_1,T_2]} > R \,\middle|\, ||\mathbf{X}||_{0,[0,1]} < \varepsilon\right)\\
&\quad\le \frac{\mathbb{P}^0\left(||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon;\, ||\mathbf{X}||_{0,[T_2,1]} < \varepsilon\right)}{\mathbb{P}^0\left[||\mathbf{X}||_{0,[0,1]} < \varepsilon\right]}.
\end{aligned}$$

By using the Markov-property and the above lemma, writing $\lambda^{(\varepsilon)} = \lambda^{a;\varepsilon}$, this equals

$$\begin{aligned}
&\frac{\mathbb{E}^0\left[\mathbb{P}^{\mathbf{X}_{T_2}}\left(||\mathbf{X}||_{0,[0,1-T_2]} < \varepsilon\right);\, ||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right]}{\mathbb{P}^0\left[||\mathbf{X}||_{0,[0,1]} < \varepsilon\right]}\\
&\le C e^{\lambda^{(\varepsilon)}\varepsilon^{-2}}\mathbb{E}^0\left[e^{-\lambda^{(\varepsilon)}(1-T_2)\varepsilon^{-2}};\, ||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right]\\
&= C e^{\lambda^{(\varepsilon)}T_2\varepsilon^{-2}}\mathbb{P}^0\left[||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right]
\end{aligned}$$

where constants were allowed to change in insignificant ways. If $\mathbf{X}$ had indepedent increments in the group (such as is the case for Enhanced Brownian motion $\mathbf{B}$) $\mathbb{P}^0[\ldots]$ would split up immediately. This is not the case here but the Markov property serves as a substitute; using the Dirichlet heat kernel $p^a_{B(0,\varepsilon)}$we can write

$$\begin{aligned}
&\mathbb{P}^0\left[||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right]\\
&\quad= \int_{B(0,\varepsilon)} dx\;\, p^a_{B(0,\varepsilon)}(T_1, 0, x)\,\mathbb{P}^x\left[||\mathbf{X}||_{\alpha,[0,T_2-T_1]} > R\right].
\end{aligned}$$

Then, scaling and the usual Fernique-type estimates for the Hölder norm of $\mathbf{X}$ gives

$$\sup_x \mathbb{P}^x\left[||\mathbf{X}||_{\alpha,[0,T_2-T_1]} > R\right] \le C \exp\left(-\frac{1}{C}\left(\frac{R}{\varepsilon^{1-2\alpha}}\right)^2\right),$$

where we used $T_2 - T_1 = 2\varepsilon^2$, and we obtain

$$\begin{aligned}
&\mathbb{P}^0\left[||\mathbf{X}||_{\alpha,[T_1,T_2]} > R;\, ||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right] \\
&\quad\leq C\exp\left(-\frac{1}{C}\left(\frac{R}{\varepsilon^{1-2\alpha}}\right)^2\right)\mathbb{P}^0\left[||\mathbf{X}||_{0,[0,T_1]} < \varepsilon\right] \\
&\quad\leq C\exp\left(-\frac{1}{C}\left(\frac{R}{\varepsilon^{1-2\alpha}}\right)^2\right)\mathrm{e}^{-\lambda^{(\varepsilon)}T_1\varepsilon^{-2}}.
\end{aligned}$$

Putting things together we have

$$\begin{aligned}
\mathbb{P}^0\left(||\mathbf{X}||_{\alpha,[T_1,T_2]} > R\,\middle|\,||\mathbf{X}||_{0,[0,1]} < \varepsilon\right) &\leq C\mathrm{e}^{\lambda^{(\varepsilon)}(T_2-T_1)\varepsilon^{-2}}\exp\left(-\frac{1}{C}\left(\frac{R}{\varepsilon^{1-2\alpha}}\right)^2\right) \\
&\leq C\mathrm{e}^{2\lambda_{\max}}\exp\left(-\frac{1}{C}\left(\frac{R}{\varepsilon^{1-2\alpha}}\right)^2\right)
\end{aligned}$$

and the proof is finished. □

Corollary 46 *Let* $\alpha \in [0, 1/2)$. *For all* $R > 0$ *the ball* $\left\{\mathbf{x} : \|\mathbf{x}\|_{\alpha-\text{Höl};[0,1]} < R\right\}$ *has positive* $\mathbb{P}^{a,0}$*-measure and*

$$\lim_{\epsilon\to 0}\mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} < R\,\middle|\,\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right) \to 1. \tag{17}$$

Proof We first observe that the uniform conditioning allows to localise the Hö lder norm. More precisely, take $s < t$ in $[0, 1]$ with $t - s \geq \epsilon^2$ and note that from $\|\mathbf{X}\|_{0;[0,1]} < \varepsilon$ we get $\left\|\mathbf{X}_{s,t}\right\| / |t-s|^\alpha \leq \epsilon^{1-2\alpha}$. It follows that for fixed R and ϵ small enough,

$$\begin{aligned}
&\mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} \geq R\,\middle|\,\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right) \\
&\quad= \mathbb{P}^{a,0}\left(\sup_{|t-s|<\varepsilon^2}\frac{\left\|\mathbf{X}_{s,t}\right\|}{|t-s|^\alpha} \geq R\,\middle|\,\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right)
\end{aligned}$$

and the preceding proposition shows convergence to zero with ϵ and (17) follows. Finally,

$$\begin{aligned}
\mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} < R\right) &\geq \mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} < R\,\middle|\,\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right) \\
&\quad\times\mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right) \\
&\geq \mathbb{P}^{a,0}\left(\|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right)/2 \quad \text{(for } \epsilon \text{ small enough)}
\end{aligned}$$

and this is positive by either Proposition 44 or Proposition 41. □

Corollary 47 *Let $Y = \pi\left(0, y_0; \mathbf{X}^{a;0}\right) \equiv \pi(\mathbf{X})$ denote the $\mathbb{R}^e$-valued (random) RDE solution driven by $\mathbf{X}^{a,x}$ along fixed $\mathrm{Lip}^{2+\varepsilon}$ vector fields $V_1, \ldots, V_d$ on $\mathbb{R}^e$ and started at time 0 from y_0 fixed. Then, for any $R > 0$,*

$$\mathbb{P}^{a,0}\left(|Y|_{\alpha-\text{Höl};[0,1]} > R \middle| \|\mathbf{X}\|_{0;[0,1]} < \varepsilon\right) \to 0 \quad \text{with } \epsilon \to 0.$$

Proof From, Lyons' limit theorem, $\|\mathbf{X}\|_{\alpha-\text{Höl};[0,1]} \to 0$ implies, deterministically, $|Y|_{\alpha-\text{Höl};[0,1]} \to 0$. □

8.3 The Stroock–Varadhan support theorem for Markov RDEs

Let $h \in C_0^{1-\text{var}}\left([0,1], \mathbb{R}^d\right)$. Give a uniformly elliptic $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, so that $a \circ \pi_1 \in \Xi(\Lambda)$ we know that $T_h(\mathbf{X}^a)$ is Markov. Furthermore, $\mathbf{X}_{0,\cdot} = \mathbf{X}^a_{0,\cdot} \in C_0^{\gamma-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$ for $\gamma \in (1/3, 1/2)$ and from basic facts of the translation operator we also have

$$T_h(\mathbf{X}) \in C_0^{\gamma-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$$

This "step-2" γ-Hölder rough path lifts uniquely and continuously to any step-N rough path

$$S_N\left(T_h(\mathbf{X})\right) \in C_0^{\gamma-\text{Höl}}\left([0,1], g^N\left(\mathbb{R}^d\right)\right).$$

Obviously, specializing to $h = 0$ and it is clear that $S_N(\mathbf{X})$ is also α-Hölder for $1/(N+1) < \alpha < 1/N$ and thus a "step-N" α-rough path in its own right. By basic consistency properties of rough differential equations, the solutions corresponding to driving $S_N(\mathbf{X})$, as step-N rough path, and $\mathbf{X}$ as step-2 rough path, coincide. Hence, it is good enough to obtain a support description for $S_N(\mathbf{X})$ in α-Hölder topology and we are able to do this with $N = 6$ and any $\alpha < 1/6$.

Lemma 48 *Let $\mathbf{y} \in C_0^{\gamma-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$, $\gamma \in (1/3, 1/2)$. For every integer N, there exists K_N such that for all $s < t$ in $[0,1]$,*

$$\|S_N(\mathbf{y})\|_{\gamma-\text{Höl};[s,t]} \leq K_N \|\mathbf{y}\|_{\gamma-\text{Höl};[s,t]}$$

(Notice that the respective γ-Hölder "norms" are with respect to $g^N\left(\mathbb{R}^d\right)$-valued paths on the left-hand-side and with respect to $g^2\left(\mathbb{R}^d\right)$-valued paths on the right-hand-side.)

Proof See [17, p. 242]. □

Proposition 49 *Let $\gamma \in [0, 1/2)$. Let $h \in C_0^{1-var}\left([0,1], \mathbb{R}^d\right)$ and $\mathbf{y} \in C_0^{\gamma-\text{Höl}}\left([0,1], g^2\left(\mathbb{R}^d\right)\right)$. Then, there exists a constant $C_{49} > 0$ such that for all $0 \leq s < t \leq 1$,*

$$\left\|T_h(\mathbf{y})_{s,t}\right\| \leq C_{49}\left(\|\mathbf{y}\|_{\gamma-\text{Höl};[s,t]} |t-s|^\gamma + |h|_{1-\text{var},[s,t]}\right). \tag{18}$$

In particular, if $h \in \mathcal{H}$*, the usual Cameron–Martin space with* $|h|_{\mathcal{H}} \equiv |\dot{h}|_{L^2([0,1],\mathbb{R}^d)}$*, then*

$$\|T_h(\mathbf{y})\|_{\gamma-\text{Höl};[s,t]} \leq C_{49}\left(\|\mathbf{y}\|_{\gamma-\text{Höl};[s,t]} + |h|_{\mathcal{H}}|t-s|^{\frac{1}{2}-\gamma}\right). \tag{19}$$

Proof It is easy to see that $\left\|T_h(\mathbf{y})_{s,t}\right\|$ is less equal than a constant times

$$\left|h_{s,t}+\mathbf{y}^1_{s,t}\right| + \sqrt{|\pi_2(\mathbf{y}_{s,t})|} + \sqrt{\left|\int_s^t h_{s,r}\otimes \mathrm{d}h_r\right|} + \sqrt{\left|\int_s^t h_{s,r}\otimes \mathrm{d}y_r\right|} + \sqrt{\left|\int_s^t y_{s,r}\otimes \mathrm{d}h_r\right|}.$$

The first three summand are easy to estimate. To deal with the last two it suffices to note

$$\sqrt{\left|\int_s^t y_{s,r}\otimes \mathrm{d}h_r\right|} \leq \sqrt{\sup_{r\in[s,t]}\left|y_{s,r}\right| . |h|_{1-\text{var},[s,t]}} \leq \frac{1}{2}\sup_{r\in[s,t]}\left|y_{s,r}\right| + \frac{1}{2}|h|_{1-\text{var},[s,t]},$$

then use integration by parts for the last summand. Finally, $|h|_{1-\text{var};[s,t]} \leq |h|_{\mathcal{H}}|t-s|^{\frac{1}{2}}$ implies (19). □

Let us remark that Proposition 41 and remains valid with identical proof in the step-N setting. (The toolbox of Dirchlet forms applies immediately with $g^N(\mathbb{R}^d)$ instead of $g^2(\mathbb{R}^d)$. Constants may depend on N, but $N = 6$ will suffice for us.

Theorem 50 *Let h be a Lipschitz path and* $\alpha \in [0, 1/6)$*. Then, for all* $\varepsilon > 0$,

$$\mathbb{P}^a\left(\|S_6(T_h(\mathbf{X}))\|_{\alpha-\text{Höl};[0,1]} < \varepsilon\right) > 0.$$

Proof By take α close enough to $1/6$ we may assyme that $N = [1/\alpha] = 6$. We shall choose a (good) Hölder exponent $\gamma = \gamma(\alpha) \in (1/3, 1/2)$, to be chosen below ($\gamma = 1/3 + (1/6 - \alpha)/2$ will do). For any $p > 0$ (to be choose large later on),

$$\begin{aligned}
&\mathbb{P}\left(\|S_N(T_h(\mathbf{X}))\|_{\alpha-\text{Höl};[0,1]} > \varepsilon\right)\\
&\leq \mathbb{P}\left(\sup_{|t-s|<\varepsilon^p}\frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^{\alpha}} > \varepsilon\right) + \mathbb{P}\left(\sup_{|t-s|\geq\varepsilon^p}\frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^{\alpha}} > \varepsilon\right)\\
&\leq \mathbb{P}\left(\sup_{|t-s|<\varepsilon^p}\frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^{\gamma}} > \varepsilon\right) + \mathbb{P}\left(\sup_{|t-s|\geq\varepsilon^p}\frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^{\alpha}} > \varepsilon\right).
\end{aligned}$$

Using $\mathbb{P}(A) \leq \mathbb{P}(B) + \mathbb{P}(C) \implies \mathbb{P}(A^c) \geq -\mathbb{P}(B) + \mathbb{P}(C^c)$ we see that

$$\begin{aligned}\mathbb{P}\left(\|S_N(T_h(\mathbf{X}))\|_{\alpha-\text{Höl};[0,1]} \leq \varepsilon\right) &\geq -\mathbb{P}\left(\sup_{|t-s|<\varepsilon^p} \frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^\gamma} > \varepsilon\right) \\ &\quad + \mathbb{P}\left(\ \|S_N(T_h(\mathbf{X}))\|_{\infty;[0,1]} \leq \varepsilon.\left(\varepsilon^p\right)^\alpha\right) \\ &\equiv -(I) + (II).\end{aligned}$$

and the proof will be finished if we can find $p = p(\alpha)$ such that $(I)/(II) \to 0$ as $\varepsilon \to 0$. It follows from Proposition 41 that as $\varepsilon \to 0$,

$$(II) \geq \frac{1}{c_1} \exp\left(-c_1\left(\varepsilon^{1+p\alpha}\right)^{-2}\right) = \frac{1}{c_1} \exp\left(-c_1\left(\frac{1}{\varepsilon}\right)^{2+2p\alpha}\right)$$

which we express without the irrelevant positive constants as

$$\log(II) \gtrsim -\left(\frac{1}{\varepsilon}\right)^{2+2p\alpha}. \tag{20}$$

At the same time,

$$\begin{aligned}(I) &= \mathbb{P}\left(\sup_{|t-s|<\varepsilon^p} \frac{\left\|S_N(T_h(\mathbf{X}))_{s,t}\right\|}{|t-s|^\gamma} > \varepsilon\right) \\ &\leq \sum_{k=1}^{\lceil 1/\varepsilon^p \rceil} \mathbb{P}\left(\|S_N(T_h(\mathbf{X}))\|_{\gamma-\text{Höl};[(k-1)\varepsilon^p,(k+1)\varepsilon^p]} > \varepsilon\right) \\ &\leq \sum_{k=1}^{\lceil 1/\varepsilon^p \rceil} \mathbb{P}\left(\|T_h(\mathbf{X})\|_{\gamma-\text{Höl};[(k-1)\varepsilon^p,(k+1)\varepsilon^p]} > \frac{\varepsilon}{c_2}\right)\end{aligned}$$

where $c_2 = K_N$ is the constant from Lemma 48. (Here we used $\gamma > 1/3$.) By Proposition 49 this estimate continues with

$$\begin{aligned}&\leq \sum_{k=1}^{\lceil 1/\varepsilon^p \rceil} \mathbb{P}\left(\|\mathbf{X}\|_{\gamma-\text{Höl};[(k-1)\varepsilon^p,(k+1)\varepsilon^p]} + \underbrace{|h|_{\mathcal{H}}\left(2\varepsilon^p\right)^{\frac{1}{2}-\gamma}} > \frac{\varepsilon}{c_3}\right) \\ &\leq \sum_{k=1}^{\lceil 1/\varepsilon^p \rceil} \mathbb{P}\left(\|\mathbf{X}\|_{\gamma-\text{Höl};[(k-1)\varepsilon^p,(k+1)\varepsilon^p]} > \frac{\varepsilon}{c_4}\right)\end{aligned}$$

where the term indicated by the curley bracket can indeed by omitted as $\varepsilon \to 0$ provided p is chosen large enough so that $p(1/2-\gamma) > 1$. With scaling and Fernique estimates we see that

$$\sum_{k=1}^{\lceil 1/\varepsilon^p \rceil} \mathbb{P}\left(\|\mathbf{X}\|_{\gamma-\text{Höl};[(k-1)\varepsilon^p,(k+1)\varepsilon^p]} > \frac{\varepsilon}{c_4} \right) \leq c_5 \varepsilon^p \exp\left(-\frac{1}{c_5} \left(\frac{\varepsilon}{(\varepsilon^p)^{1/2-\gamma}} \right)^2 \right)$$

Focusing on the decay rate of (I) and again ignoring irrelevant positive constants, we see that

$$\log (I) \lesssim -\left(\frac{\varepsilon}{(\varepsilon^p)^{1/2-\gamma}} \right)^2 = -\left(\frac{1}{\varepsilon} \right)^{-2+p(1-2\gamma)}.$$

Recalling $\log (II) \gtrsim -(1/\varepsilon)^{2+p(2\alpha)}$ it is clear that, by choosing p large enough, $(I) / (II) \to 0$ as $\varepsilon \to 0$ *provided that* $1 - 2\gamma > 2\alpha$. Our only constraint is $\gamma > 1/3$ and we now see that this is precisely possible when $\alpha < 1/6$ and so the proof is finished. □

Corollary 51 *The support of the law of* $S_6 \left(\mathbf{X}_{0,\cdot}^{a,0} \right)$ *in* α*-Hölder topology,* $\alpha \in [0, 1/6)$*, equals* $C_0^{0,\alpha-\text{Hölder}} \left([0, 1], g^6 \left(\mathbb{R}^d \right) \right)$.

Proof Given $\alpha \in (0, 1)$, $N = [1/\alpha]$, a fixed Lipschitz h and $\mathbf{x} \in C_0^{0,\alpha-\text{Höl}} \left([0, 1], g^N \left(\mathbb{R}^d \right) \right)$ we know [19, p57] that

$$\mathbf{x} \mapsto T_h (\mathbf{x})$$

is continuous under $d_{\alpha-\text{Höl}}$ on the pathspace $C_0^{0,\alpha-\text{Höl}}$. It then easily follows that

$$T_{-h} \left(\mathbf{x}^n \right) \to 0 \iff \mathbf{x}^n \to S_N (h).$$

Indeed, "$\Longleftarrow$" comes from continuity of $\mathbf{x} \mapsto T_{-h} (\mathbf{x})$ and $T_{-h} (S_N (h)) = S_N (h-h) = 0$ while "$\Longrightarrow$" follows from

$$\underbrace{T_h T_{-h} \left(\mathbf{x}^n \right)}_{=\mathbf{x}^n} \to \underbrace{T_h (0)}_{S_N(H)}.$$

Then use Theorem 50. □

Corollary 52 (Stroock–Varadhan) *Let* $Y = \pi (0, y_0; \mathbf{X}^{a,x}) \equiv \pi (\mathbf{X}^{a,x})$ *denote the* $\mathbb{R}^e$*-valued (random) RDE solution driven by* $\mathbf{X}^{a,x}$ *along fixed* $\text{Lip}^{6+\varepsilon}$ *vector fields* $V_1, \ldots, V_d$ *on* $\mathbb{R}^e$ *and started at time* 0 *from* y_0 *fixed. Let* $\mathbb{Q}$ *denote the law of* $(Y_t : 0 \leq t \leq 1)$*. Then the support of* $\mathbb{Q}$ *in uniform topology is the closure of all control ODE solution,*

$$\mathcal{S} = \left\{ \pi (0, y_0, h) : h \in C^1 \left([0, 1], \mathbb{R}^d \right) \right\}.$$

Here $y \equiv \pi(0, y_0, h)$ *denotes the unique solution, started at time* 0 *from* y_0, *of the ordinary differential equation*

$$\mathrm{d}y = \sum_{i=1}^{d} V_i(y)\mathrm{d}h^i.$$

Proof Y is obtained as RDE solution driven by a $\mathbf{X}^{a,0}$. By a basic consistency properties of RDE solutions, it is *also* the RDE solution driven by $S_6\left(\mathbf{X}^{a,0}\right)$. By continuity of the Itô-Lyons map, the support description of the later implies the Stroock–Varadhan support description for Y. □

Acknowledgements The authors would like to thank T. J. Lyons, S. R. S. Varadhan and J. R. Norris for helpful discussions. The first author is grateful to T. Coulhon, A. A. Grigor'yan and E. B. Davies for conversations during the writing of Sect. 8.2.

References

1. Aronson, D.G.: Bounds for the fundamental solution of a parabolic equation. Bull. Am. Math. Soc. **73**, 890–896 (1967)
2. Bass, R.F., Kumagai, T.: Laws of the iterated logarithm for some symmetric diffusion processes. Osaka J. Math. **37**(3), 625–650 (2000)
3. Arous, G.B., Grădinaru, M., Ledoux, M.: Hölder norms and the support theorem for diffusions. Ann. Inst. H. Poincaré Probab. Stat. **30**(3), 415–436 (1994)
4. Carlen, E.A., Kusuoka, S., Stroock, D.W.: Upper bounds for symmetric Markov transition functions. Ann. Inst. H. Poincaré Probab. Stat. **23**(2, suppl), 245–287 (1987)
5. Davies, E.B.: Heat Kernels and Spectral Theory, Vol. 92 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1989)
6. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, Vol. 38 of Applications of Mathematics (New York), Second edition. Springer, New York (1998)
7. Deuschel, J.-D., Stroock, D.W.: Large Deviations, Vol. 137 of Pure and Applied Mathematics. Academic, Boston (1989)
8. Fabes, E.B., Stroock, D.W.: A new proof of Moser's parabolic Harnack inequality using the old ideas of Nash. Arch. Rational Mech. Anal. **96**(4), 327–338 (1986)
9. Friz, P., Lyons, T., Stroock, D.: Lévy's area under conditioning. Ann. Inst. H. Poincaré Probab. Stat. **42**(1), 89–101 (2006)
10. Friz, P., Victoir, N.: A note on the notion of geometric rough paths. Probab. Theory Related Fields **136**, 395–416 (2006)
11. Friz, P., Victoir, N.: Euler estimates for rough differential equations. J. Differ. Equ. (2007). doi:10.1016/j.jde.2007.10.008
12. Fukushima, M., Ōshima, Y., Takeda, M.: Dirichlet Forms and Symmetric Markov Processes, Vol. 19 of de Gruyter Studies in Mathematics. Walter de Gruyter, Berlin (1994)
13. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes, Second edition. North-Holland, Amsterdam (1989)
14. Jerison, D.: The Poincaré inequality for vector fields satisfying Hörmander's condition. Duke Math. J. **53**(2), 503–523 (1986)
15. Lejay, A.: Stochastic differential equations driven by processes generated by divergence form operators I: a Wong–Zakai theorem. ESAIM Probab. Stat. **10**, 356–379 (2006)
16. Lejay, A.: Stochastic differential equations driven by processes generated by divergence form operators II: Convergence results. ESAIM Probab. Stat. (2007, accepted)
17. Lyons, T.: Differential equations driven by rough signals. Rev. Mat. Iberoamericana **14**(2), 215–310 (1998)
18. Lyons, T.: St. Flour Lectures on Rough Paths, 2004. Handwritten notes available at www.sag.maths.ox.ac.uk/tlyons/st-flour/

19. Lyons, T., Qian, Z.: System Control and Rough Paths. Oxford University Press, Oxford Mathematical Monographs, (2002)
20. Lyons, T., Stoica, L.: The limits of stochastic integrals of differential forms. Ann. Probab. **27**(1), 1–49 (1999)
21. Millet, A., Sanz-Solé, M.: A simple proof of the support theorem for diffusion processes. In Séminaire de Probabilités, XXVIII, volume 1583 of Lecture Notes in Math., p. 36–48. Springer, Berlin, (1994)
22. Montgomery, R.: A tour of subriemannian geometries, their geodesics and applications, Vol. 91 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, (2002)
23. Moser, J.: On a pointwise estimate for parabolic differential equations. Commun. Pure Appl. Math. **24**, 727–740 (1971)
24. Moser, J.: A Harnack inequality for parabolic differential equations. Commun. Pure Appl. Math. **17**, 101–134 (1964)
25. Ramírez, J.A.: Short-time asymptotics in Dirichlet spaces. Commun. Pure Appl. Math. **54**(3), 259–293 (2001)
26. Robinson, D.W.: Elliptic Operators and Lie Groups. Oxford Mathematical Monographs. Clarendon, Oxford University Press, New York, 1991. Oxford Science Publications
27. Saloff-Coste, L., Stroock, D.W.: Opérateurs uniformément sous-elliptiques sur les groupes de Lie. J. Funct. Anal. **98**(1), 97–121 (1991)
28. Stroock, D.W.: Diffusion semigroups corresponding to uniformly elliptic divergence form operators. In Séminaire de Probabilités, XXII, Vol. 1321 of Lecture Notes in Math., pp. 316–347. Springer, Berlin, (1988)
29. Stroock, D.W., Varadhan, S.R.S.: On the support of diffusion processes with applications to the strong maximum principle. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, CA, 1970/1971), Vol. III: Probability theory, pp. 333–359, Berkeley, 1972. University California Press
30. Sturm, K.T.: Analysis on local Dirichlet spaces. III. The parabolic Harnack inequality. J. Math. Pures Appl. (9), **75**(3), 273–297 (1996)
31. Sturm, K.-T.: Analysis on local Dirichlet spaces. II. Upper Gaussian estimates for the fundamental solutions of parabolic equations. Osaka J. Math. **32**(2), 275–312 (1995)
32. Sturm, K.-T.: On the geometry defined by Dirichlet forms. In Seminar on Stochastic Analysis, Random Fields and Applications (Ascona, 1993), volume 36 of Progr. Probab., pp. 231–242. Birkhäuser, Basel, (1995)
33. ter Elst, A.F.M., Robinson, D.W.: Second-order subelliptic operators on Lie groups. II. Real measurable principal coefficients. In Semigroups of operators: theory and applications (Newport Beach, CA, 1998), vol. 42 of Progr. Nonlinear Differential Equations Appl., pp. 103–124. Birkhäuser, Basel, (2000)
34. Varadhan, S.R.S.: Diffusion processes in a small time interval. Commun. Pure Appl. Math. **20**, 659–685 (1967)
35. Varadhan, S.R.S.: Lectures on diffusion problems and partial differential equations, Vol. 64 of Tata Institute of Fundamental Research Lectures on Mathematics and Physics. Tata Institute of Fundamental Research, Bombay, 1980. With notes by Pl. Muthuramalingam and Tara R. Nanda
36. Varopoulos, N.Th.: Small time Gaussian estimates of heat diffusion kernels. II. The theory of large deviations. J. Funct. Anal. **93**(1), 1–33 (1990)
37. Varopoulos, N.Th., Saloff-Coste, L., Coulhon, T.: Analysis and Geometry on Groups, Vol. 100 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1992)