



# A regularity method for lower bounds on the Lyapunov exponent for stochastic differential equations

Jacob Bedrossian<sup>1</sup> · Alex Blumenthal<sup>2</sup> ·  
Sam Punshon-Smith<sup>3</sup>

Received: 30 July 2020 / Accepted: 24 August 2021 / Published online: 13 September 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of  
Springer Nature 2021

**Abstract** We put forward a new method for obtaining quantitative lower bounds on the top Lyapunov exponent of stochastic differential equations. Our method combines (i) a new identity connecting the top Lyapunov exponent to a Fisher information-like functional of the stationary density of the Markov process tracking tangent directions with (ii) a novel, quantitative version of Hörmander’s hypoelliptic regularity theory in an  $L^1$  framework which estimates this (degenerate) Fisher information from below by a  $W_{loc}^{s,1}$  Sobolev norm. This method is applicable to a wide range of systems beyond the reach of currently existing mathematically rigorous methods. As an initial application, we prove the positivity of the top Lyapunov exponent for a class of weakly-

---

J.B. was supported by National Science Foundation CAREER grant DMS-1552826 and National Science Foundation RNMS #1107444. A.B. was supported by National Science Foundation grant DMS-2009431. S.P. was supported by the National Science Foundation grant DMS-1803481.

---

✉ Alex Blumenthal  
ablumenthal6@gatech.edu

Jacob Bedrossian  
jacob@math.umd.edu

Sam Punshon-Smith  
spsmith@math.ias.edu

<sup>1</sup> Department of Mathematics, University of Maryland, College Park, MD 20742, USA

<sup>2</sup> School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

<sup>3</sup> School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540, USA

dissipative, weakly forced stochastic differential equations; in this paper we prove that this class includes the Lorenz 96 model in any dimension, provided the additive stochastic driving is applied to any consecutive pair of modes.

**Mathematics Subject Classification** Primary: 37H15 · 35H10, Secondary: 37D25 · 58J65 · 35B65

## Contents

1	Introduction	430
2	Preliminaries	443
3	Fisher information formula for Lyapunov exponents	449
4	Uniform hypoelliptic regularity via the Fisher information	456
5	Projective hypoellipticity and irreducibility for Euler-like systems	479
6	Rigidity of invariant measures of the deterministic, projective process	490
A	Basic geometric preliminaries and compactness criterion	496
B	Qualitative properties of the projective stationary measure	502
C	Properties of projective lifts and the Furstenberg–Khasminskii formula	508
	References	513

## 1 Introduction

Many nonlinear systems of physical origin exhibit chaotic behavior. Although there is no single mathematical definition of ‘chaos’, a commonly observed and studied aspect of chaotic behavior is *sensitivity with respect to the initial conditions*. One way to quantify this sensitivity is the *Lyapunov exponent* along a given trajectory, i.e., the asymptotic exponential rate at which nearby trajectories converge (a negative exponent) or diverge (a positive exponent, implying sensitivity). Despite the ubiquity of chaotic behavior in systems of physical interest, and in contrast with the rather well-developed abstract theory for the description of chaotic states and associated statistical properties, it is notoriously challenging to verify, for a given system, whether or not a positive Lyapunov exponent is present along a large (e.g., positive-volume) subset of phase space.

The purpose of this paper is to put forward a method for providing quantitative lower bounds for the Lyapunov exponents of weakly-damped, weakly-driven stochastic differential equations. Our method combines two new ingredients:

- (i) An apparently new identity (Theorem A) connecting the largest Lyapunov exponent to a certain degenerate Fisher information functional on the stationary density of the Markov process tracking tangent directions; and
- (ii) A quantitative hypoellipticity argument in  $L^1$  for showing that this Fisher information uniformly controls local fractional Sobolev regularity of the

tangent-direction stationary density (Theorem B), implying by (i) a quantitative lower bound on the Lyapunov exponent in terms of the local regularity of this stationary density.

As a first application of our methods, we prove positivity of the largest Lyapunov exponent for a broad class of weakly-dissipative weakly-driven SDE on  $\mathbb{R}^n$  (with fluctuation dissipation scaling) with bilinear drift term that conserves both volume and the norm on  $\mathbb{R}^n$  (Theorem C). This result incorporates several interesting models of fundamental interest, such as the Lorenz-96 (L96) model (Corollary D) and Galerkin truncations of the Navier–Stokes equation (GNSE) on a periodic box (supplemented by a recent and separate work [19] of the first and third author; see the beginning of Sect. 1.3). These results constitute the first mathematically rigorous results affirming a positive Lyapunov exponent for either of these models, even in the stochastic case.

Below in Sect. 1.1 we give an informal, nontechnical overview of the main results (Theorems A–C and Corollary D), followed by a brief discussion of context within prior work (Sect. 1.2) and a discussion of the future outlook for the ideas introduced in this paper (Sect. 1.3). See Sect. 1.4 for a brief outline of the rest of the paper.

## 1.1 Overview of results

Although several of our most general results (Theorem A and Theorem B) apply to SDEs posed on essentially any (potentially non-compact) geodesically complete manifold  $M$ , in order to simplify the exposition and to fix ideas, here we will consider just SDEs posed on  $\mathbb{R}^n$  (already sufficient for Theorem C and Corollary D). See the main body of the paper for more general statements.

Let  $n \geq 1$ , and consider the stochastic process  $(x_t)_{t \geq 0}$  in  $\mathbb{R}^n$  defined by the Stratonovich SDE

$$dx_t = X_0(x_t) dt + \sum_{k=1}^r X_k(x_t) \circ dW_t^k, \quad (1.1)$$

where  $\{X_k\}_{k=0}^r$  are a family of smooth vector fields on  $\mathbb{R}^n$  and  $\{W^k\}_{k=1}^r$  are independent standard Wiener processes.

Under mild conditions, trajectories of the Markov process  $(x_t)$  are realized by an associated stochastic flow of diffeomorphisms  $\Phi_\omega^t : \mathbb{R}^n \rightarrow \mathbb{R}^n$  indexed by the corresponding sample path  $\omega \in \Omega$ , where  $\Omega$  is the canonical path space for  $\{W^k\}$ .

Under the (natural) assumption that there is a unique stationary probability measure  $\mu$  for  $(x_t)$ , it follows from standard tools in ergodic theory that the limits

$$\begin{aligned}\lambda_1 &= \lim_{t \rightarrow \infty} \frac{1}{t} \log |D_x \Phi_\omega^t|, \\ \lambda_\Sigma &= \lim_{t \rightarrow \infty} \frac{1}{t} \log |\det D_x \Phi_\omega^t|\end{aligned}$$

exist with probability 1 for  $\mu$  almost every  $x \in \mathbb{R}^n$ , while the limiting values  $\lambda_1, \lambda_\Sigma$  are constant (independent of  $x$ ) and deterministic (independent of the random sample  $\omega$ ); for more details, see Theorem 2.4 below.

The value  $\lambda_1$  is known as the *top Lyapunov exponent*; the condition  $\lambda_1 > 0$  implies exponential sensitivity with respect to initial conditions, as well as local moving-frame saddle-type behavior, for the (random) trajectories of  $\mu$ -typical initial  $x$  corresponding to almost every random sample  $\omega \in \Omega$  (see [11, 94]; see also [6, 49, 63] for emphasis on random dynamics). This nonlinear moving-frame behavior is the subject of *smooth ergodic theory*, and leans on the Multiplicative Ergodic Theorem [74, 80, 90], which, roughly speaking, provides a decomposition of the tangent bundle  $T\mathbb{R}^d$  into (random) sub-bundles along which various exponential growth rates (Lyapunov exponents) are realized. The value  $\lambda_\Sigma$  is the *sum Lyapunov exponent* and describes the asymptotic exponential rate at which Lebesgue volume is contracted/expanded by the dynamics. For more information, see, e.g., the expositions [93, 95].

### Fisher information identity

The purpose of this paper is to put forward a new method for obtaining lower bounds on  $\lambda_1$  based on the regularity of a certain stationary density. Results are framed in terms of the augmented Markov process  $(x_t, v_t)$  tracking a trajectory in phase space  $(x_t)$  and the tangent direction

$$v_t := \frac{D_x \Phi_\omega^t v}{|D_x \Phi_\omega^t v|},$$

for a fixed initial unit vector  $v_0 := v$ . The process  $(v_t)$  is natural in the study of Lyapunov exponents, on noticing by the chain rule that convergence of the finite-time Lyapunov exponents  $\frac{1}{n} \log |D_x \Phi_\omega^n v|$  is connected to a strong law of large numbers for the observable  $g_\omega(x, v) = \log |D_x \Phi_\omega^1 v|$  of the Markov process  $(x_t, v_t)$ . We call  $(w_t) = (x_t, v_t)$  the *projective process* on the unit tangent bundle<sup>1</sup>  $\mathbb{SR}^n = \mathbb{R}^n \times \mathbb{S}^{n-1}$ . For more background, see, e.g., [59, 90]; see also Theorem III.1.2 of [49] and [26] in the context of random dynamical systems. The projective process  $(w_t)$  solves a corresponding SDE on  $\mathbb{SR}^n$ ,

<sup>1</sup> The distinction between  $v_t$  or  $-v_t$  is irrelevant for Lyapunov exponents, and so morally  $(w_t)$  should be thought of as evolving on the projective bundle  $P\mathbb{R}^n = \mathbb{R}^n \times P^{n-1}$ , where  $P^{n-1} = \mathbb{S}^{n-1}/\pm$  is the projective space of  $\mathbb{R}^n$ . However, in this work we adopt the (technically more convenient) convention that  $(w_t)$  lives on  $\mathbb{SR}^n$ .

$$dw_t = \tilde{X}_0(w_t)dt + \sum_{k=1}^r \tilde{X}_k(w_t) \circ dW_t^k, \quad (1.2)$$

where  $\tilde{X}_j$  denotes the ‘lift’ of the vector field  $X_j$  to a vector field on  $\mathbb{SR}^n$  (see Sect. 2.1 for a precise definition). Our first main result is an identity connecting stationary densities of  $(w_t)$  to the exponent  $\lambda_1$  through a partial *Fisher information*-type quantity defined for a probability density  $f$  on  $\mathbb{SR}^n$  by

$$FI(f) := \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{SM}} \frac{|\tilde{X}_k^* f|^2}{f} dq,$$

where  $dq$  is the (Riemannian) volume measure on  $\mathbb{SR}^n$  and  $\tilde{X}^*$  denotes the  $L^2(dq)$ -adjoint of  $\tilde{X}$  as a differential operator.

**Theorem A** *Assume  $(w_t)$  admits a unique stationary density  $f > 0$  on  $\mathbb{SR}^n$  satisfying some additional mild moment estimates (see Proposition 3.2 for details). Let  $\rho$  denote the corresponding stationary density of  $(x_t)$  on  $\mathbb{R}^n$ . Then,*

$$\begin{aligned} FI(f) &= n\lambda_1 - 2\lambda_\Sigma, \\ FI(\rho) &= -\lambda_\Sigma. \end{aligned}$$

This formula can be interpreted as the time-infinitesimal analogue of the well-known relative entropy formula for Lyapunov exponents for SDE [12, 35] (see also [60]) and in many ways provides a missing link between these relative entropy formulas and the Furstenberg-Khasminskii formula (Lemma C.4; see also [6, 47]). For more on the relative entropy formula, see equation (1.10) and the discussion in Sect. 1.2. A complete statement and proof of Theorem A is given in Sect. 3.

### *Fisher information and hypoelliptic regularity*

If, hypothetically,  $\{\tilde{X}_1, \dots, \tilde{X}_k\}$  were to span the tangent space of  $\mathbb{SR}^n$  everywhere in a ‘uniform’ way, then we would obtain the easy lower-bound

$$\|\nabla f\|_{L^1}^2 \lesssim \|f\|_{L^1} FI(f) = FI(f), \quad (1.3)$$

i.e., the Fisher information  $FI(f)$  dominates  $W^{1,1}$  regularity of the stationary density  $f$  over all of  $\mathbb{SR}^n$ , thereby providing a non-vanishing lower bound on the Lyapunov exponent. Unfortunately, for most cases of interest, the vectors

$\{\tilde{X}_k(w)\}_{k=1}^r$  do not span  $T_w \mathbb{SR}^n$  at each  $w \in \mathbb{SR}^n$  (even if  $\{X_k(x)\}_{k=1}^r$  do span  $T_x \mathbb{R}^n$ ), and so the functional  $FI(f)$  only directly controls *partial regularity* along the forced directions  $\{\tilde{X}_k\}_{k=1}^r$  and is degenerate along unforced directions. In particular, there is no simple way to relate  $FI(f)$  to, e.g., Sobolev regularity of  $f$  in all directions on  $\mathbb{SR}^n$  without using more information about  $f$ .

Our second main result shows that this degeneracy in  $FI(f)$  can be resolved if one assumes that the degenerate elliptic Kolmogorov equation governing  $f$ ,

$$X_0^* f + \frac{1}{2} (X_k^*)^2 f = 0, \quad (1.4)$$

is *hypoelliptic*. Hypoellipticity of this kind was studied by Hörmander [41], who showed that under a checkable *Lie bracket spanning* condition on the Lie algebra generated by the vector fields  $\{\tilde{X}_k\}_{k=0}^r$  at each point  $w \in \mathbb{SR}^n$ , regularity of  $f$  along the  $\{\tilde{X}_k\}_{k=1}^r$  directions propagates to regularity along *all* directions in  $\mathbb{SR}^n$ . One of our main contributions in this paper is an analogous and quantitative version of Hörmander's result for the Fisher information  $FI(f)$ : assuming  $\{\tilde{X}_k\}_{k=0}^r$  satisfies the parabolic version of Hörmander's bracket spanning condition (Definition 2.1 below) and that  $f$  is a stationary probability density for  $(w_t)$  solving the PDE (1.4), we establish that there exists a regularity parameter  $s \in (0, 1)$  such that for any open, bounded ball  $U \subset \mathbb{SR}^n$ , there exists a constant  $C = C_U > 0$  with

$$\|\chi_U f\|_{W^{s,1}}^2 \leq C(1 + FI(f)), \quad (1.5)$$

where  $\chi_U$  is a smooth cutoff subordinate<sup>2</sup> to  $U$ .

Equation (1.5) is not immediately useful in isolation, as it is unclear how to directly use this to obtain information about  $n\lambda_1$  from  $FI(f)$ . For instance, the constant  $C$  depends on the details of precisely how the Lie brackets are put together to span the tangent space and is difficult to control. Moreover (1.5) does not preclude the vanishing of  $FI(f)$  as an inequality like (1.3) would. However, we can obtain a far more useful estimate applicable for SDE in small noise regimes, e.g.,

$$dx_t^\epsilon = X_0^\epsilon(x_t^\epsilon) dt + \sqrt{\epsilon} \sum_{k=1}^r X_k^\epsilon(x_t^\epsilon) \circ dW_t^k,$$

<sup>2</sup> i.e.  $\exists V \subset \mathbb{SR}^n$  open and bounded such that  $\overline{U} \subset\subset V$ ,  $\chi_U(x) = 1$  for  $x \in U$  and  $\chi_U(x) = 0$  for  $x \notin V$ . The constant in (1.5) will depend on the choice of  $V$  and  $\chi_U$ , but this dependence has been suppressed as it is unimportant.

where  $\epsilon \in (0, 1)$  is a small parameter and  $\{X_k^\epsilon\}_{k=0}^r$  is a parametrized family of uniformly smooth vector fields. In this context, the Fisher information identity in Theorem A reads

$$FI(f^\epsilon) = \frac{n\lambda_1^\epsilon - 2\lambda_\Sigma^\epsilon}{\epsilon}, \quad (1.6)$$

provided that for all  $\epsilon$ , the corresponding projective process  $(x_t^\epsilon, v_t^\epsilon)$  admits a unique stationary density  $f^\epsilon$  on  $\mathbb{SR}^n$ . Our second main result is that assuming that  $\{\tilde{X}_0^\epsilon, \dots, \tilde{X}_r^\epsilon\}$  satisfies Hörmander's hypoellipticity condition uniformly-in- $\epsilon$  (locally in  $x$ ), we may obtain an estimate which is independent of  $\epsilon$ .

**Theorem B** *Assume that  $\{\tilde{X}_k^\epsilon\}_{k=0}^r$  satisfies Hörmander's parabolic bracket spanning condition uniformly in  $\epsilon \in (0, 1)$  (see Definition 4.1). Then, there exists  $s \in (0, 1)$  and for each bounded open set of the form  $U := B_R(x_0) \times \mathbb{S}^{n-1} \subset \mathbb{SR}^n$ , a constant  $C = C_U > 0$  (independent of  $\epsilon > 0$ ) such that for all  $\epsilon \in (0, 1)$  and for any absolutely continuous stationary measure  $f^\epsilon$  of  $(x_t^\epsilon, v_t^\epsilon)$ , we have the estimate*

$$\|\chi_U f^\epsilon\|_{W^{s,1}}^2 \leq C(1 + FI(f^\epsilon)).$$

In view of (1.6), it is clear that Theorem B is most useful for weakly-damped systems, for example, systems where  $\lambda_\Sigma = -O(\epsilon)$ . Crucially however, it does not have to be exactly zero. In this manner, we can treat systems which are close to, but not exactly, volume preserving (see the discussion in Sect. 1.2 for more detail).

The proof of Theorem B is carried out in Sect. 4 and requires a significant reworking of Hörmander's work to pass from the original  $L^2$  framework to an  $L^1$ -compatible framework more suited to estimates involving Fisher information. Of course, there is a large literature of works extending Hörmander's theory in various ways, e.g., to handle rough coefficients: we refer the reader to, e.g., [1, 3, 18, 38, 51, 56, 71] and the references therein. However, as far as the authors are aware, there are no previous works that fundamentally rework the theory into  $L^1$ .

### Application to “Euler-like” bilinear systems

We anticipate that Theorems A and B are applicable to a wide class of weakly damped, weakly-driven SDE. In this manuscript, we present an application to a natural class of examples with bilinear drift term, which we call *Euler-like* systems:

$$dx_t^\epsilon = (B(x_t^\epsilon, x_t^\epsilon) + \epsilon Ax_t^\epsilon)dt + \sum_{k=1}^r X_k dW_t^k. \quad (1.7)$$

Here,  $\{X_k\}_{k=1}^r$  is a collection of constant ( $x$ -independent) forcing vector fields, while  $B : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  is bilinear, nontrivial (not identically zero), and satisfies

$$\operatorname{div} B = 0, \quad x \cdot B(x, x) = 0,$$

so in particular the  $\epsilon = 0$  dynamics is norm (“energy”) and volume preserving. Meanwhile, the term  $\epsilon A$  provides weak linear damping, where  $A$  is assumed to be a symmetric, negative-definite  $n \times n$  matrix. Stochastically forced versions of both Lorenz 96 and Galerkin truncations of 2d and 3d Navier–Stokes on a torus can be cast in this form.

*Remark 1.1* Note the absence of  $\sqrt{\epsilon}$  in front of the noise term in (1.7); in this class of models, the dynamics is subjected to weak dissipation of order  $\epsilon$  at a constant level of noisy driving. However, due to the bilinear form of the nonlinearity in the drift term, (1.7) is equivalent to the weakly-driven, weakly damped form

$$dx_t^\epsilon = (B(x_t^\epsilon, x_t^\epsilon) + \epsilon Ax_t^\epsilon)dt + \sqrt{\epsilon} \sum_{k=1}^r X_k dW_t^k, \quad (1.8)$$

by rescaling  $x_t \mapsto \sqrt{\epsilon} x_{\sqrt{\epsilon} t}^\epsilon$ , replacing  $\epsilon \mapsto \epsilon^{3/2}$ , and using the self-similarity of Brownian motion. This rescaling does not affect our results on Lyapunov exponents, since upon setting  $\hat{\epsilon} = \epsilon^{3/2}$ , the Lyapunov exponent  $\hat{\lambda}_1^{\hat{\epsilon}}$  of (1.8) with parameter  $\hat{\epsilon}$  is related to the Lyapunov exponent  $\lambda_1^\epsilon$  of (1.7) by the identity

$$\frac{\hat{\lambda}_1^{\hat{\epsilon}}}{\hat{\epsilon}} = \frac{\lambda_1^\epsilon}{\epsilon}.$$

Moreover, while (1.7) is common among models of complex real-world systems, we note that in this scaling, the stationary measure  $\mu$  has characteristic energy  $\int |x|^2 d\mu(x) \approx \epsilon^{-1}$ . Since we are concerned with the regime  $\epsilon \ll 1$ , it is natural to consider the weakly-damped, weakly-driven rescaling (1.8) which has characteristic energy  $O(1)$ .

For this class of systems (1.7), we give a sufficient condition for a positive Lyapunov exponent in terms of the bracket spanning condition for the lifted vector fields  $\{\tilde{X}_0^\epsilon, \tilde{X}_1, \dots, \tilde{X}_r\}$  on  $\mathbb{SR}^n$  corresponding to the projective process  $(x_t^\epsilon, v_t^\epsilon)$ , where here we follow the convention  $X_0^\epsilon(x) = B(x, x) + \epsilon Ax$ .

**Theorem C** Assume  $\{\tilde{X}_k\}_{k=0}^r$  satisfies Hörmander's parabolic bracket spanning condition uniformly in  $\epsilon \in (0, 1)$ , and additionally, assume that for all  $\epsilon \in (0, 1)$  we have that  $(x_t^\epsilon, v_t^\epsilon)$  admits a unique stationary density  $f^\epsilon$ . Then,

$$\lim_{\epsilon \rightarrow 0} \frac{\lambda_1^\epsilon}{\epsilon} = \infty.$$

The basic idea of the proof is as follows. If  $\liminf \lambda_1^\epsilon / \epsilon < \infty$ , then Theorem B implies that the stationary densities  $f^\epsilon$  of the projective process are uniformly bounded in  $W_{\text{loc}}^{s,1}$  along some subsequence of  $\epsilon \rightarrow 0$ . Using the locally compact embedding of  $W^{s,1}$  in  $L^1$  (Lemma A.3), one easily establishes that the sequence of  $f^\epsilon$  are precompact<sup>3</sup> in  $L^1$ , and on refining the subsequence we can find an  $L^1$  limit  $f^0$ , which as one easily checks, is an invariant probability density for the zero noise, deterministic projective process  $(x_t^0, v_t^0)$ . In contrast with the noisy setting, the existence of an invariant density for the projective process of a deterministic flow is extremely rigid and is equivalent, roughly speaking, to the flow being an isometry with respect to a certain Riemannian metric (Theorem 6.2). This can be ruled out for the zero-noise flow due to shearing between energy surfaces  $\{|x| = \text{const}\}$  (a natural consequence of the quadratic nonlinearity), ruling out the possibility that  $(D_x \Phi^t)$  is an isometry. Hence, the proof of Theorem C is by a compactness-rigidity argument. For more details and the full proof of Theorem C, see Sect. 6.

In this manuscript we confirm the sufficient conditions in Theorem C for the Lorenz 96 (L96) model<sup>4</sup>

$$du_m = ((u_{m+1} - u_{m-2})u_{m-1} - \epsilon u_m)dt + q_m d\hat{W}_t^m, \quad 1 \leq m \leq J, \quad (1.9)$$

on  $\mathbb{R}^J$ , where  $\{q_m\}$  are fixed parameters and the  $u_m$  are  $J$ -periodic in  $m$ , i.e.,  $u_{m+kJ} := u_m$  [65].

**Corollary D** Assume  $J \geq 7$  and that  $q_1, q_2 \neq 0$ . Then, the top Lyapunov exponent  $\lambda_1^\epsilon$  of the L96 model (1.9) satisfies  $\lambda_1^\epsilon / \epsilon \rightarrow \infty$  as  $\epsilon \rightarrow 0$ . In particular,  $\lambda_1^\epsilon > 0$  for all  $\epsilon$  sufficiently small.

Remarkably, the problem of proving  $\lambda_1^\epsilon > 0$  for L96 was previously open in spite of overwhelming numerical evidence to support this [23, 46, 67, 75,

<sup>3</sup> We stress that it is not enough that  $\{f^\epsilon dq\}$  is merely tight as a sequence of measures on  $\mathbb{SR}^n$ , which would not ensure that the limiting measure has a density. Rather, we vitally use strong  $L^1$  convergence coming from the uniform regularity estimate to ensure that the limiting measure we obtain has a density w.r.t.  $dq$ . Of course, one could replace strong convergence in  $L^1$  with weak  $L^1$  convergence via an weaker uniform integrability assumption and still obtain a limit in  $L^1$ .

<sup>4</sup> Note that L96 is distinct from the Lorenz 63 “butterfly attractor” model, an ODE on  $\mathbb{R}^3$ , introduced in Lorenz's seminal 1963 work [64].

[77]. To the best of the authors’ knowledge, ours is the first mathematically rigorous result in this direction. Moreover,  $\lambda_1^\epsilon > 0$  does not follow from existing techniques, e.g., Furstenberg’s criterion for the Lyapunov exponents of a random system; see Sect. 1.2 for more discussion. See also the beginning of Sect. 1.3 for discussion of a very recent application of Theorem C to Galerkin truncations of the 2d Navier–Stokes equations on a periodic box in the recent work [19] of the first and third authors of this paper. See Sect. 6 for the proof of Corollary D.

## 1.2 Context within prior work

As remarked earlier, for a given system it can be extremely challenging to estimate its Lyapunov exponents and provide a mathematically rigorous account of its time-asymptotic behavior. Indeed, in principle Lyapunov exponents require infinitely precise information on infinitely many trajectories, and in practice the convergence of Lyapunov exponents to their ‘true’ values can exhibit long stretches of intermittent behavior. This is especially so for deterministic systems in the absence of stochastic driving, for which one anticipates that “chaotic” and “orderly” regimes coexist in a convoluted way in both phase space as well as ‘parameter space’, i.e., as the underlying dynamical system is varied: we refer the interested reader to, e.g., work on Newhouse phenomena in dissipative systems [72, 73]; the proliferation of elliptic islands in volume-preserving systems [33]; known coexistence of chaotic and ordered regimes for the quadratic map family [66]; and  $C^1$  generic dichotomies [21, 22]. For more background on this rich topic, see, e.g., [29, 78, 93, 95].

Although it still presents significant challenges, the situation for Lyapunov exponents of stochastically forced systems is notably more tractable. To start, let us first address the body of work à la Furstenberg which describes necessary conditions for ‘degeneracy’ of the Lyapunov exponents of a random dynamical system. As before, consider a stochastic flow of diffeomorphisms  $\Phi_\omega^t$  on  $\mathbb{R}^n$  arising from an SDE. Let  $\mu$  be the (presumed unique) stationary measure for  $x_t := \Phi_\omega^t(x_0)$ , and let  $\lambda_1, \lambda_\Sigma$  be the corresponding Lyapunov exponents.

Note that unconditionally we have  $n\lambda_1 - \lambda_\Sigma \geq 0$ . In this context, and brushing aside technical details, the criterion à la Furstenberg is due to a variety of authors (e.g., [27, 60, 84, 89]), and can be stated as follows: if  $\nu \in \mathcal{P}(\mathbb{S}M)$  is a stationary measure for the projective process  $(x_t, v_t)$  and  $d\nu(x, v) = d\nu_x(v)d\mu(x)$  the disintegration of  $\nu$ , then for all  $t > 0$  there holds (c.f. Theorem 3.6)

$$\mathbf{E} \int_M H(D_x \Phi_*^t \nu_x | \nu_{\Phi^t(x)}) d\mu(x) \leq t(n\lambda_1 - \lambda_\Sigma), \quad (1.10)$$

where  $H$  denotes the relative entropy, defined for any two probability measures  $\eta, \lambda$  with  $\eta \ll \lambda$ , by

$$H(\eta|\lambda) := \int \log \left( \frac{d\eta}{d\lambda} \right) d\eta, \quad (1.11)$$

Unconditionally, relative entropy satisfies  $H(\eta|\lambda) \geq 0$ , and  $H(\eta|\lambda) = 0$  iff  $\eta = \lambda$  by Jensen's inequality. From this we see that either

$$n\lambda_1 - \lambda_\Sigma > 0, \quad (1.12)$$

or the probabilistic law governing the stochastic flow admits a strong ‘degeneracy’ in the sense that

$$(D_x \Phi_\omega^t)_* \nu_x = \nu_{\Phi_\omega^t(x)}, \quad (1.13)$$

with probability 1 for all  $t \geq 0$  and  $\mu$ -typical  $x$ . That this situation is very ‘degenerate’ follows from the fact that for fixed  $x$  and  $t$ , the above right-hand side depends only on the time- $-t$  position  $\Phi_\omega^t(x)$ , while the left-hand side depends additionally on the entire noise path  $\omega|_{[0,t]}$ .

Observe that in the weakly-damped, weakly-driven setting of (1.7),  $\lambda_\Sigma^\epsilon = \epsilon \operatorname{tr} A < 0$  and so (1.12) is agnostic as to whether  $\lambda_1^\epsilon > 0$  or not. Indeed, the techniques in the above-mentioned works are ‘soft’ as the identity (1.13) is non-quantitative in the parameters of the underlying system. Although (1.10) does at least provide some kind of formula for  $n\lambda_1 - \lambda_\Sigma$ , it is unclear how to glean useful quantitative information directly from (1.10).

Interestingly, our Fisher-information identity in Theorem A, (specifically (3.2) below), is essentially the time-infinitesimal analogue of (1.10), as we show below in Sect. 3.2. Hence, like (1.10), our Theorem A admits an interpretation in terms of the *rate at which the degeneracy (1.13) fails to hold* for the stochastic flow  $\Phi_\omega^t$ . However, Theorem A recasts the information in terms of the generator of  $(w_t)$ , which is more amenable now to the use of hypoelliptic PDE methods such as those employed in Theorem B. This motivates the claim that the methods in this paper constitute a first step towards a quantitative à la Furstenberg theory. We remark that Fisher information-type quantities also commonly appear as the time derivatives of the relative entropy in the study of gradient flows and logarithmic Sobolev inequalities (see e.g. [10, 58, 81, 87]).

Beyond Furstenberg's criterion and its descendants, there is by now a large literature on the Lyapunov exponents of particular models for which we cannot do justice in this space. Instead, we will focus on a class of results most closely related to ours (Theorem C): small-noise expansions of Lyapunov exponents for weakly-driven stochastic systems. To frame the discussion, consider the abstract linear SDE

$$dV_t = A_t^\epsilon V_t dt + \sqrt{\epsilon} \sum_{k=1}^r B_t^k V_t \circ dW_t^k, \quad (1.14)$$

where  $A_t^\epsilon, B_t^k$  are, in general, time-varying and/or themselves randomly driven, and  $A_t^\epsilon$  may or may not exhibit some vanishingly weak damping as  $\epsilon \rightarrow 0$ . There are many works studying the scaling behavior of Lyapunov exponent  $\lambda_1^\epsilon := \lim_{t \rightarrow \infty} \frac{1}{t} \log |V_t|$  of such systems, e.g., [9, 42, 70, 76, 79] in the constant coefficient case, and [8, 15, 16, 68] when the  $A_t, B_t^k$  are coupled to some other stochastic process. To the authors' best knowledge, however, all of these results are restricted to settings where the  $\epsilon = 0$  dynamics are relatively simple and essentially completely known. In comparison, our results are indifferent to any detailed description of the zero-noise dynamics. On the other hand, the sacrifice for our level of generality is that our estimate  $\lambda_1^\epsilon/\epsilon \rightarrow \infty$  is far weaker than an asymptotic expansion, and is likely to be sub-optimal for many models of interest.

Of particular interest is that among models of the form (1.14), scaling laws of the form  $\lambda_1^\epsilon \sim \epsilon^\gamma, \gamma \geq 1$  tend to be associated with zero-noise dynamics which are *rigid isometries* (exhibiting no shearing) [9, 14, 15, 76]. Meanwhile, laws of the form  $\lambda_1^\epsilon \sim \epsilon^\gamma, \gamma < 1$  are associated with zero-noise dynamics exhibiting some shearing mechanism. By way of example, [9, 79] derive such scaling laws when  $A_t$  as above is given by

$$A_t \equiv \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B_t \equiv \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

corresponding to the constant application of a horizontal shear in conjunction with a small, stochastically driven vertical shear. This analysis was extended to the setting of fluctuation-dissipation zero-noise limits of certain 2d completely integrable Hamiltonian systems in the work [16]. Despite the consideration of high dimensional systems which are unlikely to be completely integrable, our Theorem C is in fact related to these works, in the sense that the scaling  $\lambda_1^\epsilon/\epsilon \rightarrow \infty$  is derived by taking advantage of *shearing between the energy surfaces*  $|x| = R, R > 0$  (see Sect. 6.2 for details).

Of course, shearing has long been regarded as a potential mechanism for the generation of chaotic behavior. As early as the late 70's it was realized that chaotic attractors could arise from time-periodic driving of a system undergoing a Hopf bifurcation [96], while subsequent mathematically rigorous work has confirmed this mechanism (see, e.g., [92] for an overview of this program). We also point out the work [62], which provides a mix of heuristics, numerics, mathematical analysis and conjectures demonstrating the shearing mechanism as a source of chaotic behavior, as well as the works [31, 34] resolving some of the open problems in [62] pertaining to white-noise models.

### 1.3 Conclusion and outlook for future work

We provide here some discussion on potential future extensions and applications of the ideas in this paper.

**Applications to different systems.** As mentioned earlier, the class of Euler-like systems (1.7) includes Galerkin truncations of the Navier–Stokes equation on the periodic box in 2d and 3d, while Theorem C provides a positive lower bound on the top Lyapunov exponent provided that the corresponding projective process satisfies a uniform Hörmander bracket spanning condition. Recently, the first and third author [19] affirmed this spanning condition for the 2d Galerkin–Navier–Stokes equations with additive forcing on rectangular 2d tori when only a small number of modes are directly forced and assuming that the Galerkin truncation is sufficiently large (i.e. the results hold for all sufficiently high dimensions).

In contrast with the Lorenz 96 model, the coupling between modes of Galerkin truncations of PDE models is more ‘global’, with low modes strongly coupled to high ones, and so for models of this kind it is substantially more difficult to check the projective Hörmander’s condition; the proof in [19] is in fact computer assisted, but applies to all sufficiently high dimensional Galerkin truncations and contains ideas that should be useful for verifying the projective spanning for other (sufficiently high dimensional) Galerkin truncations of PDEs. However, these ideas alone are likely inadequate for other kinds of high dimensional systems of interest, such as many-particle systems with weak dissipation.

Another class of systems of interest is small random perturbations of completely integrable Hamiltonian flows, where estimates of the form  $\lambda_1^\epsilon \gg \epsilon$  are far closer to optimal. For these, one would replace shearing between energy shells as in the proof of Theorem C with shearing between invariant tori, which in practice can be read directly off of the action-angle coordinates for the system.

**Tighter hypoelliptic regularity estimates.** It is of natural interest to attempt to improve the scaling  $\lambda_1^\epsilon \gg \epsilon$  that naturally falls out from Theorem B. One way to do this is to attempt to strengthen the hypoelliptic regularity estimate by refining the  $\epsilon$  scaling to derive something like

$$\|f^\epsilon\|_{W^{s,1}}^2 \lesssim 1 + \frac{n\lambda_1^\epsilon - 2\lambda_\Sigma^\epsilon}{\epsilon^{1-\gamma}},$$

for some constant  $\gamma > 0$ . If such an estimate were true, the same compactness-rigidity argument of Theorem C would imply a scaling like  $\lambda_1^\epsilon \gtrsim \epsilon^{1-\gamma}$ , a significant improvement. For more discussion on potential improvements to Theorem B, see the discussion in Sect. 4.

**Beyond compactness-rigidity.** Another way to strengthen the results of this paper is to work exclusively with  $\epsilon > 0$  without passing the limit as  $\epsilon \rightarrow 0$ . This was essentially the approach of works [16, 79] estimating Lyapunov exponents for systems for which a nearly-complete understanding of the pathwise random dynamics was available. Although we largely lack such detailed information about the pathwise dynamics of high-dimensional models such as L96, there is some hope for a ‘middle ground’, e.g., partial information such as some finite-time exponential growth mechanism resulting in a lower bound on  $\|f^\epsilon\|_{W^{s*,1}}$ . An approach with a similar flavor for random perturbations of discrete-time systems, including the Chirikov standard map, was carried out in the previous work [20].

**Finer dynamical information: moment Lyapunov exponents.** Lyapunov exponents themselves provide asymptotic exponential growth rates, but the timescales along which these rates are realized can be quite long. Some quantitative control is provided by large deviations principles in the convergence of the sequences  $\frac{1}{t} \log |D_x \Phi_\omega^t(v)|$ , the rate function of which is the Legendre transform of the *moment Lyapunov exponent function*  $p \mapsto \Lambda(p) := \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbf{E} |D_x \Phi_\omega^t v|^p$  (the limit defining  $\Lambda(p)$  exists and is independent of  $(x, v)$  under natural conditions; see, e.g., [5]). It would be highly interesting to see if ideas similar to those presented in this work could provide quantitative estimates on the moment Lyapunov exponents of weakly-driven systems.

**More general noise models.** One can also expand the noise models to which our work applies. One simple example of this is to extend Theorem C to different types of multiplicative noise. While Theorems A and B apply to multiplicative noise, certain aspects of Theorem C are specialized to additive noise, such as the arguments for projective spanning and irreducibility in Sect. 5 (and in [19]). Another extension would be to noise models which are not white-in-time, e.g., jump processes or the models used in [53, 54]. Our work is clearly deeply tied to the infinitesimal generator  $\mathcal{L}$ , and so any such extension will be rather non-trivial. A simpler candidate for non-white forcing is constructed from ‘towers’ of coupled Ornstein-Uhlenbeck processes, which can be built to be  $C^k$  in time for any  $k \geq 0$  (see, e.g., [17] for details). While such a noise model is built out of SDE and so largely falls under the purview of the analysis in this paper, it is not clear how to prove the analogue of Theorem C. This will be considered in future research.

## 1.4 Plan for remainder of the paper

Preliminaries on the SDE setting in this paper are provided in Sect. 2, while the complete statement and proof of the Fisher information identity in Theorem A is provided in Sect. 3. In Sect. 4 we turn attention to the full statement and

proof of the hypoelliptic regularity estimate in Theorem B. The remainder of the paper is devoted to the proof of Theorem C and Corollary D: in Sect. 5 we address how some of the assumptions of Theorem A and B (projective spanning and uniqueness of stationary densities) are checked in practice, while the compactness rigidity argument is carried out in Sect. 6.

## 2 Preliminaries

We present here preliminaries on SDE, including a discussion of stationary measures and their properties; the existence of Lyapunov exponents; and remarks on the projective process, a crucial tool in Theorem A.

Throughout,  $(M, g)$  is a smooth, connected, geodesically complete and orientable Riemannian manifold without boundary (not necessarily bounded); let  $n = \dim M$ . Throughout, we abbreviate the Riemannian/Lebesgue volume on  $(M, g)$  by  $dx$ . Here, and everywhere below unless specified otherwise, we use the notation  $\langle a, b \rangle_x = g_x(a, b)$  for  $a, b \in T_x M, x \in M$ .

We consider the stochastic process

$$dx_t = X_0(x_t) dt + \sum_{k=1}^r X_k(x_t) \circ dW_t^k, \quad (2.1)$$

where  $\{X_k\}_{k=0}^r$  are a family of smooth vector fields (potentially unbounded) on  $M$  and  $\{W_t^k\}_{k=1}^r$  are independent standard Wiener processes with respect to a canonical stochastic basis  $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbf{P})$ .

### 2.1 Background on SDEs

The following are standing hypotheses imposed throughout the paper, and ensure the existence with probability 1 of the limits defining Lyapunov exponents.

**Assumption 1** (i) For each initial datum  $x \in M$ , equation (1.1) has a unique global solution  $(x_t)$  with probability 1. The (random) solution maps  $x \mapsto x_t =: \Phi_\omega^t(x)$ ,  $t \geq 0$  comprise a (stochastic) flow of  $C^r$  diffeomorphisms  $(\Phi_\omega^t)$  on  $M$ ,  $r \geq 2$ .

(ii) The Markov process  $(x_t)$  admits a unique, absolutely continuous stationary probability measure  $\mu$  on  $M$ . We write  $\rho = \frac{d\mu}{dx}$  for the density of  $\mu$ .

(iii) The measure  $\mu$  has the integrability condition<sup>5</sup>

$$\mathbf{E} \int_M [\log^+ |D_x \Phi^t| + \log^+ |(D_x \Phi^t)^{-1}|] d\mu(x) < \infty.$$

Although not automatic, Assumption 1(i) is well-studied and follows from mild conditions on (1.1), although careful checking is required in the case when  $M$  is noncompact and the  $\{X_k\}$  are unbounded; see, e.g., [6, 55]. Given (i) and (ii), item (iii) follows under mild moment assumptions [48].

For (ii), we separately address existence, absolute continuity and uniqueness. Existence of stationary measures when  $M$  is compact is immediate from compactness in the narrow topology on the space of probability measures on  $M$ , while when  $M$  is noncompact some additional assumption is needed to ensure tightness as  $t \rightarrow \infty$  of the distribution  $\text{Law}(x_t)$  of  $x_t$ , usually via a Lyapunov-Foster drift condition [69].

Absolute continuity is usually checked using Hörmander's parabolic condition for the vector fields  $\{X_k\}$ , which we briefly recall below. For a manifold  $\mathcal{M}$ , let  $\mathfrak{X}(\mathcal{M})$  denote the set of smooth vector fields on  $\mathcal{M}$ . Elements  $X \in \mathfrak{X}(\mathcal{M})$  are regarded in the usual way as first-order differential operators acting on observables  $w : \mathcal{M} \rightarrow \mathbb{R}$  via  $Xw = dw(X)$ . For vector fields  $X, Y$ , we write  $[X, Y]$  for the standard Lie bracket of  $X$  and  $Y$ .

**Definition 2.1** Given a collection of vector fields  $Z_0, Z_1, \dots, Z_r$  on a manifold  $\mathcal{M}$ , we define collections of vector fields  $\mathcal{X}_0 \subseteq \mathcal{X}_1 \subseteq \dots$  recursively by

$$\begin{aligned}\mathcal{X}_0 &= \{Z_j : j \geq 1\}, \\ \mathcal{X}_{k+1} &= \mathcal{X}_k \cup \{[Z_j, Z] : Z \in \mathcal{X}_k, \quad j \geq 0\}.\end{aligned}$$

We say that  $\{Z_i\}_{i=0}^r$  satisfies the *parabolic Hörmander condition* (also called bracket spanning) if there exists  $k$  such that for all  $w \in \mathcal{M}$ ,

$$\text{span} \{Z(w) : Z \in \mathcal{X}_k\} = T_w \mathcal{M}.$$

Next, we recall a version of Hörmander's theorem for transition kernels.

**Theorem 2.2** (Hörmander's theorem [41]) Assume  $\{X_k\}_{k=0}^r$  satisfy the parabolic Hörmander condition. Then, the transition kernels  $P^t(x, K) := \mathbf{P}(x_t \in K | x_0 = x)$  are all absolutely continuous with respect to  $dx$ .

Given existence and absolute continuity, the Doob-Khasminskii theorem (see e.g. [30]) ensures that uniqueness follows if we have some additional

<sup>5</sup> Here, for  $a > 0$  we write  $\log^+ a := \max\{\log a, 0\}$  for the positive part of  $\log$ .

topological irreducibility condition ensuring that all initial conditions in  $M$  can “access” any open set  $O \subset M$  with positive probability. The version we take in Definition 2.3 is stronger than necessary but suffices for our purposes here.

**Definition 2.3** (*Topological irreducibility*) Consider a stochastic process  $(z_t)$  defined on a complete metric space  $\mathcal{Z}$ . We say  $(z_t)$  is (topologically) irreducible if for every open set  $O \subset \mathcal{Z}$ , initial condition  $z \in \mathcal{Z}$ , and  $t > 0$  there holds

$$\mathbf{P}(z_t \in O | z_0 = z) > 0.$$

For Markov processes coming from SDEs on manifolds, a very common method of proving topological irreducibility is through the Stroock-Varadhan support theorem [86] (see Theorem 5.8 below), which connects supports of transition kernels to a *controllability problem* treating the Brownian paths  $W_t^k$  as control parameters. For more information, see the discussion in Sect. 5.3, where irreducibility for the class of Euler-like models is treated in more detail.

We close this discussion by addressing the existence of Lyapunov exponents for the stochastic flow of diffeomorphisms  $\Phi_\omega^t$ .

**Theorem 2.4** *Assume (1.1) satisfies Assumption 1. Then there exist positive, deterministic constants  $\lambda_1$  and  $\lambda_\Sigma$ , independent of both the random sample  $\omega$  as well as  $x \in M$ , such that for  $\mathbf{P} \otimes \mu$  almost every  $(\omega, x) \in \Omega \times M$  the following limits hold:*

$$\begin{aligned}\lambda_1 &= \lim_{t \rightarrow \infty} \frac{1}{t} \log |D_x \Phi_\omega^t|, \\ \lambda_\Sigma &= \lim_{t \rightarrow \infty} \frac{1}{t} \log |\det D_x \Phi_\omega^t|.\end{aligned}$$

Theorem 2.4 is classical and follows from the Kingman subadditive ergodic theorem [50] and some basic ergodic theory for random dynamical systems [49].

## 2.2 The projective process

As remarked in Sect. 1, the Fisher-information identity we obtain (Theorem A) is framed in terms of the *projective process*, i.e., the process on tangent directions

$$v_t := \frac{D_x \Phi_\omega^t v}{|D_x \Phi_\omega^t v|},$$

so that the full process  $w_t := (x_t, v_t)$  lives on the unit tangent bundle  $\mathbb{S}M \subset TM$  with fibers  $\mathbb{S}_x M \subset T_x M$ . Our aim in this Sect. 2.2 will be to lay out preliminaries framing  $(w_t)$  as the solution to an SDE on  $\mathbb{S}M$ .

We begin with some basic geometry for the manifold  $\mathbb{S}M$ . Using the Riemannian structure on  $M$  and the Levi-Civita connection  $\nabla$ , we equip  $\mathbb{S}M$  with the unique lifted metric  $\tilde{g}$  (the Sasaki metric [85]) such that the bundle projection  $\pi : \mathbb{S}M \rightarrow M$  is a Riemannian submersion. In particular, for each  $w = (x, v) \in \mathbb{S}M$  we can decompose  $T_w \mathbb{S}M$  into a horizontal subspace  $H_w \mathbb{S}M$  of directions transverse to the fibers, identified with  $T_x M$ , and a vertical  $V_w \mathbb{S}M$  subspace of directions along the fibers identified with  $T_v(\mathbb{S}_x M)$ , which itself is isomorphic to the orthogonal complement of  $v$  in  $T_x M$ . The spaces  $H_w \mathbb{S}M$  and  $V_w \mathbb{S}M$  are orthogonal with respect to  $\tilde{g}$ , giving the orthogonal decomposition

$$T_w \mathbb{S}M = T_x M \oplus T_v(\mathbb{S}_x M).$$

With these preliminaries in place, the process  $(v_t)$  satisfies the SDE

$$dv_t = V_{\nabla X_0(x_t)}(x_t, v_t)dt + \sum_{k=1}^r V_{\nabla X_k(x_t)}(x_t, v_t) \circ dW_t^k,$$

where  $\nabla$  denotes the covariant derivative and, for  $x \in M$  and  $A : T_x M \rightarrow T_x M$  linear, the ‘vertical’ vector field  $V_A$  on  $\mathbb{S}M$ ,  $V_A(x, v) \in T_v(\mathbb{S}_x M)$  for each  $(x, v) \in \mathbb{S}_x M$ , is defined by

$$V_A(x, v) := Av - v \langle v, Av \rangle_x =: \Pi_{(x, v)} Av.$$

The full projective process  $(w_t)$  evolves according to

$$dw_t = \tilde{X}_0(w_t)dt + \sum_{k=1}^r \tilde{X}_k(w_t) \circ dW_t^k, \quad (2.2)$$

where the  $\tilde{X}_k$  are vector fields on  $\mathbb{S}M$ , which when expressed in terms of the orthogonal horizontal/vertical splitting read as

$$\tilde{X}_k(x, v) := (X_k(x), V_{\nabla X_k(x)}(x, v)).$$

Throughout, we take on the following assumption regarding  $(w_t)$ .

**Assumption 2** The SDE (2.2) defining the process  $(w_t)$  satisfies Assumptions 1 (i) and (ii). In particular, the SDE defining  $(w_t)$  is globally well-posed for a.e. random sample and every initial data; and the Markov process  $(w_t)$  admits a unique, absolutely continuous stationary measure  $\nu$  on  $\mathbb{S}M$ .

As with other SDE, it is common to check existence and uniqueness of stationary measures via Hörmander's parabolic condition and topological irreducibility, respectively. In particular, bracket spanning for the projective process  $(w_t)$  appears routinely in the random dynamics literature (see, e.g. [12, 32]).

### 2.2.1 Bracket spanning for the projective process

Below we give a sufficient condition for the projective process  $(w_t)$  to satisfy bracket spanning assuming bracket spanning for the base process  $(x_t)$ .

In what follows, we will find it convenient to reformulate the parabolic Hörmander condition using slightly different notation. Below, we write  $\mathfrak{X}(M)$  for the space of smooth vector fields on  $M$ . Recall that given  $X \in \mathfrak{X}(M)$ , the *adjoint representation*  $\text{ad}(X)$  is the linear operator  $\mathfrak{X}(M) \rightarrow \mathfrak{X}(M)$  sending  $Y \mapsto \text{ad}(X)Y := [X, Y]$ .

**Definition 2.5** For a given collection  $\mathcal{F} \subseteq \mathfrak{X}(M)$  define the *Lie algebra generated by  $\mathcal{F}$*  by

$$\text{Lie}(\mathcal{F}) := \text{span}\{\text{Lie}^m(\mathcal{F}) : m \geq 1\},$$

where

$$\text{Lie}^m(\mathcal{F}) := \text{span}\{\text{ad}(X_r) \dots \text{ad}(X_2)X_1 : X_i \in \mathcal{F}, 1 \leq r \leq m\}.$$

**Definition 2.6** Let  $Z_0, Z_1, \dots, Z_r \in \mathfrak{X}(M)$  be smooth vector fields on  $M$ , and define  $\mathcal{X} = \{Z_1, \dots, Z_r\}$ . The *zero-time ideal*  $\text{Lie}(Z_0; \mathcal{X})$  is defined to be the Lie algebra generated by  $\mathcal{X}$  and  $[\mathcal{X}, Z_0] := \{[Z, Z_0] : Z \in \mathcal{X}\}$ ; that is,

$$\text{Lie}(Z_0; \mathcal{X}) = \text{Lie}(\mathcal{X}, [\mathcal{X}, Z_0]).$$

The ideal  $\text{Lie}(Z_0; \mathcal{X})$  plays a significant role in geometric control theory, which we will revisit in Sect. 5.3 when we discuss irreducibility for Euler-like systems. We also note that it is straightforward to check, using the Jacobi identity, that  $\{Z_0, Z_1, \dots, Z_r\}$  satisfies the parabolic Hörmander condition as in Definition 2.1 if

$$\text{Lie}_x(X_0; \mathcal{X}) := \{X(x) : X \in \text{Lie}(X_0; \mathcal{X})\} = T_x M,$$

for all  $x \in M$ .

We now turn to a general sufficient condition (Proposition 2.7 below) on the vector fields  $\{X_k\}_{k=0}^r$  so that their lifts  $\{\tilde{X}_k\}_{k=0}^r$  satisfy the parabolic Hörmander condition on  $\mathbb{S}M$ . Given  $X \in \mathfrak{X}(M)$ , define

$$M_X(x) := \nabla X(x) - \frac{1}{n} \operatorname{div} X(x) I. \quad (2.3)$$

We view  $M_X(x)$  as an element of  $\mathfrak{sl}(T_x M)$ , the Lie algebra of traceless linear operators on  $T_x M$ . Observe that since the projective vector field  $V_{\nabla X}(v)$  includes a projection orthogonal to  $v$ , we have the identity  $V_{\nabla X} \equiv V_{M_X}$ . Define

$$\mathfrak{m}_x(X_0; X_1, \dots, X_r) := \{M_X(x) : X \in \operatorname{Lie}(X_0; X_1, \dots, X_r), X(x) = 0\}. \quad (2.4)$$

Lemma C.1 in the Appendix implies that  $\mathfrak{m}_x(X_0; X_1, \dots, X_r)$  is indeed a Lie sub-algebra of  $\mathfrak{sl}(T_x M)$  with respect to the usual commutator  $[A, B] = AB - BA$  for linear operators.

The following relates bracket spanning for  $\{\tilde{X}_k\}_{k=0}^r$  to the matrix Lie algebra  $\mathfrak{m}_x(X_0; X_1, \dots, X_r)$ .

**Proposition 2.7** *Let  $\{X_k\}_{k=0}^r$  be a collection of smooth vector fields on  $M$ . Their lifts  $\{\tilde{X}_k\}_{k=0}^r$  satisfy the parabolic Hörmander condition on  $\mathbb{S}M$  if and only if  $\{X_k\}_{k=0}^r$  satisfy the parabolic Hörmander condition on  $M$  and for each  $(x, v) \in \mathbb{S}M$  we have*

$$\{V_A(x, v) : A \in \mathfrak{m}_x(X_0; X_1, \dots, X_r)\} = T_v \mathbb{S}_x M. \quad (2.5)$$

Proposition 2.7 was stated without proof in Baxendale's paper [12]. For the sake of completeness, a self-contained proof is included in Appendix C.1.

*Remark 2.8* In the theory of Lie algebra actions on manifolds, the condition (2.5) means that  $\mathfrak{m}_x$  acts transitively on  $\mathbb{S}_x M$  through the Lie algebra action  $A \mapsto V_A$ . It is straightforward to show that the Lie algebra  $\mathfrak{so}(T_x M)$  of skew-symmetric linear operators (depending on the metric) also acts transitively on  $\mathbb{S}_x M$ , and therefore a sufficient condition for transitive action of  $\mathfrak{m}_x(X_0; X_1, \dots, X_r)$  on  $\mathbb{S}_x M$  is that

$$\mathfrak{so}(T_x M) \subseteq \mathfrak{m}_x(X_0; X_1, \dots, X_r).$$

### 2.2.2 Generator for the projective process

We close this preliminary section with some brief comments on the generator for the projective process. Observe that the infinitesimal generator for  $(w_t)$  on  $\mathbb{S}M$  has the following Hörmander form:

$$\tilde{\mathcal{L}} = \tilde{X}_0 + \frac{1}{2} \sum_k \tilde{X}_k^2.$$

In particular, when they exist, stationary densities  $f = \frac{d\nu}{dq}$  solve the Kolmogorov equation

$$\tilde{\mathcal{L}}^* f = \tilde{X}_0^* f + \frac{1}{2} \sum_{k=1}^r (\tilde{X}_k^*)^2 f = 0, \quad (2.6)$$

where  $dq$  denotes the Riemannian volume on  $\mathbb{S}M$  and for a given vector field  $\tilde{X}$  on  $\mathbb{S}M$ ,  $\tilde{X}^*$  denotes the formal adjoint operator with respect to  $L^2(dq)$ . Note that the differential operator  $\tilde{X}^*$  can be related to  $\tilde{X}$  and  $\operatorname{div} \tilde{X}$  through the relation

$$\tilde{X}^* h = -\tilde{X} h - (\operatorname{div} \tilde{X}) h, \quad h \in C_c^\infty(\mathbb{S}M). \quad (2.7)$$

### 3 Fisher information formula for Lyapunov exponents

Here we give a full, general statement of our first main result, a version of Theorem A on manifolds relating the values  $\lambda_1, \lambda_\Sigma$  to the partial Fisher information

$$FI(\phi) := \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} \frac{|\tilde{X}_k^* \phi|^2}{\phi} dq,$$

for  $\phi = f$ , the density of the stationary measure  $\nu$  of  $(w_t)$ , and for  $\phi = \rho$ , the density of the stationary measure  $\mu$  for  $(x_t)$  (abusing notation somewhat and viewing  $\rho$  as a function on  $\mathbb{S}M$ ).

*Remark 3.1* Note that  $FI(\phi)$  is well defined even when the density  $\phi$  is not supported everywhere. Indeed, it can be re-written as

$$FI(\phi) = \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} |X_k \log \phi + \operatorname{div} X_k|^2 \phi dq,$$

which is implicitly integrated over  $\{\phi > 0\}$  so that  $\log \phi$  is well-defined (see Remark 3.4 below for how this logarithmic gradient form relates to metric independence of the Fisher information).

Throughout, the setting is as laid out in the beginning of Sect. 2.

In what follows, for a given partition of unity  $\{\chi_i\}_{i \geq 1}$  of  $\mathbb{S}M$ , we denote  $\chi_{\leq N} = \sum_{i=1}^N \chi_i$ .

**Proposition 3.2** (Fisher Information Identity) *Let Assumptions 1 and 2 hold. Moreover, assume that*

(a) there exists a smooth partition of unity  $\{\chi_i\}_{i \geq 1}$  of  $\mathbb{S}M$  and a measurable function  $m : \mathbb{S}M \rightarrow \mathbb{R}_+$  such that  $|\tilde{\mathcal{L}}\chi_{\leq N}(w)| \leq m(w)$  and

$$\int_{\mathbb{S}M} f |\log f| m \, dq < \infty;$$

(b)  $Q \in L^1(\mu)$  and  $\tilde{Q} \in L^1(\nu)$ , where  $Q, \tilde{Q}$  are defined by

$$\begin{aligned} Q(x) &:= \operatorname{div} X_0(x) + \frac{1}{2} \sum_{k=1}^r X_k \operatorname{div} X_k(x), \\ \tilde{Q}(w) &:= \operatorname{div} \tilde{X}_0(w) + \frac{1}{2} \sum_{k=1}^r \tilde{X}_k \operatorname{div} \tilde{X}_k(w). \end{aligned}$$

Then, the following formulas hold:

$$\begin{aligned} FI(\rho) &= - \int_M Q \, d\mu = -\lambda_\Sigma, \\ FI(f) &= - \int_{\mathbb{S}M} \tilde{Q} \, d\nu = n\lambda_1 - 2\lambda_\Sigma. \end{aligned} \tag{3.1}$$

Equivalently, writing  $h_x(v) = f(x, v)/\rho(x)$  for the conditional density on  $\mathbb{S}_x M$ , we have

$$\begin{aligned} FI(f) - FI(\rho) &= \frac{1}{2} \sum_{k=1}^r \int_M \left( \int_{\mathbb{S}_x M} \frac{|(X_k - V_{\nabla X_k(x)}^*) h_x(v)|^2}{h_x(v)} \, dv \right) d\mu(x) \\ &= n\lambda_1 - \lambda_\Sigma. \end{aligned} \tag{3.2}$$

In each line of (3.1), the second equality is a version of the famous Furstenberg-Khasminskii formula (see, e.g., [6,47]) for the Lyapunov exponents of an SDE satisfying Assumptions 1 and 2; for the convenience of the reader we recall this formula in Lemma C.4 in the Appendix. Assumptions (a) and (b) are technical and needed only when  $M$  is noncompact to justify a certain integration-by-parts step; see below for details.

After some remarks, Proposition 3.2 is derived below. We give two proofs: the first (Sect. 3.1) is a combination of the Furstenberg-Khasminskii formula [6,47] with the Kolmogorov equation (2.6) for  $f$ ; the second proof (Sect. 3.2) is a sketch connecting  $FI(f)$  to a certain relative entropy formula for Lyapunov exponents [12,35] (see also [60]) intimately connected with Furstenberg's criterion (see Sect. 1.2).

*Remark 3.3* Equation (3.2) is an equivalent formulation highlighting the relation to the natural quantity  $n\lambda_1 - \lambda_\Sigma$  and similarities to Furstenberg's criterion for the Lyapunov exponents of stochastic systems (Sect. 1.2). In particular, note that one always has  $FI(f) - FI(\rho) \geq 0$  and that  $n\lambda_1 > \lambda_\Sigma$  if and only if  $FI(f) - FI(\rho) > 0$ .

*Remark 3.4* We emphasize that the Fisher information  $FI(f)$  does not depend on the choice of Riemannian metric  $g$  on  $M$ , as one would expect: after all, it is straightforward to check that Lyapunov exponents themselves are independent of the Riemannian metric  $g$  when  $M$  is compact or under a mild integrability condition when  $M$  is noncompact.

To see this, we will cast  $FI(f) = \int_{\mathbb{S}M} (|\tilde{X}_k^* f|^2/f) dq$  as a *logarithmic derivative* of  $\nu$  with respect to  $\tilde{X}_k^*$ . To make this precise, define the weak derivative  $\tilde{X}_k^* \nu$ , a signed measure on  $\mathbb{S}M$ , by duality (c.f. Chapter 3 of [24]) via

$$\int \phi \, d(\tilde{X}_k^* \nu) = \int \tilde{X}_k \phi \, d\nu, \quad \phi \in C_c^\infty(\mathbb{S}M).$$

When  $\tilde{X}_k^* \nu \ll \nu$ , the Radon–Nikodym derivative  $\beta_{\tilde{X}_k^*}^\nu := \frac{d\tilde{X}_k^* \nu}{d\nu}$  exists and satisfies

$$\beta_{\tilde{X}_k^*}^\nu = \frac{\tilde{X}_k^* f}{f} = -X_k \log f - \operatorname{div} X_k,$$

when  $d\nu = f \, dq$ , as we assume in this paper. This means that we can write the Fisher information as

$$FI(f) = \frac{1}{2} \sum_{k=1}^r \left\| \beta_{\tilde{X}_k^*}^\nu \right\|_{L^2(\nu)}^2,$$

which has no explicit dependence on the Riemannian metric  $\tilde{g}$ . Remarkably, this form makes sense even when  $\nu$  is not absolutely continuous with respect to  $dq$  and only seems to require differentiability of the measure along  $\tilde{X}_k$ . This suggests the potential extension of Proposition 3.2 to infinite dimensional settings where no notion of volume measure  $dq$  is available. However, this is beyond our scope here and left to future work.

*Remark 3.5* The Fisher information is a fundamental quantity in the theory of statistical inference and information geometry (see [2]). It is typically used to measure the amount of information a parametrized family of laws (e.g., the law of one variable conditioned on the value of another) carries about the inference parameter. In our case, the Fisher information formula (3.2) can be interpreted

as the amount of information about  $x$ , on average, which can be inferred by making observations on the family of conditional densities  $\{h_x(\cdot)\}_{x \in M}$  in the  $v$  coordinate.

### 3.1 Proof of Proposition 3.2

We begin with a formal argument. Pairing the Kolmogorov equation (2.6) for  $f$  with  $\log f$  and integrating gives

$$-\frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} (\log f)(\tilde{X}_k^*)^2 f \, dq = \int_{\mathbb{S}M} (\log f)\tilde{X}_0^* f \, dq.$$

Ignoring for the moment that  $f$  is not compactly supported, integrating by parts and using (2.7) gives for the left hand side

$$\begin{aligned} -\frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} (\log f)(\tilde{X}_k^*)^2 f \, dq &= -\frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} \frac{(\tilde{X}_k f)(\tilde{X}_k^* f)}{f} \, dq \\ &= FI(f) + \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} (\operatorname{div} \tilde{X}_k) X_k^* f \, dq \\ &= FI(f) + \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} (\tilde{X}_k \operatorname{div} \tilde{X}_k) f \, dq. \end{aligned}$$

For the right hand side integration by parts gives

$$\int (\log f)\tilde{X}_0^* f \, dq = \int \tilde{X}_0 f \, dq = - \int (\operatorname{div} \tilde{X}_0) f \, dq,$$

where we used the fact that  $\int_{\mathbb{S}M} \tilde{X}_0^* f \, dq = 0$ . Putting these two identities together yields  $FI(f) = - \int \tilde{Q} \, dv$  and therefore (3.1). The formula for  $FI(\rho)$  with  $\rho = \frac{d\mu}{dx}$  the stationary density for  $(x_t)$ , follows from an identical argument, omitted for brevity, once one observes that  $\rho$  solves the Kolmogorov equation

$$X_0^* \rho + \frac{1}{2} \sum_{j=1}^r (X_j^*)^2 \rho = 0.$$

Making the above argument rigorous will require us to justify the integration-by-parts steps, which along the way will make use of hypothesis (a) regarding integrability of  $f \log f$  with respect to  $mdq$ . Let  $\{\chi_i\}_{i=1}^\infty$  be

the partition of unity as in (a) and recall that  $\chi_{\leq N} := \sum_{i=1}^N \chi_i$ . Multiplying both sides of (2.6) by  $(\log f)\chi_{\leq N}$  and repeating the computation above gives

$$\frac{1}{2} \sum_k \int_{\mathbb{S}M} \frac{|\tilde{X}_k^* f|^2}{f} \chi_{\leq N} \, dq = - \int_{\mathbb{S}M} \tilde{Q} f \chi_{\leq N} \, dq + \int_{\mathbb{S}M} (\tilde{\mathcal{L}} \chi_{\leq N})(f \log f - f) \, dq.$$

As  $N \rightarrow \infty$ , the LHS converges to  $FI(f)$  by the monotone convergence theorem. The first RHS term converges to  $-\int_{\mathbb{S}M} \tilde{Q} \, dv$  by the dominated convergence theorem, using that  $\tilde{Q} \in L^1(v)$  (hypothesis (b)). Finally, the second RHS term above converges to 0 using the dominated convergence theorem and the facts that  $|\tilde{\mathcal{L}} \chi_{\leq N}| \leq m$  and  $(f \log f - f)m \in L^1(dq)$  (hypothesis (a)).

It remains to check the conditional version (3.2). Here we give only the formal proof, which can be justified under hypotheses (a) and (b) using an argument parallel to that presented above. For this we observe that

$$\tilde{X}_k^*(h\rho) = (V_{\nabla X_k}^* h - \tilde{X}_k h)\rho + (X_k^* \rho)h$$

and therefore since  $\int_{\mathbb{S}_x M} h_x(v) \, dv = 1$  (writing  $dv$  for Lebesgue measure on  $\mathbb{S}_x M$ ) we find

$$\begin{aligned} FI(f) - FI(\rho) &= \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} \frac{|(V_{\nabla X_k}^* h - \tilde{X}_k h)\rho + (X_k^* \rho)h|^2}{h\rho} \, dq \\ &\quad - \frac{1}{2} \sum_{k=1}^r \int_{\mathbb{S}M} \frac{|X_k^* \rho|^2}{\rho} h \, dq \\ &= \frac{1}{2} \sum_{k=1}^r \int_M \left( \int_{\mathbb{S}_x M} \frac{|(X_k - V_{\nabla X_k(x)}^*)h_x(v)|^2}{h_x(v)} \, dv \right) d\mu(x) \\ &\quad + \int_{\mathbb{S}M} (V_{\nabla X_k}^* h - X_k h) (X_k^* \rho) \, dq. \end{aligned}$$

Meanwhile, for the second term on the RHS, we have

$$\int_{\mathbb{S}M} (V_{\nabla X_k}^* h - X_k h) (X_k^* \rho) \, dq = - \int_{\mathbb{S}M} X_k h (X_k^* \rho) \, dq = - \int_{\mathbb{S}M} h (X_k^*)^2 \rho \, dq = 0,$$

completing the proof of Proposition 3.2.

### 3.2 Relation to Baxendale's relative entropy formula

This section is purely for exposition purposes to highlight the connection between the Fisher information identity in Proposition 3.2 and the relative entropy formula (1.10). Readers who are not interested may skip to Sect. 4.

In this section we give an argument of the Fisher information identity from the relative entropy formula (1.10) measuring the *degree to which the degeneracy* (1.13) fails to hold. We already have given a complete proof above, and so our purpose here is provide some additional intuition behind the meaning of Proposition 3.2. For this reason, we will not provide results in their full generality, and for technical simplicity we only consider the case in which  $M$  is compact and boundaryless (recall that Proposition 3.2 holds for noncompact  $M$ ).

In preparation, recall that given two measures  $\lambda, \eta$  on  $\mathbb{S}M$ ,  $\eta \ll \lambda$ , we write  $H(\eta|\lambda)$  for the relative entropy of  $\eta$  given  $\lambda$ , defined as in (1.11) above. In what follows, we abuse notation somewhat and write  $H(f|g)$  for the relative entropy of  $f \mathrm{d}q$  given  $g \mathrm{d}q$  when  $f, g$  are densities on  $\mathbb{S}M$ . Recall that  $H(f|g) = 0$  iff  $f \equiv g$  almost everywhere.

In what follows, we let  $\hat{\Phi}_\omega^t$  be the stochastic flow of diffeomorphisms on  $\mathbb{S}M$  induced by the SDE governing the projective process  $(w_t)$ . Given a smooth density  $f \in L^1(\mathbb{S}M)$ , let

$$f_t := (\hat{\Phi}_\omega^t)_* f = f \circ (\hat{\Phi}_\omega^t)^{-1} |\det D(\hat{\Phi}_\omega^t)^{-1}|$$

be the pushforward of  $f$  as a density on  $\mathbb{S}M$ , and analogously,  $\rho_t := (\Phi_\omega^t)_* \rho$ . In [12], Baxendale derived the following continuous-time version of a formula derived by Furstenberg [35] in the context of IID compositions of random matrices.

**Theorem 3.6** (Baxendale [12]) *Under Assumptions 1 & 2, one has the following:*

$$\begin{aligned} \mathbf{E} H(\rho_t|\rho) &= -t\lambda_\Sigma, \\ \mathbf{E} (H(f_t|f) - H(\rho_t|\rho)) &= t(n\lambda_1 - \lambda_\Sigma). \end{aligned} \tag{3.3}$$

It is straightforward to show that the second line of (3.3) is an equivalent formulation of (1.10), with “ $\leq$ ” replaced by “ $=$ ”, and directly encodes the Furstenberg criterion (1.13); that is,  $n\lambda_1 - \lambda_\Sigma = 0$  only if  $f_t \equiv f$  for all  $t$  and with full probability.

We will provide here an independent argument showing the following:

**Proposition 3.7** *Assume  $M$  is compact and boundaryless. Then,*

$$FI(\rho) = \frac{1}{t} \mathbf{E} H(\rho_t|\rho), \quad \text{and} \quad FI(f) = \frac{1}{t} \mathbf{E} H(f_t|f). \tag{3.4}$$

Proposition 3.2 for compact  $M$  immediately follows. The identity (3.4) confirms that the Fisher information identity for Lyapunov exponents is the time-infinitesimal analogue of (3.3).

*Proof sketch for Proposition 3.7* We present here the proof for  $f_t$ ; the proof for  $\rho_t$  follows the same lines and is omitted.

The density  $f_t$  can be readily seen to solve the stochastic continuity equation

$$df_t = \tilde{\mathcal{L}}^* f_t dt + \sum_{k=1}^r \tilde{X}_k^* f_t dW_t^k,$$

in Itô form, while  $\bar{f}_t = \mathbf{E}f_t$  solves the forward Kolmogorov equation

$$\partial_t \bar{f}_t = \tilde{\mathcal{L}}^* \bar{f}_t.$$

Therefore, since the initial data  $\bar{f}_0 = f$  is a stationary density satisfying  $\tilde{\mathcal{L}}^* f = 0$ , we have  $\mathbf{E}f_t = f$  for all  $t \geq 0$ . Using the formula for  $f_t$ , we can apply Itô's lemma to obtain the following stochastic equation that holds pointwise on  $\mathbb{S}M$ :

$$f_t \log(f_t/f) = \frac{1}{2} \sum_k \int_0^t \frac{|X_k^* f_s|^2}{f_s} ds + \int_0^t (\tilde{\mathcal{L}}^* f_s) (\log(f_s/f) + 1) ds + M_t;$$

here, the first RHS term is the Itô correction and  $M_t$  is a mean-zero martingale whose exact form is not important. Integrating over  $\mathbb{S}M$  and averaging with respect to  $\mathbf{E}$  gives

$$\frac{1}{t} \mathbf{E} H(f_t|f) = \frac{1}{t} \mathbf{E} \int_0^t FI(f_s) ds + \frac{1}{t} \mathbf{E} \int_0^t \int_{\mathbb{S}M} f_s \tilde{\mathcal{L}} \log(f_s/f) dq ds, \quad (3.5)$$

having used duality to bring  $\tilde{\mathcal{L}}$  onto the log. A standard calculation (using the fact that  $\tilde{\mathcal{L}}^* f = 0$ ) implies the second term on the right-hand side is just the relative entropy dissipation  $FI(f_s|f)$  (or relative Fisher information)

$$\int_{\mathbb{S}M} f_s \tilde{\mathcal{L}} \log(f_s/f) dq = -FI(f_s|f) := -\frac{1}{2} \sum_k \int_{\mathbb{S}M} \frac{|X_k(f_s/f)|^2}{f_s/f} f dq.$$

Moreover by a straightforward manipulation involving basic calculus and integration by parts,  $FI(f_s|f)$  can be expressed as

$$FI(f_s|f) = FI(f_s) - \frac{1}{2} \int_{\mathbb{S}M} \frac{|X_k^* f|^2}{f^2} f_s dq - \int_{\mathbb{S}M} X_k^2 \left( \frac{f_s}{f} \right) f dq.$$

Using Fubini and that  $\mathbf{E} f_s = f$  (equivalent to the fact that  $f$  is stationary), hence  $\mathbf{E} X_k^2(f_s/f) = 0$ , gives

$$\mathbf{E} FI(f_s|f) = \mathbf{E} FI(f_s) - FI(f).$$

Substituting this into (3.5) gives the result.  $\square$

## 4 Uniform hypoelliptic regularity via the Fisher information

We now turn our attention to the proof of Theorem B, which estimates the Fisher information quantity  $FI(f)$  from below in terms of fractional Sobolev regularity.

### 4.1 Full statement of Theorem B

As discussed in Sect. 1, our results are stated for parametrized families of SDE of the form

$$dx_t^\epsilon = X_0^\epsilon(x_t^\epsilon) dt + \sqrt{\epsilon} \sum_{k=1}^r X_k^\epsilon(x_t^\epsilon) \circ dW_t^k,$$

where for each  $\epsilon \in (0, 1]$ , the vector fields  $\{X_k^\epsilon\}_{k=0}^r$  on  $M$  are smooth. The following is a “uniform-in- $\epsilon$ ” version of the parabolic Hörmander condition (Definition 2.1).

**Definition 4.1** (*Uniform parabolic Hörmander*) Let  $\mathcal{M}$  be a manifold, and let  $\{Z_0^\epsilon, Z_1^\epsilon, \dots, Z_r^\epsilon\} \subset \mathfrak{X}(\mathcal{M})$  be a set of vector fields parameterized by  $\epsilon \in (0, 1]$ . With  $\mathcal{X}_k$  defined as in Definition 2.1 we say  $\{Z_0^\epsilon, Z_1^\epsilon, \dots, Z_r^\epsilon\}$  satisfies the uniform parabolic Hörmander condition on  $\mathcal{M}$  if  $\exists k \in \mathbb{N}$ , such that for any open, bounded set  $U \subseteq \mathcal{M}$  there exist constants  $\{K_n\}_{n=0}^\infty$ , such that for all  $\epsilon \in (0, 1]$  and all  $x \in U$ , there is a finite subset  $V(x) \subset \mathcal{X}_k$  such that  $\forall \xi \in T_x \mathcal{M}$ ,

$$|\xi| \leq K_0 \sum_{Z \in V(x)} |Z(x) \cdot \xi| \quad \sum_{Z \in V(x)} \|Z\|_{C^n} \leq K_n.$$

Equivalently, in the notation of Sect. 2.2.1, we could stipulate that the subset  $V(x)$  be drawn from  $\text{Lie}^k(Z_0^\epsilon; Z_1^\epsilon, \dots, Z_r^\epsilon)$  for some fixed  $k \geq 1$ .

In what follows, we will assume that the SDE defining the projective process  $w_t = (x_t, v_t)$  satisfies this uniform spanning condition:

**Assumption 3** (*Projective spanning condition*) The vector fields  $\{\tilde{X}_0^\epsilon, \tilde{X}_1^\epsilon, \dots, \tilde{X}_r^\epsilon\}$  satisfy the parabolic Hörmander condition on  $\mathbb{S}M$  uniformly in  $\epsilon \in (0, 1]$  on bounded sets in the sense of Definition 4.1.

We are now positioned to state the general version of Theorem B relating Fisher information and weak Sobolev regularity. Below,  $B_r(w_0; \mathbb{S}M)$  is the geodesic ball of radius  $r > 0$  centered at  $w_0 \in \mathbb{S}M$ .

**Theorem 4.2** *Let  $(M, g)$  be as in the beginning of Sect. 2, and let  $\{X_0^\epsilon, \dots, X_r^\epsilon\}$  be smooth vector fields. Assume that for all fixed  $\epsilon \in (0, 1]$ , these vector fields satisfy Assumptions 1, 2, and that the parametrized family  $\{X_k^\epsilon\}_{\epsilon \in (0, 1]}$  satisfies Assumption 3. Let  $f^\epsilon$  denote the unique stationary density for the projective process  $w_t = (x_t, v_t)$  on  $\mathbb{S}M$ , and assume that for all  $\epsilon \in (0, 1]$  we have  $FI(f^\epsilon) < \infty$ . Then  $\exists s_* \in (0, 1)$  such that  $\forall s \in (0, s_*], \forall R > 0, \forall w_0 \in \mathbb{S}M, \exists C > 0$  such that  $\forall \epsilon \in (0, 1]$ ,*

$$\|\chi_R f^\epsilon\|_{W^{s,1}}^2 \leq C (1 + FI(f^\epsilon)),$$

where  $\chi_R$  is a smooth cutoff function equal to 1 on  $B_R(w_0; \mathbb{S}M)$  and zero on  $\mathbb{S}M \setminus B_{R+1}(w_0; \mathbb{S}M)$ . Moreover, the constant  $C$  can be chosen to depend only on  $s, R, w_0, \dim M$  and the constants  $k$  and  $\{K_n\}_{n=0}^J$  (for a  $J$  depending only on  $k$ ) in Definition 4.1.

The definition of fractional Sobolev space  $W^{s,1}$  is recalled for functions on manifolds in (A.1).

**Remark 4.3** Theorem 4.2 requires only a uniform Hörmander condition concerning spanning of the Lie algebra  $\text{Lie}(X_0^\epsilon, \dots, X_r^\epsilon)$  rather than the slightly stronger parabolic Hörmander condition concerning the zero-time ideal  $\text{Lie}(X_0^\epsilon; X_1^\epsilon, \dots, X_r^\epsilon)$ . However, the parabolic Hörmander condition is required anyway for other purposes such as verifying Assumption 2, c.f. the discussion in Sect. 2.

**Remark 4.4** Above, the value  $s_*$  is determined exclusively by the number of ‘generations’ of brackets needed to satisfy Assumption 3 (though note this will generally depend on the dimension of the manifold  $M$  itself). One can check that the proof gives  $\frac{1}{2k} < s_* \leq \frac{1}{k}$  where  $k$  is as in Definition 4.1, being the maximum depth of brackets needed to uniformly span.

**Remark 4.5** Lastly, the hypoelliptic lower bound Theorem B can potentially be improved. In the proof of the theorem below, the duality semi-norm  $\mathfrak{D}(f)$  is of size  $O(\epsilon \sqrt{FI})$ . In analogy with other kinds of interpolation inequalities, it is reasonable to seek a version of the Hörmander inequality (4.7) that gives extra weight to  $\mathfrak{D}$  relative to  $FI$ , i.e. something along the lines of  $\|\chi_U f\|_{W^{s,1}}^2 \lesssim 1 + \epsilon^{-\alpha} \mathfrak{D}(f)^2 + \epsilon^\beta FI(f) \lesssim 1 + \frac{n\lambda_1 - 2\lambda_\Sigma}{\epsilon^{1-\gamma}}$ , thus improving the scaling available

to the compactness-rigidity scheme. Perhaps in this scheme one can use a weaker regularity on the left-hand-side but nevertheless is good enough to prove equi-integrability (all that is really required for the compactness-rigidity approach).

The remainder of this section is devoted to proving Theorem 4.2.

#### 4.1.1 Notation and setup

Theorem 4.2 is not directly related to the projective process, and so for the sake of conceptual clarity we present a proof here establishing a more general version. Throughout the remainder of Sect. 4, we will assume  $(\mathcal{M}, \tilde{g})$  is a smooth, connected, geodesically complete, orientable manifold without boundary. We write  $d = \dim \mathcal{M}$  and  $dq$  for the volume element on  $\mathcal{M}$ . The geodesic ball of radius  $r > 0$  at a point  $x \in \mathcal{M}$  is written  $B_r(x; \mathcal{M})$ , while we write  $B_r(0) \subset \mathbb{R}^d$  for the Euclidean ball of radius  $r$ .

Abusing notation somewhat, we will assume  $\{X_k^\epsilon\}_{k=0}^r$ ,  $\epsilon \in (0, 1]$ , is a family of smooth vector fields on  $\mathcal{M}$  satisfying the uniform parabolic Hörmander condition (Definition 4.1), while  $f^\epsilon$  is any<sup>6</sup> smooth solution of the Kolmogorov equation

$$(X_0^\epsilon)^* f^\epsilon + \frac{\epsilon}{2} \sum_{j=1}^r ((X_j^\epsilon)^*)^2 f^\epsilon = 0. \quad (4.1)$$

Recall that given  $X, Y \in \mathfrak{X}(\mathcal{M})$ , the adjoint action of  $X$  on  $Y$  is defined through the Lie bracket  $\text{ad}(X)Y = [X, Y]$ . For a multi-index  $I = (i_1, \dots, i_k)$ ,  $i_j \in \{0, \dots, r\}$  for each  $1 \leq j \leq k$ , we denote

$$X_I = \text{ad}(X_{i_1}) \dots \text{ad}(X_{i_{k-1}}) X_{i_k}.$$

In what follows, set  $s_0 = \frac{1}{2}$ ,  $s_j = 1$ ,  $1 \leq j \leq r$  and for a multi-index  $I = (i_1, \dots, i_k)$  we write,

$$m(I) := \frac{1}{s(I)} := \sum_{j=1}^k \frac{1}{s_{i_j}}.$$

Note that  $m(I)$  provides a measure of how “deep” a bracket is (i.e. the larger  $m(I)$  the more brackets that were taken), weighted in a way that will be consistent with available regularity.

<sup>6</sup> Uniqueness of  $f^\epsilon$  is not used in the proof of Theorem 4.2 and so we need not assume any topological irreducibility property of the corresponding SDE.

We denote by  $\mathfrak{X}^s(\mathcal{M}) \subset \mathfrak{X}(\mathcal{M})$  the  $C^\infty(\mathcal{M})$ -submodule of vector fields generated from successive brackets with  $s \leq s(I)$ , that is,

$$\mathfrak{X}^s(\mathcal{M}) = \left\{ Z \in \mathfrak{X}(\mathcal{M}) : Z = \sum_j h_j X_{I_j}, \quad s(I_j) \geq s, \quad h_j \in C^\infty(\mathcal{M}) \right\}.$$

Recall that  $\{X_j\}_{j=0}^r = \{X_j^\epsilon\}_{j=0}^r \subset \mathfrak{X}(\mathcal{M})$  depend in a general manner on a parameter  $\epsilon \in (0, 1)$ , hence  $\mathfrak{X}^s(\mathcal{M})$  also depends on  $\epsilon$ . However, assuming that  $\{X_j^\epsilon\}$  satisfies the uniform parabolic Hörmander condition, a simple consequence is that  $\exists s > 0$  such that  $\forall \epsilon \in (0, 1)$ ,  $\mathfrak{X}^s(\mathcal{M}) = \mathfrak{X}(\mathcal{M})$ <sup>7</sup>. Once and for all, fix  $s_* > 0$  so that  $\mathfrak{X}^{s_*}(\mathcal{M}) = \mathfrak{X}(\mathcal{M})$ .

Theorem 4.2 estimates the Fisher information of  $f^\epsilon$  from below by fractional Sobolev regularity  $W^{s_*, 1}$  (defined on general manifolds in (A.1)). However, for our purposes, it is easier to work with an  $L^1$  Hölder-type regularity class  $\Lambda^s$  embedded in  $W^{s', 1}$  for all  $0 < s' < s < 1$  (Lemma A.2). Due to subtleties involved in intrinsic definitions of Besov spaces on manifolds (see e.g. [88]), we will construct the  $\Lambda^s$  seminorm on bounded sets in  $\mathcal{M}$  using a coordinate atlas, as follows.

Fix  $R > 0$ , set  $U := B_R(z_0; \mathcal{M})$ <sup>8</sup> at some arbitrary  $z_0 \in \mathcal{M}$ , and set  $U' = B_{R+1}(z_0; \mathcal{M})$ . Note that by the Hopf-Rinow theorem,  $\overline{U}$  and  $\overline{U'}$  is compact, hence there exists  $\delta > 0$  and a finite family  $\{\mathbf{x}_j\}$  smooth injective mappings  $\mathbf{x}_j : B_{4\delta}(0) \rightarrow U$  such that both  $\tilde{U}_j := \mathbf{x}_j(B_{4\delta}(0))$  and  $U_j := \mathbf{x}_j(B_\delta(0))$  are open covers of  $U$ . Let  $\{\chi_j\}$  be a smooth partition of unity on  $B_R(z_0; \mathcal{M})$  subordinate to the cover  $\{\tilde{U}_j\}$ , i.e., (i)  $0 \leq \chi_j \leq 1$  everywhere, (ii)  $\chi_j|_{U_j} \equiv 1$ , and (iii)  $\chi_j$  is supported in  $\tilde{U}_j$ . With this, we define the following semi-norm for  $w \in C_c^\infty(U)$ :

$$\|w\|_{\Lambda^s} = \|w\|_{L^1} + \sup_{h \in \mathbb{R}^d : |h| < \delta} \sum_j \int_{\mathbb{R}^d} \frac{|\tilde{w}_j(x+h) - \tilde{w}_j(x)|}{|h|^s} J_j(x) dx, \quad (4.2)$$

where  $\tilde{w}_j = (\chi_j w) \circ \mathbf{x}_j$  and  $J_j : B_\delta(0) \rightarrow \mathbb{R}_{>0}$  is the coordinate representation of the volume element in the chart  $(U_j, \mathbf{x}_j)$ . Note that  $J_j$  The choice of  $z_0, R$

<sup>7</sup> Indeed, if  $\{Z_1, \dots, Z_m\}$  span  $T_w \mathcal{M}$  then  $\exists \delta > 0$  such that  $\forall v \in B_\delta(w) = \{v \in \mathcal{M} : d(w, v) < \delta\}$  the same vector fields span, and so for  $V \in \mathfrak{X}(\mathcal{M})$ ,  $\exists c_j \in C^\infty$  such that  $V = \sum c_j Z_j$  on  $B_\delta$ . The result then follows by a suitable partition of unity.

<sup>8</sup> More general sets  $U$  can be used, but this is sufficient for all our purposes. In fact, at this step, any bounded, open set can be used.

and the coordinate charts  $\{\mathbf{x}_j\}$  is henceforth fixed and thus the dependence of  $\Lambda^s$  on these parameters<sup>9</sup> will be suppressed in what follows.

The embedding

$$\|w\|_{W^{s',1}} \lesssim_U \|w\|_{\Lambda^s}$$

for all  $0 < s' < s < 1$  and  $w \in C_c^\infty(U)$  is intuitively clear. However, for the sake of completeness we include a proof in the Appendix (Lemma A.2; see also [pg 301; [88]]).

#### 4.1.2 Outline of the proof of Theorem 4.2

Crucial to both Hörmander's original approach and our own is the ability to measure *partial* regularity of a function along some given set of directions. Given  $Y \in \mathfrak{X}(\mathcal{M})$ , let  $Y^*$  denote its formal adjoint<sup>10</sup> and let  $e^{tY^*}$  denote the linear propagator solving the partial differential equation  $\partial_t - Y^* = 0$  (this is well-defined as long as  $t > 0$  is taken sufficiently small depending on  $Y$  and  $U$ ). For  $s > 0$ , we define the family of 'partial'  $L^1$  Hölder seminorms<sup>11</sup>

$$|w|_{Y,s} = \sup_{|t| \leq \delta_0} |t|^{-s} \|e^{tY^*} w - w\|_{L^1}.$$

Note the dependence on the parameter  $\delta_0 > 0$ : in practice, given  $U$ , this parameter is fixed and depends only on the regularity of  $\{X_I\}$  as  $I$  ranges over the multi-indices with  $s(I) \geq s_*$ . We may henceforth fix  $\delta_0$ , as the vector fields in the proof vary in a uniformly bounded set in  $C^J$  for a  $J$  depending only the constants in Definition 4.1.

<sup>9</sup> While the value of  $\|w\|_{\Lambda^s}$  depends on the choice of coordinate chart, for any alternative choice of charts  $\{\mathbf{x}'_j\}$ , note that the following holds:  $\forall w \in C_c^\infty(U), \|w\|_{\Lambda^s(U, \{\mathbf{x}_j\})} \approx \|w\|_{\Lambda^s(U, \{\mathbf{x}'_j\})}$ .

<sup>10</sup> Note that  $Y^* f = -Yf - (\operatorname{div} Y)f$  and in particular  $Y^*$  is not a vector field, but a more general first order differential operator.

<sup>11</sup> Note that these seminorms are slightly different from those used in [41], where the linear propagator  $e^{tY}$  solving  $\partial_t - Y = 0$  is used directly. Note, though, that the regularity defined is essentially the same in the sense that

$$\|w\|_{L^1} + \sup_{|t| \leq \delta_0} |t|^{-s} \|e^{tY^*} w - w\|_{L^1} \approx_{\delta_0, U, H} \|w\|_{L^1} + \sup_{|t| \leq \delta_0} |t|^{-s} \|e^{tY} w - w\|_{L^1}.$$

Turning back to the proof of Theorem 4.2: ultimately, for  $f = f^\epsilon$  solving the Kolmogorov equation (4.1), we seek to control  $\|f^\epsilon\|_{\Lambda^{s_*}(U)}$  from above in terms of  $FI(f^\epsilon)$ . Starting from the latter, it is straightforward (Lemma 4.6) to obtain the general functional inequality

$$\sum_{j=1}^r |w|_{X_j,1} \lesssim \|w\|_{L^1}^{1/2} \sqrt{FI(w)}, \quad (4.3)$$

for any  $w \in C_c^\infty(U)$ . Hence, for all intents and purposes it suffices to control the regularity  $\|f\|_{\Lambda^{s_*}}$  from above in terms of  $\sum_{j \geq 1} |f|_{X_j,1}$ . For this, we turn to the ideas laid out by Hörmander. First, the spanning condition  $\mathfrak{X}^{s_*} = \mathfrak{X}$  allows to “fill in” the missing directions not spanned by the original  $\{X_0, \dots, X_r\}$ , leading to the general functional inequality:

$$\|w\|_{\Lambda^{s_*}} \lesssim_U \|w\|_{L^1} + \sum_{j=0}^r |w|_{X_j,s_j}, \quad (4.4)$$

for  $w \in C_c^\infty(U)$ . This is a relatively straightforward adaptation of [Sect. 4; [41]]—see Lemma 4.9 below.

While (4.3) controls  $|w|_{X_j,1}$ ,  $1 \leq j \leq r$  in terms of the Fisher information  $FI(w)$ , it remains (as in [41]) to obtain an upper estimate on  $|f^\epsilon|_{X_0,1/2}$ . The starting point is the derivation of an a priori estimate on  $f^\epsilon$  from (4.1). In [41], Hörmander observed that one naturally obtains an a priori regularity estimate on  $X_0 f$  in a *negative* regularity  $L^2$  space in terms of  $X_j f \in L^2$  (see also discussions in [4]). In our case, we cannot work in  $L^2$ , and instead have to work in a negative-type regularity which is essentially the dual to that in (4.3)—this is the only a priori estimate available that will be useful. Pairing (4.1) with a test function  $v \in C_c^\infty$  we obtain the following a priori estimate, which is essentially the  $W^{-1,\infty}$  norm with respect to the  $X_j^*$  directions:

$$\begin{aligned} \mathfrak{D}_\epsilon(f^\epsilon) &:= \sup \left\{ \left| \int f^\epsilon X_0^\epsilon v \, dq \right| : v \in C_c^\infty, \|v\|_{L^\infty} + \sum_{j=1}^r \|X_j v\|_{L^\infty} \leq 1 \right\} \\ &\leq \epsilon \sum_{j=1}^r \|X_j^* f^\epsilon\|_{L^1} \lesssim \epsilon \sqrt{FI}. \end{aligned} \quad (4.5)$$

Using this, the missing  $X_0$  regularity is recovered by the following, which is the main difficulty in the proof: for any  $0 < \sigma < s_*$ ,  $U$  bounded geodesic

ball and  $w \in C_c^\infty(U)$ , we show that

$$|w|_{X_{0,1/2}} \lesssim_U \sum_{j=1}^r |w|_{X_j,1} + \mathfrak{D}_\epsilon(w) + \|w\|_{\Lambda^\sigma}. \quad (4.6)$$

That is, we recover the  $|w|_{X_{0,1/2}}$  regularity by a combination of the negative  $\mathfrak{D}_\epsilon$  regularity in conjunction with the positive  $|w|_{X_j,1}$  regularity, accruing only a remainder term  $\|w\|_{\Lambda^\sigma}$ . Combining with (4.4) (along with interpolation of  $\Lambda^\sigma$  between  $\Lambda^{s_*}$  and  $L^1$ ), we obtain the following:  $\forall U \subset \mathcal{M}$  open, bounded geodesic balls  $\exists C > 0$  such that for all  $w \in C_c^\infty(U)$ , there holds

$$\|w\|_{\Lambda^{s_*}(U)} \leq C \left( \|w\|_{L^1} + \mathfrak{D}_\epsilon(w) + \sum_{j=1}^r |w|_{X_j,1} \right). \quad (4.7)$$

From here, our estimate on  $\|f^\epsilon\|_{\Lambda^{s_*}(U)}$  in Theorem 4.2 is an easy consequence of the functional inequality (4.3) and the apriori estimate in (4.5).

In Sect. 4.2 we review the available apriori estimates and basic functional inequalities that are used in the proof. In Sect. 4.3 we briefly recall (4.4) and a closely related inequality which are straightforward adaptations of estimates in [Sect. 4; [41]]. In Sect. 4.4 we give the proof of (4.6), leaving the main lemma to be proved in Sect. 4.5. As in the corresponding step in [41], (4.6) is based on a careful regularization procedure, though it is more subtle to perform this procedure in the  $W^{-1,\infty}$ -type framework we work with here. Section 4.5 is dedicated to the details of this regularization.

## 4.2 Preliminary estimates

To start, we record some useful estimates for the  $L^1$  Hölder-type seminorms  $|\cdot|_{Y,s}$ . Let  $Y \in \mathfrak{X}(M)$  and let  $e^{tY}$  be the linear propagator of the partial differential operator  $\partial_t - Y$ . By the method of characteristics, the smooth family of diffeomorphisms  $h_Y(t) : \mathcal{M} \rightarrow \mathcal{M}$  solving the initial value problem  $\dot{x} = Y(x)$  satisfies the identity

$$e^{tY} w = w \circ h_Y(t).$$

With  $Y^*$  the formal adjoint of  $Y$ , again by the method of characteristics there is a smooth family of strictly positive densities  $H_Y(t) : \mathcal{M} \rightarrow (0, \infty)$  such that

$$e^{tY^*} w = H_Y(t) e^{-tY} w. \quad (4.8)$$

In particular, for  $|t| \lesssim 1$ ,

$$|H_Y(t) - 1| \lesssim |t| , \quad (4.9)$$

with similar estimates on higher derivatives. The choice of the parameter  $\delta_0$  above will be such that  $h_Y(t, x)$  is well-defined (i.e. no trajectories reach infinity in finite time) for all  $t \in (-\delta_0, \delta_0)$  and  $x \in U$  for every vector field that appears in the proof which follows (more accurately, we will re-scale the fields so that we may take  $\delta_0 = 1$ ). Accordingly, we have for  $t \in (-\delta_0, \delta_0)$  and  $w \in C_c^\infty(U)$  we have,

$$\|e^{tY} w\|_{L^p} + \|e^{tY^*} w\|_{L^p} \lesssim \|w\|_{L^p} . \quad (4.10)$$

Next, we prove (4.3): that  $\|X_j^* w\|_{L^1}$  controls one derivative in the  $L^1$ -Hölder norms.

**Lemma 4.6** *Let  $U \subset \mathcal{M}$  be a bounded geodesic ball. Then,  $\forall w \in C_c^\infty(U)$  there holds,*

$$\|w\|_{X_j, 1} \lesssim_U \|X_j^* w\|_{L^1} \lesssim \|w\|_{L^1}^{1/2} \sqrt{FI(w)} .$$

*Proof* Let  $v \in L^\infty$ , then

$$\left| \int_{\mathcal{M}} v(e^{tX_j^*} w - w) dq \right| = \left| \int_0^t \int_{\mathcal{M}} v e^{sX_j^*} X_j^* w dq ds \right| \leq |t| \|v\|_{L^\infty} \|X_j^* w\|_{L^1} .$$

Taking the supremum over  $\|v\|_{L^\infty} \leq 1$  and dividing by  $|t|$  gives the first inequality whereas the second follows by Cauchy-Schwarz.  $\square$

Lastly, we record the simple observation that the negative regularity  $\mathfrak{D}_\epsilon$  can be localized.

**Lemma 4.7** *Let  $U \subseteq \mathcal{M}$  be a bounded geodesic ball and  $\chi \in C_c^\infty(U)$ . Then, for any  $h \in L^1(\mathcal{M})$ , we have*

$$\mathfrak{D}_\epsilon(\chi h) \lesssim_U \|h\|_{L^1} + \mathfrak{D}_\epsilon(h) .$$

*Proof* Set  $w = \chi h$ . For test functions  $v \in C_c^\infty(U)$ , we estimate

$$\left| \int (X_0 v) w dq \right| \leq \left| \int v (X_0 \chi) h dq \right| + \left| \int X_0(\chi v) h dq \right| \lesssim \|h\|_{L^1} + \mathfrak{D}_\epsilon(h) .$$

Note that the estimate is uniform in  $\|X_0\|_{C^1(U)}$ .  $\square$

### 4.3 Controlling $\Lambda^s$ with $|\cdot|_{X_j, s_j}$

The first steps to Theorem 4.2 are several lemmas that are nearly the same as those in [Sect. 4; [41]], except (A) we need them in  $L^1$ , (B) we need them uniform in the parameter  $\epsilon$  hidden in  $X_0$ , (C) we need to generalize the proof to compact subsets of manifolds  $(\mathcal{M}, g)$ . However, these changes are seen to be relatively small a careful reading of [41] and so we only sketch the changes below after the statements; see [18] for more discussion on the uniformity.

**Lemma 4.8** *Let  $U$  be a bounded geodesic ball and  $0 < \sigma < s_*$  with  $\mathfrak{X}^{s_*}(\mathcal{M}) = \mathfrak{X}(\mathcal{M})$ . For all  $\delta > 0$ ,  $\exists C_\delta > 0$  such that for all multi-indices  $I$  such that at least one index is zero, the following holds  $\forall w \in C_c^\infty(U)$ ,*

$$|w|_{X_I, s(I)} \leq \delta |w|_{X_0, \frac{1}{2}} + C_\delta \left( \sum_{j=1}^r |w|_{X_j, 1} + \|w\|_{\Lambda^\sigma} \right).$$

Moreover,  $C_\delta$  depends on  $\{X_0, X_1, \dots, X_r\}$  only in the manner stated in Theorem 4.2.

The next lemma shows that one can control regularity in  $\Lambda^s$  by controlling the original vector fields.

**Lemma 4.9** *Let  $U$  be an open, bounded geodesic ball and  $s_*$  be such that  $\mathfrak{X}^{s_*}(\mathcal{M}) = \mathfrak{X}(\mathcal{M})$ . Then, for  $w \in C_c^\infty(U)$  there holds*

$$\|w\|_{\Lambda^{s_*}} \lesssim_U \|w\|_{L^1} + \sum_{j=0}^r |w|_{X_j, s_j}.$$

Moreover, the implicit constant depends on  $\{X_0, X_1, \dots, X_r\}$  only in the manner stated in Theorem 4.2.

*Summary of Proofs of Lemma 4.8 and Lemma 4.9* First we observe that the analogues of [Lemma 4.1; [41]] and [Lemma 4.2; [41]] can both be extended to  $\Lambda^s(U)$  for general locally bounded geometries (see the proof of Lemma 4.16 below for more details on how this would be done). With this update, the proof of Lemma 4.8 follows as in [Lemma 4.6; [41]] (note the remark that follows the proof) whereas the proof of Lemma 4.9 follows as in [Theorem 4.3; [41]]; indeed the only steps specific to  $\mathbb{R}^d$  are [Lemma 4.1; [41]] and [Lemma 4.2; [41]] and all steps apply equally well for any  $L^p$ ,  $p \in [1, \infty)$ . The uniformity is also seen noting how Hörmander's condition is applied and by noting that the errors depend only on a fixed, finite number of local  $C^k$  norms of the vector fields (see [18] for more discussion).  $\square$

#### 4.4 Positive $X_0$ regularity from negative $X_0$ and positive $X_j$ regularity

In this subsection, we prove the apriori estimate (4.6) and then use it to complete the proof of Theorem 4.2. Fix  $0 < \sigma < s_*$  arbitrary. Having fixed  $U$  we may, by rescaling  $\{X_0, X_1, \dots, X_r\}$ , assume that  $e^{tX_I}$  (and hence  $e^{tX_I^*}$ ) is well-posed for  $w \in C_c^\infty(U)$  for  $t \in [-1, 1]$  for  $\sigma \leq s(I)$  (and hence we may choose  $\delta_0 = 1$ ).

Analogous to [Sect. 5; [41]], the primary intermediate step is to first deduce the estimate assuming the natural control on essentially all other vector fields in  $\mathfrak{X}^\sigma$ .

**Definition 4.10** Denote by  $\mathcal{J}$  the set of all multi-indices  $I$  with  $\sigma \leq s(I)$  except for the singleton  $\{0\}$ .

Note this definition is slightly different from that in [41]. Define the following semi-norm

$$|w|_M := \sum_{I \in \mathcal{J}} |w|_{X_I, s(I)}.$$

The main step in the proof of (4.6) (and hence Theorem 4.2 as a whole) is to prove the following.

**Lemma 4.11** *For any bounded, geodesic ball  $U \subset \mathcal{M}$ , and every  $w \in C_c^\infty(U)$ , the following holds uniformly in  $\epsilon$ ,*

$$|w|_{X_0, \frac{1}{2}} \lesssim_U |w|_M + \|w\|_{\Lambda^\sigma} + \mathfrak{D}_\epsilon(w).$$

As in the corresponding [Sect. 5; [41]] (and in [4]), we use an approach based on a carefully selected regularization, but our choice is even a little more delicate than [41]. As the regularization procedure is quite technically subtle, we first give the proof of Lemma 4.11 assuming the existence of a regularizer satisfying the desired properties.

**Lemma 4.12** *There exists a family of uniformly bounded smoothing operators  $S_\tau : L^p \rightarrow L^p$  for  $\tau \in (0, 1)$  and  $p \in [1, \infty]$  with the following properties: for all  $w \in C_c^\infty(U)$ ,*

$$\begin{aligned} \|S_\tau^* w - w\|_{L^1} &\lesssim \tau |w|_M, \\ \sum_{j=1}^r \|X_j S_\tau w\|_{L^\infty} &\lesssim \frac{1}{\tau} \|w\|_{L^\infty}, \\ \|[X_0, S_\tau]^* w\|_{L^1} &\lesssim \frac{1}{\tau} (|w|_M + \|w\|_{\Lambda^\sigma}). \end{aligned}$$

Assuming this lemma for now, we proceed.

*Proof of Lemma 4.11 assuming Lemma 4.12* We will first obtain regularity estimates by evaluating the fractional time derivative of  $e^{tX_0^*}w$  (omitting the  $\epsilon$  for notational simplicity). Observe that for any  $t, \tau > 0$ ,

$$\|e^{tX_0^*}w - w\|_{L^1} \leq \|e^{tX_0^*}(S_\tau^*w - w)\|_{L^1} + \|S_\tau^*w - w\|_{L^1} + \|e^{tX_0^*}S_\tau^*w - S_\tau^*w\|_{L^1}. \quad (4.11)$$

Therefore, by Lemma 4.12 and  $L^1$  boundedness of the group  $e^{-tX_0^*}$  on  $U$ ,

$$\sup_{|t| \leq 1} \|e^{tX_0^*}(S_\tau^*w - w)\|_{L^1} \lesssim \|S_\tau^*w - w\|_{L^1} \lesssim \tau |w|_M. \quad (4.12)$$

This will suffice for the first two terms in (4.11). Next, we estimate the last term in (4.11). We will do this using the inequality,

$$\|e^{tX_0^*}S_\tau^*w - S_\tau^*w\|_{L^1} \leq \sup_{\|v\|_{L^\infty} \leq 1} \left| \int_0^t \int_{\mathcal{M}} (e^{sX_0^*}v) X_0^*S_\tau^*w \, dq \, ds \right|, \quad (4.13)$$

which can be easily deduced from  $\|w\|_{L^1} = \sup_{\|v\|_{L^\infty} \leq 1} \int v w \, dq$  and the property  $e^{tX_0^*} = (e^{tX_0})^*$ . For a fixed  $v \in L^\infty$ , we find that

$$\begin{aligned} \left| \int_{\mathcal{M}} (e^{sX_0}v) X_0^*S_\tau^*w \, dq \right| &\leq \left| \int_{\mathcal{M}} (e^{sX_0}v)[X_0, S_\tau]^*w \, dq \right| + \left| \int_{\mathcal{M}} (S_\tau e^{sX_0}v) X_0^*w \, dq \right| \\ &\leq \|e^{sX_0}v\|_{L^\infty} \| [X_0, S_\tau]^*w \|_{L^1} + \left( \|S_\tau e^{sX_0}v\|_\infty + \sum_{j=1}^r \|X_j S_\tau e^{sX_0}v\|_{L^\infty} \right) \mathfrak{D}(w). \end{aligned}$$

Using Lemma 4.12 and the boundedness of  $e^{tX_0}$  in  $L^\infty(U)$ , we conclude that

$$\left| \int_{\mathcal{M}} (e^{sX_0}v) X_0^*S_\tau^*w \, dq \right| \lesssim_U \frac{1}{\tau} \|v\|_{L^\infty} (|w|_M + \|w\|_{\Lambda^\sigma} + \mathfrak{D}(w))$$

and from (4.13) we deduce

$$\|e^{tX_0^*}S_\tau^*w - S_\tau^*w\|_{L^1} \lesssim_U \frac{|t|}{\tau} (|w|_M + \|w\|_{\Lambda^\sigma} + \mathfrak{D}(w)).$$

Therefore, setting  $\tau = \sqrt{|t|}$  and using (4.12) implies

$$\|e^{tX_0^*}w - w\|_{L^1} \lesssim \sqrt{|t|} (|w|_M + \mathfrak{D}(w)).$$

By (4.8), (4.9), and the boundedness of  $U$ , this implies the desired result.  $\square$

To complete the section, we explain in more detail how Lemma 4.11 implies Theorem 4.2.

*Proof of Theorem 4.2* By Lemma 4.11, followed by Lemma 4.8 to absorb the effect of the higher order brackets by choosing  $\delta$  sufficiently small, implies (4.6), that is for any  $w \in C_c^\infty(U)$ ,

$$\|w\|_{X_0, \frac{1}{2}} \lesssim \|w\|_{\Lambda^\sigma} + \sum_{j=1}^r |w|_{X_j, 1} + \mathfrak{D}_\epsilon(w).$$

Applying Lemma 4.9 then implies

$$\|w\|_{\Lambda^{s_*}} \lesssim \sum_{j=1}^r |w|_{X_j, 1} + \|w\|_{\Lambda^\sigma} + \mathfrak{D}_\epsilon(w). \quad (4.14)$$

Next, note the interpolation (from Hölder's inequality and Definition 4.2):  $\forall \sigma \in (0, s^*)$  and all  $\delta > 0$ ,  $\exists C_\delta$  such that

$$\|w\|_{\Lambda^\sigma} \leq \delta \|w\|_{\Lambda^{s_*}} + C_\delta \|w\|_{L^1},$$

which by (4.14) implies Hörmander inequality (4.7). Let  $U \subset\subset U' \subset \mathcal{M}$  where  $U'$  is another bounded (open) geodesic ball and let  $\chi \in C_c^\infty(U')$  with  $\chi(x) = 1$  for all  $x \in U$ . Then, Lemma 4.7 implies

$$\|\chi f^\epsilon\|_{\Lambda^{s_*}} \lesssim 1 + \sum_{j=1}^r |\chi f^\epsilon|_{X_j, 1} + \mathfrak{D}_\epsilon(f^\epsilon). \quad (4.15)$$

Putting Lemma 4.6 together with (4.15) and (4.5), completes the proof of Theorem 4.2.  $\square$

## 4.5 Regularization: Lemma 4.12

In this subsection we prove Lemma 4.12. First, we define a suitable “isotropic” mollifier via the parameterization. Let  $\phi \in C_0^\infty((-1, 1))$  with  $\phi \geq 0$ ,  $\int_{-1}^1 \phi(t) dt = 1$ , and  $\phi(-t) = \phi(t)$ , denoting  $\tilde{w}_j = \chi_j w \circ \mathbf{x}_j$ , and for each  $x \in \mathbb{R}^d$  let  $\phi_\tau(x) = \frac{1}{\tau^d} \phi(|x|/\tau)$ . We define the regularization of  $\chi_j w$  as follows for  $0 < \tau \leq \delta$ ,

$$\Phi_\tau^j w \circ \mathbf{x}_j := \frac{1}{J(x)} \int_{\mathbb{R}^d} \phi_\tau(|x - y|) \tilde{w}_j(y) J_j(y) dy,$$

where as above we write  $J_j = (\det \tilde{g})^{1/2}$ , the volume element on  $\mathcal{M}$  in local coordinates. We write

$$\Phi_\tau w(q) = \sum_j \Phi_\tau^j w(q),$$

and note that since  $w$  is compactly supported, the above summation is finite. Note that  $\Phi_\tau$  is not  $L^2(dq)$  self adjoint with the adjoint given by  $(\Phi_\tau)^* = \sum_j (\Phi_\tau^j)^*$

$$(\Phi_\tau^j)^* w \circ \mathbf{x}_j = \int_{\mathbb{R}^d} \phi_\tau(|x - y|) \tilde{w}_j(y) dy.$$

The basic properties of these kinds of mollifiers are classical, however, we include sketches for completeness. Due to the compatibility between definitions (4.2) and (4.16), and the fact that the properties we are interested in are purely local, the results follow from the corresponding statements on  $\mathbb{R}^d$ . We sketch the details of this in the first lemma, which is the most delicate of the estimates we require.

**Lemma 4.13** *For all  $\sigma \in [0, 1)$ ,  $U \subset \mathcal{M}$  bounded, geodesic balls, there holds the following uniformly in  $\tau \in (0, \delta)$  and uniformly in  $C^3$  bounded sets of  $Y \in \mathfrak{X}(U)$ , for all  $w \in C_c^\infty(U)$  (identifying  $\Lambda^0 = L^1$ ), then the commutator  $[Y, \Phi_\tau] = Y\Phi_\tau - \Phi_\tau Y$  satisfies*

$$\|[Y, \Phi_\tau]^* w\|_{L^1} \lesssim_U \tau^\sigma \|w\|_{\Lambda^\sigma}.$$

*Proof* It suffices to show that the lemma holds for all  $\Phi^j$ . Additionally, since  $Y^* = -Y - \operatorname{div} Y$  and by the fact that  $Y$  is smooth and bounded, we easily obtain

$$\|[\operatorname{div} Y, (\Phi^j)^*] w\|_{L^1} \lesssim_U \tau \|w\|_{L^1},$$

and therefore since  $[Y, \Phi_\tau]^* = -[Y, \Phi_\tau^*] - [\operatorname{div} Y, \Phi_\tau^*]$  it suffices to prove that

$$\|[Y, (\Phi_\tau^j)^*] w\|_{L^1} \lesssim_U \tau^\sigma \|w\|_{\Lambda^\sigma}. \quad (4.16)$$

With this goal in mind, writing  $Y(x) = a_k(x) \partial_{x_k}$  (using Einstein notation summation) as the local  $\mathbf{x}_j$  parameterization of the vector field  $Y$ , we find

$$\begin{aligned} & \| [Y, (\Phi_\tau^j)^*] w \|_{L^1} \\ &= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \left( a_k(x) \partial_{x_k} \phi_\tau(|x-y|) \tilde{w}_j(y) - \phi_\tau(|x-y|) a_k(y) \partial_{y_k} \tilde{w}_j(y) \right) dy \right| J_j(x) dx. \end{aligned}$$

Integrating by parts in  $\partial_{y_k}$ , using  $\partial_{x_k} \phi_\tau(|x-y|) = -\partial_{y_k} \phi_\tau(|x-y|)$ ,  $\partial_{y_k} \tilde{w}_j(x) = 0$  and that

$$\int_{\mathbb{R}^d} a_k(x) \partial_{x_k} \phi_\tau(|x-y|) \tilde{w}_j(y) dy = 0,$$

we obtain

$$\begin{aligned} & \| [Y, (\Phi_\tau^j)^*] w \|_{L^1} \lesssim \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (a_k(x) - a_k(y)) \partial_{x_k} \phi_\tau(|x-y|) (\tilde{w}_j(y) - \tilde{w}_j(x)) dy \right| J_j(x) dx \\ &+ \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \partial_{y_k} (a_k(y)) \phi_\tau(|x-y|) (\tilde{w}_j(y) - \tilde{w}_j(x)) dy \right| J_j(x) dx. \end{aligned}$$

Using that  $|a_k(x) - a_k(y)| \lesssim |x-y|$  and  $\partial_{y_k} (a_k(y)) \lesssim 1$  gives

$$\begin{aligned} & \| [Y, (\Phi_\tau^j)^*] w \|_{L^1} \lesssim \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left( \left| \frac{|x-y|}{\tau^{d+1}} \phi' \left( \frac{|x-y|}{\tau} \right) \right| \right. \\ &+ \left. \left| \frac{1}{\tau^d} \phi \left( \frac{|x-y|}{\tau} \right) \right| \right) |\tilde{w}_j(y) - \tilde{w}_j(x)| J_j(x) dy dx. \end{aligned}$$

Then, making the change of variables  $y = x + h$ , we obtain from the definition of  $\Lambda^\sigma$  (4.2), we obtain (4.16).  $\square$

Next we prove the following regularization estimate.

**Lemma 4.14** *For all  $\sigma \in [0, 1)$ , for all  $U \subset \mathcal{M}$  open, bounded geodesic balls, there holds uniformly over  $\tau \in (0, 1)$  and uniformly over bounded  $C^2$  sets of  $Y \in \mathfrak{X}(K)$ , for all  $w \in C_c^\infty(U)$  and  $p \in [1, \infty]$ ,*

$$\| \tau Y \Phi_\tau^* w \|_{L^p} + \| \tau Y \Phi_\tau w \|_{L^p} \lesssim_U \| w \|_{L^p} \quad (4.17)$$

$$\| \tau Y \Phi_\tau^* w \|_{L^1} + \| \tau Y \Phi_\tau w \|_{L^1} \lesssim_U \tau^\sigma \| w \|_{\Lambda^\sigma}. \quad (4.18)$$

$$\| \Phi_\tau^* \tau Y w \|_{L^1} + \| \Phi_\tau \tau Y w \|_{L^1} \lesssim_U \tau^\sigma \| w \|_{\Lambda^\sigma}. \quad (4.19)$$

*Proof* We will only consider the case of  $\Phi$ ; the corresponding estimates involving  $\Phi^*$  are similar. We proceed with a proof that is similar to, but much simpler than, that used in Lemma 4.13. We consider first (4.18); (4.19) follow from similar arguments. Using the average zero property again,

$$\begin{aligned} \left\| \tau Y \Phi_\tau^j w_j \right\|_{L^1} &\lesssim \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \tau a_k(x) \partial_{x_k} \phi_\tau(|x-y|) (\tilde{w}_j(y) - \tilde{w}_j(x)) J_j(x) dy \right| J_j(x) dx \\ &\quad + \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \tau a_k(x) \partial_{x_k} \phi_\tau(|x-y|) (J_j(y) - J_j(x)) \tilde{w}_j(y) dy \right| J_j(x) dx \\ &\lesssim \tau^\sigma \|w\|_{\Lambda^\sigma}, \end{aligned}$$

which is the desired estimate.

Turn next to (4.17). The  $p = 1$  case follows from a straightforward variant of the argument used for (4.18). For the end-point  $p = \infty$  we have by Young's inequality

$$\left| \tau Y \Phi_\tau^j w_j \circ \mathbf{x}_j(x) \right| \lesssim \int_{\mathbb{R}^d} \tau |a_k(x) \partial_{x_k} \phi_\tau(|x-y|) \tilde{w}_j(y)| dy \lesssim \|w\|_{L^\infty}.$$

The intermediate  $p$  follows by the Riesz-Thorin interpolation theorem.  $\square$

Next, we introduce directional regularizations with respect to a given vector field  $Y \in \mathfrak{X}$ , as done in [Sect. 5; [41]]. Accordingly, for each  $\varphi \in C_c^\infty([-1, 1])$  and  $\tau \in (0, 1)$  define

$$\varphi_{\tau Y} w := \int_{\mathbb{R}} (e^{tY} w) \varphi_\tau(t) dt,$$

where  $\varphi_\tau(t) := \frac{1}{\tau} \varphi(\tau^{-1}t)$ . Note that since  $(e^{tY})^* = (e^{tY^*})$ , we have

$$(\varphi_{\tau Y})^* w = \varphi_{\tau Y^*} w = \int_{\mathbb{R}} (e^{tY^*} w) \varphi_\tau(t) dt,$$

a property that will be used repeatedly in the sequel.

First we record the basic property that these regularizers are bounded on  $L^p$ . The proof is straightforward and is omitted for brevity.

**Lemma 4.15** *For any  $Y \in \mathfrak{X}$ , for any open, bounded geodesic ball  $U \subset \mathcal{M}$ , and  $\varphi \in C_c^\infty([-1, 1])$  there holds for all  $p \in [1, \infty]$ , and  $w \in C_c^\infty(U)$ ,*

$$\begin{aligned} \|(\varphi_{\tau Y})^* w\|_{L^p} &\lesssim \|w\|_{L^p} \\ \|\Phi_\tau w\|_{L^p} + \|\Phi_\tau^* w\|_{L^p} &\lesssim \|w\|_{L^p}. \end{aligned}$$

Next, we note that the regularizations, the adjoint regularizations, and vector field exponentials are bounded in the  $\Lambda^s$  space.

**Lemma 4.16** *For  $|t| \leq 1$ ,  $\tau \in (0, 1)$  and  $\sigma \in [0, 1)$ , for all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ , there holds*

$$\|e^{tY} w\|_{\Lambda^\sigma} \lesssim \|w\|_{\Lambda^\sigma} \tag{4.20}$$

$$\begin{aligned}
\|e^{tY^*} w\|_{\Lambda^\sigma} &\lesssim \|w\|_{\Lambda^\sigma} \\
\|(\varphi_{tY})^* w\|_{\Lambda^\sigma} &\lesssim \|w\|_{\Lambda^\sigma} \\
\|\Phi_\tau w\|_{\Lambda^\sigma} + \|\Phi_\tau^* w\|_{\Lambda^\sigma} &\lesssim \|w\|_{\Lambda^\sigma}.
\end{aligned} \tag{4.21}$$

*Proof* The last estimate (4.21) follows analogously to the corresponding Hölder-type estimate on  $\mathbb{R}^d$ ; we omit the details for brevity. The second and third estimates follow easily from the first estimate, and hence we turn to the proof of (4.20).

After applying coordinate parameterization to reduce to the case of  $\mathbb{R}^d$ , the first estimate will follow from an adaptation of [Lemma 4.2; [41]] to  $L^1$  and more general geometry. We include the proof for the readers' convenience. Recalling the coordinate charts defined in for (4.2), denote  $\tilde{w}_j = (\chi_j w) \circ \mathbf{x}_j$  and note that  $\exists g_{Y,j}(t, x) \in C^\infty(\mathbb{R}^d)$  such that  $(\chi_j e^{tY} w) \circ \mathbf{x}_j = \tilde{w}_j(g_{Y,j}(t, x))$  satisfying exactly analogous properties that  $h_Y$  satisfies. Therefore, by the definition (4.2)

$$\begin{aligned}
\|e^{tY} w\|_{\Lambda^\sigma} &= \|e^{tY} w\|_{L^1} \\
&+ \sup_{\tau \in \mathbb{R}^d: |\tau| < 1} \sum_j \int_{\mathbb{R}^d} \frac{|\tilde{w}_j(g_{Y,j}(t, x + \tau)) - \tilde{w}_j(g_{Y,j}(t, x))|}{|\tau|^s} J_j(x) dx.
\end{aligned}$$

The first term is estimated by (4.10). We next turn to the latter term. By the group property of  $e^{tY}$  we may restrict  $|t| < \delta'$  to any  $\delta'$  sufficiently small depending only on  $Y$  and the ball  $U$  (i.e. local geometrical information). Furthermore, we may restrict  $|\tau| < \delta'$  (as the contribution from  $|\tau| > \delta'$  is controlled by  $L^1$ ). Next, observe that (using the same trick as in [Lemma 4.2; [41]])

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{|\tilde{w}_j(g_{Y,j}(t, x + \tau)) - \tilde{w}_j(g_{Y,j}(t, x))|}{|\tau|^s} J_j(x) dx \\
&\lesssim |\tau|^{-d} \int_{|\zeta| < |\tau|} \int_{\mathbb{R}^d} \frac{|\tilde{w}_j(g_{Y,j}(t, x + \tau)) - \tilde{w}_j(g_{Y,j}(t, x + \zeta))|}{|\zeta|^s} J_j(x) dx d\zeta + \|w\|_{\Lambda^s}.
\end{aligned}$$

Then for each fixed  $j$  we make the  $(t, \tau)$ -dependent change of variables from  $(x, \zeta) \rightarrow (\gamma, \eta)$  in the integrals via  $\eta + \gamma = g_{Y,j}(t, x + \tau)$  and  $\gamma = g_{Y,j}(t, x) + \zeta$ . Note that for  $t = 0$  this reduces to  $\gamma = x + \zeta$  and  $\eta = -\zeta$  and for  $\tau = 0$  this becomes  $\gamma = g_{Y,j}(t, x) + \zeta$  and  $\eta = -\zeta$ . Hence, for  $\delta'$  chosen small enough, the Jacobian  $\mathcal{J}_{t,\tau}$  for this change of variables satisfies  $|\mathcal{J}_{t,\tau} - 1| < 1/2$  and we obtain

$$\begin{aligned} & \int_{\mathbb{R}^d} \frac{|\tilde{w}_j(g_{Y,j}(t, x + \tau)) - \tilde{w}_j(g_Y(t, x))|}{|\tau|^s} J_j(x) dx \\ & \lesssim |\tau|^{-d} \int_{\mathbb{R}^d} \int_{\mathcal{D}_{t,\tau}} \frac{|\tilde{w}_j(\eta + \gamma) - \tilde{w}_j(\gamma)|}{|\tau|^s} d\eta d\gamma + \|w\|_{\Lambda^s}, \end{aligned}$$

over some bounded integration domain  $\mathcal{D}_{t,\tau} \subseteq \mathbb{R}^d$  satisfying  $\text{vol } \mathcal{D}_{t,\tau} \lesssim |\tau|^d$ . The desired estimate (4.20) follows upon noting that  $|\eta| \lesssim \tau$  on  $\mathcal{D}_{t,\tau}$ .  $\square$

In a similar vein, the chain rule implies the following estimates.

**Lemma 4.17** *For all bounded, geodesic balls  $U \subset \mathcal{M}$ , for all  $|\tau| \leq 1$  and  $\forall k \geq 2$ , the following holds  $\forall w \in C_c^\infty(U)$ ,*

$$\sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z e^{\tau Y} w\|_{L^\infty} \lesssim \sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z w\|_{L^\infty}, \quad (4.22)$$

$$\sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z^* e^{\tau Y^*} w\|_{L^1} \lesssim \sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z^* w\|_{L^1}, \quad (4.23)$$

$$\sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z^* (\varphi_{\tau Y})^* w\|_{L^1} \lesssim \sup_{Z \in \mathfrak{X}: \|Z\|_{C^k} \leq 1} \|Z^* w\|_{L^1}. \quad (4.24)$$

*Proof* Estimates (4.22), (4.23) follow from the chain rule and (4.24) then follows from the definition of the regularizers.  $\square$

The next lemma characterizes the regularization property of the regularizers.

**Lemma 4.18** *For all bounded, geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ ,*

$$\|(Y \varphi_{\tau Y})^* w\|_{L^1} \lesssim \sup_{|t| \leq \tau} \|e^{t Y^*} w - w\|_{L^1}.$$

*Proof* We have

$$\begin{aligned} (\varphi_{\tau Y})^* Y^* w &= \int_{\mathbb{R}} (e^{t Y^*} Y^* w) \varphi_{\tau}(t) dt = \int_{\mathbb{R}} \frac{d}{dt} (e^{t Y^*} w - w) \varphi_{\tau}(t) dt \\ &= - \int_{\mathbb{R}} (e^{t Y^*} w - w) \varphi'_{\tau}(t) dt. \end{aligned}$$

The result then follows by Minkowski's inequality.  $\square$

We will also need the  $L^\infty$  regularization property.

**Lemma 4.19** *For all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ ,*

$$\|Y \varphi_{\tau Y} w\|_{L^\infty} \lesssim \frac{1}{\tau} \|w\|_{L^\infty}.$$

*Proof* This follows by a straightforward variant of the proof of Lemma 4.18.  $\square$

Next, we show that the Hölder-type regularity classes are natural for controlling convergence of the operators. It is natural to specialize to the specific form in which we are using it.

**Lemma 4.20** *For all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ , there holds for  $\varphi \in C_c^\infty([-1, 1])$ ,  $\varphi \geq 0$  and  $\int_{\mathbb{R}} \varphi(t) dt = 1$ ,*

$$\|(\varphi_{\tau X_I})^* w - w\|_{L^1} \lesssim \sup_{|t| \leq \tau} \|e^{t X_I^*} w - w\|_{L^1}.$$

*Proof* By Minkowski's inequality,

$$\|(\varphi_{\tau X_I})^* w - w\|_{L^1} \leq \int_{\mathbb{R}} \|e^{t X_I^*} w - w\|_{L^1} \varphi_{\tau}(t) dt \leq \sup_{|t| \leq \tau} \|e^{t X_I^*} w - w\|_{L^1}.$$

$\square$

The following Lemma will be useful when measuring regularity of  $\varphi_{\tau X_I}$  with respect to  $X_J$ .

**Lemma 4.21** *For all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ , for  $I, J \in \mathcal{J}$ , there holds*

$$\begin{aligned} & \sup_{|t| \leq \tau^{m(J)}} \|e^{t X_J^*} (\varphi_{\tau^{m(I)} X_I})^* w - (\varphi_{\tau^{m(I)} X_I})^* w\|_{L^1} \\ & \lesssim \sup_{|t| \leq \tau^{m(J)}} \|e^{t X_J^*} w - w\|_{L^1} + \sup_{|t| \leq \tau^{m(I)}} \|e^{t X_I^*} w - w\|_{L^1}. \end{aligned}$$

Now, we are ready to define the regularizer  $S_\tau$ . Let us now give  $\mathcal{J}$  a total ordering so that  $m(I)$  is an increasing function of  $I \in \mathcal{J}$  and we denote  $\mathcal{J}_\infty = \mathcal{J} \cup \{\infty\}$ . We define  $S_\tau$  in terms of an *ascending*, ordered composition of regularizing operators

$$S_\tau w := \left( \prod_{I \in \mathcal{J}} \varphi_{\tau^{m(I)} X_I} \right) \Phi_{\tau^{1/\sigma}} w.$$

This regularizer is similar, but not quite exactly the same as that defined in [41] due to the inclusion of more regularization operators. However, we will ultimately use  $S_\tau^*$  as the regularizer, which is a little more subtle to work with.

Analogous to [41], we also define the truncated regularizer, for all  $J \in \mathcal{J}$ ,

$$S_\tau^J w := \left( \prod_{I \in \mathcal{J}: I \geq J} \varphi_{\tau^{m(I)} X_I} \right) \Phi_{\tau^{1/\sigma}} w,$$

with  $S_\tau^\infty := \Phi_{\tau^{1/\sigma}}$  when  $J = \infty$ . The remainder of the subsection is dedicated to proving Lemma 4.12. The first step is to obtain the  $L^1$  convergence.

**Lemma 4.22** *For all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ ,*

$$\|S_\tau^* w - w\|_{L^1} \lesssim \tau |w|_M.$$

*Proof* For any finite family of  $L^1 \rightarrow L^1$  bounded linear operators  $Z_1, Z_2, \dots, Z_k$  we have

$$\|Z_1 Z_2 \dots Z_k w - w\|_{L^1} \leq \sum_{j=1}^k \left\| (Z^1 \dots Z^{j-1})(Z_j w - w) \right\|_{L^1} \lesssim \sum_{j=1}^k \|Z_j w - w\|_{L^1}.$$

The result then follows from Lemma 4.20.  $\square$

The next Lemma is crucial for characterizing the regularization properties of  $(S_\tau^J)^*$  in  $L^1$ . This is the adjoint analogue of [Lemma 5.2; [41]], which is a little more technical.

**Lemma 4.23** *For all open, bounded geodesic balls  $U \subset \mathcal{M}$  and  $w \in C_c^\infty(U)$ , there holds for any multi-indices  $J \leq I$ ,*

$$\left\| (\tau^{1/\sigma} Y S_\tau^J)^* w \right\|_{L^1} \lesssim \tau \|w\|_{\Lambda^\sigma} \quad (4.25)$$

$$\left\| (\tau^{m(I)} X_I S_\tau^J)^* w \right\|_{L^1} \lesssim \sum_{I' \in \mathcal{J}: I' \geq J} \sup_{|t| \leq \tau^{m(I')}} \left\| e^{t X_{I'}^*} w - w \right\|_{L^1} + \tau \|w\|_{\Lambda^\sigma}. \quad (4.26)$$

Before we continue, we define for two vector fields  $X$  and  $Y$

$$e^{t \operatorname{ad}(X)} Y := (e^{t X})_\sharp Y,$$

where  $(e^{t X})_\sharp Y$  denotes the pushforward of  $Y$  as a vector field under the diffeomorphism  $e^{t X}$ . This is just the adjoint representation of group element  $e^{t X}$  on the Lie algebra of vector fields. It will be useful to expand  $e^{t \operatorname{ad}(X)} Y$  in a Taylor expansion.

**Lemma 4.24** *For two smooth vector fields  $X, Y$ ,  $t \in [-1, 1]$  and  $N \in \mathbb{N}$ , there exists a smooth bounded vector field  $Y_{N,t}$ , locally uniformly bounded in  $C^k$  ( $\forall k$ ) on  $t \in [-1, 1]$ , such that*

$$(e^{t \operatorname{ad}(X)} Y) = \sum_{0 \leq k < N} \frac{t^k}{k!} (\operatorname{ad}(X))^k Y + \frac{t^N}{N!} Y_{N,t}$$

A simple consequence of the chain rule implies that  $Y(e^{tX} w) = e^{tX} (e^{t \operatorname{ad}(X)} Y w)$  and gives the following useful representation for the smoothing operators

$$Y \varphi_{\tau X} w = \int_{\mathbb{R}} \left( e^{tX} (e^{-t \operatorname{ad}(X)} Y w) \right) \varphi_{\tau}(t) dt.$$

Lemma 4.24 then gives the following formula for  $Y \varphi_{\tau X}$  (used also in [41]).

**Lemma 4.25** *For each  $\varphi \in C_c^\infty((-1, 1))$ ,  $k \in \mathbb{N}$  and vector field  $X$  define*

$$\hat{\varphi}_{\tau X}^k w := \int_{\mathbb{R}} (e^{tX} w) \hat{\varphi}_{\tau}^k(t) dt, \quad \text{where } \hat{\varphi}^k(t) := \frac{t^k}{k!} \varphi(t) \in C_c^\infty((-1, 1)). \quad (4.27)$$

*Then for two smooth vector fields  $X, Y$ ,  $\tau \in (0, 1]$  and  $N \in \mathbb{N}$ , the following holds*

$$Y \varphi_{\tau X} w = \sum_{0 \leq k < N} \tau^k \left( \hat{\varphi}_{\tau X}^k (\operatorname{ad}(-X))^k Y w \right) + \tau^N R_{\tau X}^N w,$$

where

$$R_{\tau X}^N w := \int_{\mathbb{R}} (e^{tX} Y_{N,t} w) \hat{\varphi}_{\tau}^N(t) dt. \quad (4.28)$$

We are now ready to prove Lemma 4.23.

*Proof of Lemma 4.23* **Proof of (4.25).** We proceed by induction. For  $J = \infty$  the result follows from (4.19). Hence, we next assume that the result holds for all  $J'$  with  $J' > J$  and prove that it also holds for  $J$ . We begin with the decomposition

$$(S_{\tau}^J)^* = (S_{\tau}^{J'})^* (\varphi_{\tau^{m(J)} X_J})^*.$$

By a trivial application of Lemma 4.25 with  $N = 1$  and  $X = X_J$ , there exists a smooth bounded vector field  $Y_{1,t}$  such that (recall Definition (4.28)),

$$(\tau^{1/\sigma} Y S_\tau^J)^* = (\tau^{1/\sigma} Y S_\tau^{J'})^* (\varphi_{\tau^{m(J)} X_J})^* + \tau^{m(J)+1/\sigma} (S_\tau^{J'})^* (R_{\tau^{m(J)} X_J}^1)^*.$$

By the induction hypothesis and Lemma 4.16 we have for the first term above

$$\|(\tau^{1/\sigma} Y S_\tau^{J'})^* (\varphi_{\tau^{m(J)} X_J})^* w\|_{L^1} \lesssim \tau \|(\varphi_{\tau^{m(J)} X_J})^* w\|_{\Lambda^\sigma} \lesssim \tau \|w\|_{\Lambda^\sigma}.$$

A similar estimate holds for the second term using Minkowski's inequality

$$\begin{aligned} \|(\tau^{1/\sigma} S_\tau^{J'})^* (R_{\tau^{m(J)} X_J}^1)^* w\|_{L^1} &\leq \int_{\mathbb{R}} \|(\tau^{1/\sigma} Y_{1,t} S_\tau^{J'})^* e^{t X_J^*} w\|_{L^1} \hat{\varphi}_{\tau^{m(J)}}^1(t) dt \\ &\lesssim \tau \|e^{t X_J^*} w\|_{\Lambda^\sigma} \lesssim \tau \|w\|_{\Lambda^\sigma} \end{aligned}$$

and the estimate (4.25) now follows.

**Proof of (4.26).** First we note that if  $I = J$  then we have for  $J'$  the smallest element such that  $J' > J$  by Lemma 4.18 and the  $L^1$  boundedness of  $(S_\tau^J)^*$

$$\|(X_J S_\tau^J)^* w\|_{L^1} = \|(S_\tau^{J'})^* (X_J \varphi_{\tau^{m(J)} X_J})^* w\|_{L^1} \lesssim \sup_{|t| \leq \tau^{m(J)}} \|e^{t X_J^*} w - w\|_{L^1}.$$

When  $I > J$ , we proceed by induction. First of all, the result follows by definition of (4.16) if  $J = \infty$ . Hence, we next assume that the result holds for all  $J'$  with  $J' > J$  and prove that it also holds for  $J$  the largest element less than  $J'$ . Again writing

$$(S_\tau^J)^* = (S_\tau^{J'})^* (\varphi_{\tau^{m(J)} X_J})^*$$

and using Lemma 4.25 we obtain,  $\forall N \geq 1$ ,

$$\begin{aligned} (\tau^{m(I)} X_I S_\tau^J)^* &= \sum_{0 \leq k < N} (\tau^{m(I'_k)} X_{I'_k} S_\tau^{J'})^* (\hat{\varphi}_{\tau^{m(J)} X_J}^k)^* + (\tau^{m(I)+Nm(J)} S_\tau^{J'})^* (R_{\tau^{m(J)} X_J}^N)^* \\ &=: T_1 + T_2, \end{aligned}$$

where

$$X_{I'_k} := \text{ad}(-X_J)^k X_I, \quad \text{and} \quad m(I'_k) = m(I) + km(J).$$

If we choose  $N$  large enough so that

$$m(I) + Nm(J) \geq \frac{1}{\sigma},$$

we can treat the “error” term  $T_2$  by applying Minkowski’s inequality and (4.25),

$$\begin{aligned}\|T_2 w\|_{L^1} &\leq \int_{\mathbb{R}} \|(\tau^{1/\sigma} Y_{N,t} S_{\tau}^{J'})^* e^{tX_J^*} w\|_{L^1} \hat{\varphi}_{\tau^m(J)}^N(t) dt \lesssim \tau \|e^{tX_J^*} w\|_{\Lambda^\sigma} \\ &\leq \tau \|w\|_{\Lambda^\sigma}.\end{aligned}$$

Since  $I'_k \geq J'$  for all  $k \geq 0$ , we can use the induction hypothesis and Lemmas 4.21, 4.15 (it is straightforward to check that  $\hat{\varphi}^k$  satisfies the same) to treat the first term

$$\begin{aligned}\|T_1 w\|_{L^1} &\lesssim \sum_{0 \leq k < N} \sum_{I' \geq J'} \sup_{|t| \leq \tau^m(I')} \left\| e^{tX_{I'}^*} (\hat{\varphi}_{\tau^m(J)}^k)_{X_J}^* w - (\hat{\varphi}_{\tau^m(J)}^k)_{X_J}^* w \right\|_{L^1} + \tau \left\| (\hat{\varphi}_{\tau^m(J)}^k)_{X_J}^* w \right\|_{\Lambda^\sigma} \\ &\lesssim \sum_{I' \geq J} \sup_{|t| \leq \tau^m(I')} \|e^{tX_{I'}^*} w - w\|_{L^1} + \tau \|w\|_{\Lambda^\sigma}\end{aligned}$$

as desired.  $\square$

The main commutator estimate is a consequence of the following.

**Lemma 4.26** *For all  $J \in \mathcal{J}$ ,  $U \subset \mathcal{M}$  open, bounded geodesic ball, there holds  $\forall w \in C_c^\infty(U)$  and  $\tau \in (0, 1)$ ,*

$$\left\| [\tau^2 X_0, S_{\tau}^J]^* w \right\|_{L^1} \lesssim \sum_{I \in \mathcal{J}: I \geq J} \sup_{|t| \leq \tau^m(I)} \left\| e^{tX_I^*} w - w \right\|_{L^1} + \tau \|w\|_{\Lambda^s}.$$

*Proof* As in the proof of (4.26) above, we proceed by induction. Firstly, the estimate holds for  $J = \infty$  due to the commutator estimate Lemma 4.13. As above, assume the result holds for all  $J'$  with  $J' > J$  and prove that it also holds for  $J$ , writing

$$(S_{\tau}^J)^* = (S_{\tau}^{J'})^* (\varphi_{\tau^m(J)}_{X_J})^*.$$

Then,

$$[\tau^2 X_0, (S_{\tau}^J)]^* = [\tau^2 X_0, S_{\tau}^{J'}]^* (\varphi_{\tau^m(J)}_{X_J})^* + (S_{\tau}^{J'})^* [\tau^2 X_0, \varphi_{\tau^m(J)}_{X_J}]^*. \quad (4.29)$$

By the inductive hypothesis and Lemmas 4.21 and 4.15 we have

$$\begin{aligned}&\left\| [\tau^2 X_0, S_{\tau}^{J'}]^* (\varphi_{\tau^m(J)}_{X_J})^* w \right\|_{L^1} \\ &\lesssim \sum_{I' \geq J'} \sup_{|t| \leq \tau^m(I')} \left\| e^{tX_{I'}^*} (\varphi_{\tau^m(J)}_{X_J})^* w - (\varphi_{\tau^m(J)}_{X_J})^* w \right\|_{L^1} + \tau \left\| (\varphi_{\tau^m(J)}_{X_J})^* w \right\|_{\Lambda^\sigma}\end{aligned}$$

$$\leq c \sum_{I' \geq J} \sup_{|t| \leq \tau^{m(I')}} \|e^{tX_{I'}^*} w - w\|_{L^1} + \tau \|w\|_{\Lambda^\sigma}.$$

This term in (4.29) is consistent with the desired result.

For the second term in (4.29), by Lemma 4.25 we have (note the cancellation which eliminates the  $k = 0$  term)

$$(S_\tau^{J'})^* [\tau^2 X_0, \varphi_{\tau^m(J)X_J}]^* = \sum_{0 < k < N} (\tau^{m(I'_k)} X_{I'_k} S_\tau^{J'})^* (\hat{\varphi}_{\tau^m(J)X_J}^k)^* + (\tau^{2+Nm(J)} S_\tau^{J'})^* (R_{\tau^m(J)X_J}^N)^*$$

where

$$X_{I'_k} := \text{ad}(-X_J)^k X_0, \quad \text{and} \quad m(I'_k) = 2 + km(J).$$

The treatment of these terms is exactly the same as in the proof of (4.26) upon taking  $2 + Nm(J) \geq \frac{1}{\sigma}$ . We omit the repetitive details for the sake of brevity.  $\square$

Finally, we prove the required  $L^\infty$  regularization estimate.

**Lemma 4.27** *Let  $U \subset \mathcal{M}$  be open, bounded geodesic ball. Let  $I$  be any multi-index and  $J \leq I$ . Then,  $\forall w \in C_c^\infty(U)$ ,*

$$\left\| \tau^{m(I)} X_I S_\tau^J w \right\|_{L^\infty} \lesssim_U \|w\|_{L^\infty}$$

*Proof* This is done by induction as in previous lemmas. The case  $J = \infty$  follows from (4.17). Assume that the result holds for all  $J'$  with  $J' > J$  and write

$$S_\tau^J = \varphi_{\tau^m(J)X_J} S_\tau^{J'}.$$

**Case 1:**  $I > J$ . We apply the Taylor expansion of Lemma 4.25 (recalling definitions (4.27) and (4.28))

$$\tau^{m(I)} X_I S_\tau^J = \sum_{0 \leq k < N} \hat{\varphi}_{\tau^m(J)X_J}^k \tau^{m(I'_k)} X_{I'_k} S_\tau^{J'} + \tau^{m(I)+Nm(J)} R_{\tau^m(J)X_J}^N S_\tau^{J'},$$

where

$$X_{I'_k} := \text{ad}(-X_J)^k X_I, \quad \text{and} \quad m(I'_k) = m(I) + km(J).$$

Then we have,

$$\left\| \tau^{m(I)} X_I S_\tau^J w \right\|_{L^\infty} \lesssim \sum_{0 \leq \nu < N} \left\| \tau^{m(I'_k)} X_{I'_k} S_\tau^{J'} w \right\|_{L^\infty} + \sup_{Z \in \mathfrak{X}: \|Z\|_{C^3} \leq 1} \left\| \tau^{m(I) + Nm(J)} Z S_\tau^{J'} w \right\|_{L^\infty}.$$

For  $N$  chosen sufficiently large, the latter term is estimated by Lemma 4.17 an (4.17) to produce

$$\sup_{Z \in \mathfrak{X}: \|Z\|_{C^3} \leq 1} \left\| \tau^{m(I) + Nm(J)} Z S_\tau^{J'} w \right\|_{L^\infty} \lesssim \|w\|_{L^\infty}.$$

Since  $I'_k \geq J'$  and hence the first  $N$  terms in the summation are estimated by induction.

**Case 2:**  $I = J$ . In this case, the desired result follows immediately from Lemma 4.19.  $\square$

## 5 Projective hypoellipticity and irreducibility for Euler-like systems

Our focus for the remainder of the paper is to prove Theorem C for the Lyapunov exponents of Euler-like systems (see (1.7)), as well as Corollary D for our concrete example, the Lorenz 96 (L96) model (1.9). Our tools, the Fisher information identity (Theorem A) and accompanying hypoelliptic regularity estimate (Theorem B), depend on the unique existence of absolutely continuous stationary densities for projective processes (c.f. Assumptions 1, 2 and 3). As discussed in Sect. 2, these can be reduced to checking Hörmander's condition (Definition 2.1) and topological irreducibility (Definition 2.3) for the projective process  $w_t = (x_t, v_t)$  on  $\mathbb{S}M$ . The primary aim of this section is to obtain useful sufficient conditions for checking these for Euler-like models in general, and then to verify them for the L96 model in particular.

The plan is as follows: in Sect. 5.1 we present a sufficient condition (Corollary 5.2) for the Euler-like class to satisfy projective bracket spanning; in Sect. 5.2, we check this condition for the L96 model; and in Sect. 5.3 we show that topological irreducibility for Euler-like models can be *reduced* to bracket spanning (Proposition 5.7) under a natural cancellation-type condition satisfied by most models of interest (including the L96 and Galerkin Navier–Stokes equations).

### Standing assumptions

For the remainder of the paper, we return to the setting of Euler-like SDE

$$dx_t^\epsilon = (B(x_t^\epsilon, x_t^\epsilon) + \epsilon Ax_t^\epsilon) dt + \sqrt{\epsilon} \sum_{k=1}^r X_k dW_t^k,$$

where as in Sect. 1 we assume  $B$  is a nontrivial bilinear form on  $\mathbb{R}^n$ ,  $n \geq 1$  satisfying  $\operatorname{div} B = 0$  and  $x \cdot B(x, x) = 0$ ; the  $\{X_k\}_{k=1}^r$  are constant vector fields; and  $A$  is a symmetric negative-definite  $n \times n$  matrix. Recall the notation  $X_0^\epsilon(x) := B(x, x) + \epsilon Ax$ .

## 5.1 Projective spanning for Euler-like systems

For many systems of interest, it can be significantly harder to verify Hörmander's condition for  $\{\tilde{X}_k\}$  on  $\mathbb{SR}^n$  than to verify it for the vector fields  $\{X_k\}$  on  $\mathbb{R}^n$ , for the simple reason that the dimension increases substantially ( $\dim \mathbb{SR}^n = 2n-1$ ) and the fact that it is significantly harder to isolate 'simple' vector fields on  $\mathbb{SR}^n$ ; this is already the case for the L96 model with additive noise (which we treat in Sect. 5.2 below).

First, we show that for the class of Euler-like models (1.7), projective spanning as in Assumption 3 can be reduced to a transitivity condition on a finite dimensional matrix Lie algebra. This provides both a reasonable way to check projective spanning for *fixed* dimension and parameters using a brute force computer calculation as well as a framework to begin analytical studies that work in arbitrary dimension and over entire ranges of parameters (such as what is carried out for Lorenz 96 in Sect. 5.2 and for Galerkin–Navier–Stokes in [19]). To make this more precise, we present here an argument for reducing projective spanning to a combination of

- (i) the spanning condition for  $\{X_0^\epsilon, X_1, \dots, X_r\}$  on  $\mathbb{R}^d$ ; and
- (ii) the *purely linear* condition that  $\mathfrak{sl}(\mathbb{R}^n)$  is generated by a collection  $\{H^i\}$  of *constant-valued*  $n \times n$  real matrices (defined explicitly in terms of derivatives of the nonlinearity  $B(x, x)$ ) under the standard matrix Lie bracket.

By Proposition 2.7, we know that verifying the parabolic Hörmander condition for  $\{\tilde{X}_0^\epsilon, \tilde{X}_1, \dots, \tilde{X}_r\}$  on  $\mathbb{SR}^n = \mathbb{R}^n \times \mathbb{S}^{n-1}$  is equivalent to checking that  $\{X_0^\epsilon, X_1, \dots, X_r\}$  satisfies the parabolic Hörmander condition on  $\mathbb{R}^n$  and that the matrix Lie algebra  $\mathfrak{m}_x(X_0^\epsilon; X_1, \dots, X_r)$  defined by (2.4) satisfies the transitivity condition

$$\{Av - v\langle Av, v \rangle : A \in \mathfrak{m}_x(X_0^\epsilon; X_1, \dots, X_r)\} = T_v \mathbb{S}^{n-1}$$

for each  $(x, v) \in \mathbb{R}^n \times \mathbb{S}^{n-1}$ . In general it is a challenge to directly work with  $\mathfrak{m}_x(X_0^\epsilon; X_1, \dots, X_r)$  as it is not a simple task to classify all vector fields in  $\operatorname{Lie}(X_0^\epsilon; X_1, \dots, X_r)$  that vanish at each  $x \in \mathbb{R}^n$ . However, in  $\mathbb{R}^n$  with additive noise it is often the case that the parabolic Lie algebra generated by  $\{X_k\}_{k=0}^r$  contains a spanning collection of constant vector fields  $\{\partial_{x_k}\}_{k=1}^n$  (this is often how parabolic Hörmander as in Definition 2.1 is checked on  $\mathbb{R}^n$ ). In this case,  $\mathfrak{m}_x$  can be described more explicitly.

**Lemma 5.1** *Let  $\{Z_k\}_{k=0}^r \subseteq \mathfrak{X}(\mathbb{R}^n)$  and suppose that  $\text{Lie}(Z_0; Z_1, \dots, Z_r)$  contains the constant vector fields  $\{\partial_{x_k}\}_{k=1}^n$ . Then*

$$\mathfrak{m}_x(Z_0; Z_1, \dots, Z_r) = \{M_Z(x) : Z \in \text{Lie}(Z_0; Z_1, \dots, Z_r)\}.$$

*Proof* Our hypothesis  $\{\partial_{x_k}\}_{k=1}^n \subseteq \text{Lie}(X_0; X_1, \dots, X_r)$  implies that for each  $X \in \text{Lie}(X_0; X_1, \dots, X_r)$  and  $x \in \mathbb{R}^n$ , the vector field  $\hat{X} = X - X(x)$  also belongs to  $\text{Lie}(X_0; X_1, \dots, X_r)$  and satisfies  $\hat{X}(x) = 0$ . Since  $\nabla \hat{X} = \nabla X$ , we have  $M_{\hat{X}}(x) = M_X(x)$ , hence  $M_X(x) \in \mathfrak{m}_x$ .  $\square$

Using that  $X_0^\epsilon(x) = B(x, x) + \epsilon A x$  allows for further simplification. Define for each  $k = 1, \dots, n$  the linear operator

$$H^k := \nabla[\partial_{x_k}, X_0^\epsilon] = \partial_{x_k} \nabla B = M_{[\partial_{x_k}, X_0^\epsilon]} \in \mathfrak{sl}(\mathbb{R}^n).$$

Note that  $H^k$  is independent of both  $x \in M$  and  $\epsilon$ . Below,  $\text{Lie}(H^1, \dots, H^n)$  denotes the matrix Lie subalgebra of  $\mathfrak{sl}(\mathbb{R}^n)$  generated by  $\{H^1, \dots, H^n\}$ . It is a simple matter to check that if  $\{\partial_{x_k}\}_{k=1}^n \subseteq \text{Lie}(X_0^\epsilon; X_1, \dots, X_r)$ , then one has

$$\text{Lie}(H^1, \dots, H^n) = \mathfrak{m}_x(X_0^\epsilon; X_1, \dots, X_r).$$

Lemma 5.1 now yields the following.

**Corollary 5.2** *Assume (i)  $\{\partial_{x_k}\}_{k=1}^n \subseteq \text{Lie}(X_0^\epsilon; X_1, \dots, X_r)$  and (ii) that*

$$\text{Lie}(H^1, \dots, H^n) = \mathfrak{sl}(\mathbb{R}^n). \quad (5.1)$$

*Then,  $\{\tilde{X}_0^\epsilon, \tilde{X}_1, \dots, \tilde{X}_r\}$  satisfy the uniform parabolic Hörmander condition in the sense of Definition 4.1 as  $\epsilon$  is varied in  $(0, 1]$ .*

*Proof* By Lemma 5.1 and Proposition 2.7 it suffices to check that

$$\{Av - v\langle v, Av \rangle : A \in \mathfrak{sl}(\mathbb{R}^n)\} = T_v \mathbb{S}^{n-1}.$$

holds. This is true since  $\mathfrak{sl}(\mathbb{R}^n)$  acts transitively on  $\mathbb{R}^n \setminus \{0\}$ , i.e., for all  $w, v \in \mathbb{R}^n \setminus \{0\}$  there exists  $A \in \mathfrak{sl}(\mathbb{R}^n)$  such that  $Av = w$  (see, e.g., [25] for a complete classification of transitive algebras on  $\mathbb{R}^n \setminus \{0\}$ ). In particular, this implies transitive action of  $V_A$  on  $\mathbb{S}^{n-1}$  since for each  $v$ , the projection of a spanning set in  $\mathbb{R}^n$  onto the subspace  $T_v \mathbb{S}^{n-1} = v^\perp = \{v' \in \mathbb{S}^{n-1} : \langle v, v' \rangle = 0\}$  is still spanning. Indeed, it is not hard to show that the smaller subalgebra  $\mathfrak{so}(\mathbb{R}^n)$  of skew symmetric matrices also acts transitively on  $\mathbb{S}^{n-1}$  (though not

on  $\mathbb{R}^n \setminus \{0\}$ ), and so the conclusion of this Corollary holds under the following weakening of (5.1):

$$\mathfrak{so}(\mathbb{R}^n) \subseteq \text{Lie}(H^1, \dots, H^n),$$

c.f. Remark 2.8.  $\square$

*Remark 5.3* Let us comment briefly on how one might verify (5.1). Since  $\mathfrak{sl}(\mathbb{R}^n)$  is  $n^2 - 1$  dimensional, it is clear that one must use commutators that go several generations deep if one has any hope of generating  $\mathfrak{sl}(\mathbb{R}^n)$ . However, it can simplify things to instead look to build a suitable generating set for  $\mathfrak{sl}(\mathbb{R}^n)$  out of brackets of  $H^i$ 's. A particularly useful generating set for  $\mathfrak{sl}(\mathbb{R}^n)$  is the collection of elementary matrices

$$E^{1,2}, E^{2,3}, \dots, E^{n,1},$$

where  $E^{i,j}$  is the matrix with 1 in  $(i, j)$  entry and 0 elsewhere. For these, we have the commutation relation

$$[E^{i,j}, E^{k,\ell}] = E^{i,\ell} \delta_{j,k} - E^{k,j} \delta_{\ell,i},$$

so that, e.g.,

$$[E^{1,2}, E^{2,3}] = E^{1,3} \quad \text{and} \quad [E^{1,2}, E^{2,1}] = E^{1,1} - E^{2,2}.$$

Continuing like this allows to generate the off-diagonal matrices  $\{E^{i,j}\}_{i \neq j}$  as well as the directions  $E^{1,1} - E^{2,2}, \dots, E^{n,n} - E^{n-1,n-1}$  needed to complete a basis for  $\mathfrak{sl}(\mathbb{R}^n)$ . Therefore,  $\{E^{1,2}, E^{2,3}, \dots, E^{n,1}\}$  generates  $\mathfrak{sl}(\mathbb{R}^n)$ .

## 5.2 Projective spanning for Lorenz 96

Now we turn to verifying the uniform projective spanning for stochastically forced Lorenz 96 (1.9). Recall the stochastic Lorenz 96 is an SDE on  $\mathbb{R}^J$  defined by

$$du_\ell = (u_{\ell+1} - u_{\ell-2})u_{\ell-1}dt - \epsilon u_\ell dt + \sqrt{\epsilon}q_\ell dW_t^\ell. \quad (5.2)$$

Here, we assume a periodic ensemble of coupled oscillators, i.e.,  $u_{i+kJ} := u_i$ . Naturally we can write (5.2) in the general form (1.7) for Euler-like SDE by defining  $X_0(u) = B(u, u) + \epsilon Au$ , where the  $\ell$ -th coordinate of  $F(u) := B(u, u)$  is given by

$$F_\ell(u) = u_{\ell+1}u_{\ell-1} - u_{\ell-2}u_{\ell-1}, \quad (Au)_\ell = u_\ell.$$

First, we verify uniform hypoellipticity for the process  $(u_t)$  on  $\mathbb{R}^J$ .

**Lemma 5.4** *Let  $J \geq 2$  and assume  $q_1, q_2 \neq 0$ . Then  $\text{Lie}(F; q_1 \partial_{u_1}, q_2 \partial_{u_2})$  contains  $\{\partial_{u_j}\}_{j=1}^J$  and spans  $\mathbb{R}^J$  uniformly in  $\epsilon$  on compact sets.*

*Proof* Since the nonlinearity is bilinear, we readily observe that

$$[\partial_{u_2}, [\partial_{u_1}, F]] = -\partial_{u_3}.$$

Iterating this observation allows to generate all brackets of the form  $[\partial_{u_{i+1}}, [\partial_{u_i}, F]] = -\partial_{u_{i+2}}$ .  $\square$

In order to prove uniform projective spanning we first observe that

$$(\nabla F(u))_{\ell m} = DF_\ell(u)_m = u_{\ell-1} \delta_{m=\ell+1} + u_{\ell+1} \delta_{m=\ell} - u_{\ell-1} \delta_{m=\ell-2} - u_{\ell-2} \delta_{m=\ell-1},$$

hence it follows that for each  $k \in \{1, \dots, J\}$  we have

$$H^k = \partial_{u_k} DF(u) = E^{k+1,k+2} + E^{k-1,k-2} - E^{k+1,k-2} - E^{k+2,k+1}.$$

The following implies projective spanning for Lorenz-96 when combined Corollary 5.2.

**Lemma 5.5**

$$\text{Lie}(H^1, \dots, H^J) = \mathfrak{sl}(\mathbb{R}^J).$$

*Proof* Throughout, we regard the indices in  $E^{i,j}$  modulo  $J$ , so that  $E^{i+kJ, j+\ell J} = E^{i,j}$  for all  $i, j, k, \ell$ .

Let  $\mathfrak{g}$  denote the smallest Lie algebra containing  $\{H^k\}$ . To start, let  $1 \leq k \leq J$ . We compute

$$[H^k, H^{k+4}] = E^{k+3,k+1},$$

hence  $E^{k,k-2} \in \mathfrak{g}$  for all  $1 \leq k \leq J$ . Continuing,

$$[H^k, E^{k-2,k-4}] = E^{k-1,k-4},$$

hence  $E^{k,k-3} \in \mathfrak{g}$  for all  $k$ . Inductively, assuming  $E^{k,k-\ell} \in \mathfrak{g}$ , we have that

$$[H^k, E^{k-2,k-(\ell+2)}] = E^{k-1,k-1-(\ell+1)}, \quad (5.3)$$

hence  $E^{k,k-(\ell+1)} \in \mathfrak{g}$  for all  $k$ . The induction step in (5.3) continues to hold as long as  $k - (\ell + 2)$  is disjoint from  $\{k - 1, k + 1, k + 2\}$  modulo  $J$ , which is assured so long as  $\ell < J - 4$ .

Fix  $\ell_0 \in \{2, 3, \dots, J-5\}$  so that  $J - \ell_0$  is co-prime to  $J$ . In particular,  $\{1 - \ell_0, 1 - 2\ell_0, \dots, 1 - (J-1)\ell_0\}$  coincides with the complete set of residue classes  $\{0, 1, \dots, J-1\}$  in  $\mathbb{Z}/J\mathbb{Z}$ . We observe that

$$\left\{E^{1,1-\ell_0}, E^{1-\ell_0,1-2\ell_0}, \dots, E^{1-(J-2)\ell_0,1-(J-1)\ell_0}, E^{1-(J-1)\ell_0,1}\right\} \subset \mathfrak{g}$$

is a re-ordering of the generating set identified in Remark 5.3, and conclude  $\mathfrak{g} = \mathfrak{sl}(\mathbb{R}^J)$ .  $\square$

*Remark 5.6* We note that the proof presented here for L96 heavily relies on the ‘‘local’’ coupling of unknowns in the nonlinearity, which greatly simplifies the application of Corollary 5.2 in this case. However, for models which are the Galerkin truncations of PDEs, coupling between unknowns has a more ‘‘global’’ character, and so even with the simplifications presented here, verifying projective spanning takes significantly more work. For Galerkin Navier–Stokes, this is the topic of the recent paper [19] of the first and third author.

### 5.3 Projective irreducibility for Euler-like systems

We close this section with a criterion for irreducibility of the projective process  $w_t = (x_t, v_t)$  for our class of Euler-like systems. In what follows, we will assume that the nonlinearity  $B(x, x)$  satisfies the following *cancellation property* in addition to volume conservation  $\operatorname{div} B(x, x) = 0$  and  $x \cdot B(x, x) = 0$ :

**Assumption 4** There exists a collection of vectors  $\{e_1, \dots, e_s\} \subset \mathbb{R}^n$  with

$$\operatorname{span}\{e_1, \dots, e_s\} = \operatorname{span}\{X_1, \dots, X_r\}$$

such that for each  $1 \leq k \leq s$ ,  $B(e_k, e_k) = 0$ .

Systems for which Assumption 4 hold include the Lorenz-96 model (1.9) and Galerkin truncations of the Navier–Stokes equations. Cancellation properties for even degree nonlinearities are commonplace when dealing with controllability and also feature prominently in the works of e.g. [37, 40, 83].

The main result of the section shows that projective spanning combined with the cancellation condition in Assumption 4 is sufficient to imply the projective process  $(w_t)$  is irreducible. This is commonly used as a vital ingredient in obtaining uniqueness of the stationary measure via the Doob–Khasminskii theorem [30].

**Proposition 5.7** *Assume the bilinear term  $B$  satisfies Assumption 4. Then, the projective process  $(w_t)$  is irreducible for all  $\epsilon > 0$  if  $\{\tilde{X}_0^\epsilon, \tilde{X}_1, \dots, \tilde{X}_r\}$  satisfies the parabolic Hörmander condition  $\forall \epsilon > 0$ .*

### 5.3.1 Preliminaries on geometric control theory

To prove Proposition 5.7, we use the Stroock-Varadhan support theorem [86] to connect irreducibility to exact controllability (see Theorem 5.8 below). We first outline some basic ideas from geometric control theory, for which we mostly follow [43]. Consider the following affine control system

$$\dot{x}_t = X_0(x_t) + \sum_{k=1}^r X_k(x_t) \alpha_t^k \quad (5.4)$$

on a manifold  $M$ , where  $(\alpha_t^1, \dots, \alpha_t^r)$  are piecewise constant controls taking values in  $\mathbb{R}$  and  $\{X_0, X_1, \dots, X_r\}$  are smooth vector fields in  $\mathfrak{X}(M)$ . In what follows, let  $\mathcal{X} = \text{span}\{X_1, \dots, X_r\}$  and define the controlled distribution of vector fields

$$\mathcal{F}_0 := X_0 + \mathcal{X} = \{X_0 + X : X \in \mathcal{X}\} \subseteq \mathfrak{X}(M).$$

We also define the time  $t > 0$  accessible set starting from  $x \in M$  associated to  $\mathcal{F}_0$  by

$$\mathcal{A}_x^t(\mathcal{F}_0) := \left\{ e^{t_n Y_n} \dots e^{t_1 Y_1} x : Y_i \in \mathcal{F}_0, \quad t_i > 0, \quad \sum t_i = t, \quad n \geq 1 \right\} \subseteq M,$$

where for a vector field  $X \in \mathfrak{X}(M)$ ,  $e^{tX} : M \rightarrow M$  denotes the flow of diffeomorphisms associated to  $X$ . Note that the set  $\mathcal{A}_x^t(\mathcal{F}_0)$  is exactly the set of all points accessible by (5.4) at time  $t > 0$  starting from  $x \in M$  using piecewise constant controls. We then say that  $\mathcal{F}_0$  is *exactly controllable* if for every  $x$  and  $t > 0$  we have  $\mathcal{A}_x^t(\mathcal{F}_0) = M$ . The following celebrated theorem due to Stroock and Varadhan [86] links exact controllability to irreducibility.

**Theorem 5.8** (Support Theorem; [86]) *Suppose that the SDE*

$$dx_t = X_0(x_t) dt + \sum_{k=1}^r X_k(x_t) \circ dW_t^k$$

*has a globally defined flow on a manifold  $M$  and that  $\mathcal{F}_0 = X_0 + \text{span}\{X_k : k = 1, \dots, r\}$  is exactly controllable, then for each  $x \in M$  and  $t > 0$ , and every open  $O \subset M$*

$$P_t(x, O) = \mathbf{P}(x_t \in O \mid x_0 = x) > 0.$$

*In other words, exact controllability implies that the process  $(x_t)$  is topologically irreducible.*

Due to the rigidness of the time constraint  $\sum_k t_i = t$ , it is convenient to define the relaxed accessibility set of states reached by  $\mathcal{F}_0$  before time  $t$

$$\mathcal{A}_x^{\leq t}(\mathcal{F}_0) := \bigcup_{s \leq t} \mathcal{A}_x^s(\mathcal{F}_0).$$

We say that  $\mathcal{F}_0$  is *strongly controllable* if for each  $x \in M$  and  $t > 0$ ,  $\mathcal{A}_x^{\leq t}(\mathcal{F}_0) = M$ .

In order to go from strong to exact controllability, an important role is played by the *zero-time ideal*  $\mathcal{I}(\mathcal{F})$  of  $\text{Lie}(\mathcal{F})$ , defined for any collection of vector fields  $\mathcal{F} \subseteq \mathfrak{X}(M)$  by

$$\mathcal{I}(\mathcal{F}) := \text{span}\{X - Y : X, Y \in \mathcal{F}\} + \mathcal{D}(\mathcal{F}),$$

where  $\mathcal{D}(\mathcal{F})$  is the *derived algebra* ideal generated by  $\mathcal{F}$ , defined as the algebra of iterated brackets at least one bracket deep

$$\mathcal{D}(\mathcal{F}) := \text{span}\{\text{ad}(Y_r) \dots \text{ad}(Y_2)Y_1 : Y_i \in \mathcal{F}, r \geq 2\}.$$

One can check that  $\mathcal{I}(\mathcal{F})$  is a Lie algebra ideal of  $\text{Lie}(\mathcal{F})$  in the sense that if  $X \in \text{Lie}(\mathcal{F})$  and  $Y \in \mathcal{I}(\mathcal{F})$  then  $[X, Y] \in \mathcal{I}(\mathcal{F})$ . The follow results connects the zero-time ideal to exact controllability.

**Theorem 5.9** (Theorem 13 Chapter 3 [43]) *Suppose that  $\mathcal{F}$  is strongly controllable, then it is exactly controllable if for each  $x \in M$*

$$\mathcal{I}_x(\mathcal{F}) := \{X(x) : X \in \mathcal{I}(\mathcal{F})\} = T_x M.$$

For the controlled distribution of vector fields  $\mathcal{F}_0 = X_0 + \mathcal{X}$ , we have the following useful characterization of  $\mathcal{I}(\mathcal{F}_0)$  that intricately links exact controllability to the parabolic Hörmander condition. It is certainly known among experts in control theory, but we could not find a proof in the literature.

**Proposition 5.10** *If  $\mathcal{F}_0 = X_0 + \mathcal{X}$ , where  $\mathcal{X} = \text{span}\{X_1, \dots, X_r\}$ , we have*

$$\mathcal{I}(\mathcal{F}_0) = \text{Lie}(\mathcal{X}, [X_0]).$$

*Proof* First we note that

$$\text{span}\{X - Y : X, Y \in \mathcal{F}_0\} = \text{span}\{X - Y : X, Y \in \mathcal{X}\} = \mathcal{X}.$$

so that  $\mathcal{I}(\mathcal{F}_0) = \mathcal{X} + \mathcal{D}(\mathcal{F})$ . Next, we note that we can describe  $\text{Lie}(\mathcal{X}, [\mathcal{X}, X_0])$  as

$$\begin{aligned}\text{Lie}(\mathcal{X}, [\mathcal{X}, X_0]) &= \text{span}\{\text{ad}(Y_r) \dots \text{ad}(Y_1)X : Y_i \in \text{span } \mathcal{F}_0, X \in \mathcal{X}, r \in \mathbb{N}\} \\ &= \mathcal{X} + \text{span}\{\text{ad}(Y_r) \dots \text{ad}(Y_1)X : Y_i \in \text{span } \mathcal{F}_0, X \in \mathcal{X}, r \geq 1\},\end{aligned}$$

since for any  $X \in \mathcal{X}$  we can apply the Jacobi identity

$$\text{ad}([X_0, X]) = \text{ad}(X_0)\text{ad}(X) - \text{ad}(X)\text{ad}(X_0).$$

The proof is complete upon realizing that  $\mathcal{D}(\mathcal{F}_0)$  can also be described by

$$\mathcal{D}(\mathcal{F}_0) = \text{span}\{\text{ad}(Y_r) \dots \text{ad}(Y_1)X : Y_i \in \text{span } \mathcal{F}_0, X \in \mathcal{X}, r \geq 1\}$$

which follows from the fact that allowing  $X$  to take values in  $\text{span } \mathcal{F}_0 = \text{span}\{X_0, X_1, \dots, X_r\}$  instead of  $\mathcal{X}$  doesn't change the span above since  $\text{ad}(X_0)X_0 = 0$  and  $\text{ad}(X_i)X_0 = -\text{ad}(X_0)X_i$ .  $\square$

In order to prove strong controllability, we will use the notion of a Lie *saturate* originally introduced by Jurdjevic and Kupka [45]. Following [43], we introduce an equivalence class on subsets of  $\mathfrak{X}(M)$  induced by the mapping  $\mathcal{F} \mapsto \text{cl}(\mathcal{A}_x^{\leq t}(\mathcal{F}))$ , where  $\text{cl}(A)$  denotes the closure of a set  $A \subseteq M$ . Let  $\mathcal{F}, \mathcal{G} \subseteq \mathfrak{X}(M)$ , we say that  $\mathcal{F}$  and  $\mathcal{G}$  are equivalent, denoted by  $\mathcal{F} \sim \mathcal{G}$ , if for each  $x \in M$  and  $t > 0$

$$\text{cl}(\mathcal{A}_x^{\leq t}(\mathcal{F})) = \text{cl}(\mathcal{A}_x^{\leq t}(\mathcal{G})).$$

It is easy to see that if  $\mathcal{G} \sim \mathcal{F}$  then  $\mathcal{F} \sim \mathcal{F} \cup \mathcal{G}$ . This naturally motivates the definition of the *Lie saturate* of a family  $\mathcal{F}$  defined by

$$\text{LS}(\mathcal{F}) := \bigcup_{\mathcal{G} \sim \mathcal{F}} \mathcal{G} \cap \text{Lie}(\mathcal{F}).$$

Note that we always have  $\mathcal{F} \sim \text{LS}(\mathcal{F})$ .

We have the following very important theorem relating spanning properties of the Lie saturate to strong controllability.

**Theorem 5.11** (Theorem 12 Ch 3 [43])  $\mathcal{F}_0 = X_0 + \mathcal{X}$  is strongly controllable if for every  $x$

$$\text{LS}_x(\mathcal{F}_0) := \{X(x) : X \in \text{LS}(\mathcal{F}_0)\} = T_x M.$$

An important corollary of this is the following sufficient condition for exactly controllability, which will prove useful for what follows.

**Corollary 5.12** Suppose that  $X_0, \mathcal{X}$  satisfy the parabolic Hörmander condition. Then  $\mathcal{F}_0 = X_0 + \mathcal{X}$  is exactly controllable if

$$\text{Lie}(\mathcal{X}, [\mathcal{X}, X_0]) \subseteq \text{LS}(\mathcal{F}_0).$$

In light of this theorem, our goal will usually be to show that  $\text{LS}(\mathcal{F}_0)$  contains  $\text{Lie}(\mathcal{X}, [\mathcal{X}, X_0])$ . Typically this is done by repeatedly “enlarging” the initial set  $\mathcal{F}_0$  by vector fields that are equivalent to  $\mathcal{F}_0$ . The Lie saturate has the following very useful symmetry properties

**Proposition 5.13** (Proposition 2.24 [44]) *The Lie saturate  $\text{LS}(\mathcal{F})$  is convex and closed in the  $C^\infty(M)$  topology and is invariant under the following enlargements:*

1. *If  $\mathcal{F} \subseteq \text{LS}(\mathcal{F})$ , then the convex hull  $\text{co}(\mathcal{F})$  also belongs to  $\text{LS}(\mathcal{F})$ .*
2. *If  $\mathcal{V} \subseteq \text{LS}(\mathcal{F})$  is a vector space then  $\text{Lie}(\mathcal{V})$  also belongs to  $\text{LS}(\mathcal{F})$ .*
3. *If  $\pm X$  belongs to  $\text{LS}(\mathcal{F})$ , then the pushforward <sup>12</sup>  $(e^{\alpha X})_\sharp \mathcal{F}$  also belongs to  $\text{LS}(\mathcal{F})$  for each  $\alpha \in \mathbb{R}$ .*

*Remark 5.14* One can think of the pushforward  $(e^{\alpha X})_\sharp Y$  as an infinite dimensional version of the adjoint action  $e^{\alpha \text{ad}(X)}$  by the group element  $e^{\alpha X} \in \text{Diff}(M)$  on  $\mathfrak{X}(M)$  in the sense that  $(e^{\alpha X})_\sharp Y$  satisfies the following Taylor expansion

$$(e^{\alpha X})_\sharp Y = \sum_{k=0}^{\infty} \frac{\text{ad}(\alpha X)^k Y}{k!},$$

which expresses  $(e^{\alpha X})_\sharp$  as a Lie polynomial in  $\text{ad}(\alpha X)$ .

### 5.3.2 Proof of Proposition 5.7

We now want to apply this machinery to the bilinear system (1.7) to prove Proposition 5.7. In what follows, let

$$X_0(x) := B(x, x) + \epsilon A x$$

and  $\mathcal{X} := \text{span}\{X_k : k = 1 \dots, r\}$ . In light of the assumption of the parabolic Hörmander condition, Corollary 5.12, and Theorem 5.8, all we need to do is show that  $\text{Lie}(\tilde{\mathcal{X}}, [\tilde{\mathcal{X}}, \tilde{X}_0]) \subseteq \text{LS}(\tilde{\mathcal{F}}_0)$ .

<sup>12</sup> the push forward  $\phi_\sharp Y$  of a vector field  $Y$  by a diffeomorphism  $\phi$  is defined by  $(\phi_\sharp Y)(y) := d\phi_x(Y(x))$ , where  $x = \phi^{-1}(y)$ .

*Proof* Lemma C.2 shows that  $X \rightarrow \tilde{X}$  is a Lie algebra isomorphism on to  $\tilde{\mathfrak{X}}(M) = \{\tilde{X} : X \in \mathfrak{X}(M)\}$ , and so  $[\tilde{X}, \tilde{Y}]$  is the same as the lift of  $[X, Y]$ .

By property 1 of Proposition 5.13, for each  $\tilde{X} \in \tilde{\mathcal{X}}$  and  $\alpha \in [0, 1]$ , we have  $\alpha(\tilde{X}_0 + \alpha^{-1}\tilde{X}) = \alpha\tilde{X}_0 + \tilde{X} \in \text{LS}(\tilde{\mathcal{F}}_0)$ . By closedness of  $\text{LS}(\tilde{\mathcal{F}}_0)$ , it follows that for each  $\tilde{X} \in \tilde{\mathcal{X}}$

$$\tilde{X} = \lim_{\alpha \rightarrow 0} (\alpha\tilde{X}_0 + \tilde{X}) \in \text{LS}(\tilde{\mathcal{F}}_0).$$

We conclude that  $\tilde{\mathcal{X}} \in \text{LS}(\tilde{\mathcal{F}}_0)$ . By the spanning of the  $\{e_k\}_{k=1}^s$  in Assumption 4, it then follows that for all  $k = 1 \dots, s$ , the constant vector fields  $\pm \tilde{e}_k \in \text{LS}(\tilde{\mathcal{F}}_0)$ . Then, from property 3 of Proposition 5.13, and the fact  $\tilde{X}_0$  is a polynomial in  $x$  of degree two,  $\forall \alpha \in \mathbb{R}$  and  $k \in \mathcal{K}$ , the following vector field is also in  $\text{LS}(\tilde{\mathcal{F}}_0)$

$$(e^{\alpha\tilde{e}_k})_{\sharp}\tilde{X}_0 = \tilde{X}_0 + \alpha[\tilde{e}_k, \tilde{X}_0] + \frac{\alpha^2}{2}[\tilde{e}_k, [\tilde{e}_k, \tilde{X}_0]].$$

Using the Lie algebra isomorphism (Lemma C.2)  $[\tilde{e}_k, [\tilde{e}_k, \tilde{X}_0]] = [e_k, [e_k, X_0]]$  and cancellation condition of Assumption 4 that  $[e_k, [e_k, X_0]] = B(e_k, e_k) = 0$ , we deduce that  $[\tilde{e}_k, [\tilde{e}_k, \tilde{X}_0]] = 0$ . Therefore for each  $\lambda \in \mathbb{R}$

$$\lim_{\alpha \rightarrow \infty} \frac{1}{\alpha} (e^{\lambda\alpha\tilde{e}_k})_{\sharp}\tilde{X}_0 = \lambda[\tilde{e}_k, \tilde{X}_0] \in \text{LS}(\tilde{\mathcal{F}}_0).$$

It follows by property 1 of Proposition 5.13, that  $\text{span}\{\tilde{\mathcal{X}}, [\tilde{\mathcal{X}}, \tilde{X}_0]\}$  is contained in  $\text{LS}(\tilde{\mathcal{F}}_0)$  and so by property 2 of Proposition 5.13

$$\text{Lie}(\tilde{\mathcal{X}}, [\tilde{\mathcal{X}}, \tilde{X}_0]) \subseteq \text{LS}(\tilde{\mathcal{F}}_0).$$

Exact controllability now follows from Corollary 5.12 and irreducibility follows from Theorem 5.8, completing the proof of Proposition 5.7.  $\square$

*Remark 5.15* The spanning property  $\mathfrak{m}_x(\mathcal{F}) = \mathfrak{sl}(\mathbb{R}^n)$  actually implies a much stronger form of controllability for the linearized process on  $SL(\mathbb{R}^d)$ . Indeed, let

$$\hat{J}_t = \frac{D_x \Phi^t}{\det(D_x \Phi^t)^{1/n}} \in SL(\mathbb{R}^n),$$

then the normalized Jacobian Markov process is given by  $Z_t = (x_t, \hat{J}_t) \in \mathbb{R}^n \times SL(\mathbb{R}^n)$  and solves an SDE system

$$dZ_t = \hat{X}_0(Z_t)dt + \sum_{k=1}^r \hat{X}_k(Z_t) \circ dW_t^{(k)}, \quad (5.5)$$

defined by the lifts

$$\hat{X}(x, \hat{J}) = (X(x), M_X(x)\hat{J}),$$

where  $M_X(x) \in \mathfrak{sl}(\mathbb{R}^n)$  is defined above in (2.3). The assumption of Corollary 5.2 then implies that this lifted process is hypoelliptic on  $\mathbb{R}^n \times SL(\mathbb{R}^n)$ . The proof of Proposition 5.7 extends to this process analogously as to the projective process (replacing vector fields  $\tilde{X}$  with  $\hat{X}$ ), hence yielding exact controllability of the associated affine control problem as well as irreducibility of the SDE system (5.5) through Theorem 5.8.

## 6 Rigidity of invariant measures of the deterministic, projective process

In this section, we will complete the proof of Theorem C on the Lyapunov exponents of Euler-like systems and Corollary D for the L96 system. Throughout,  $\Phi_\omega^t$  is the stochastic flow of diffeomorphisms corresponding to an Euler-like SDE as in standing assumptions at the beginning of Sect. 5. Lastly, we assume that the SDE satisfies the Hörmander bracket spanning condition uniformly in  $\epsilon \in (0, 1)$ , and that the corresponding projective process  $w_t^\epsilon = (x_t^\epsilon, v_t^\epsilon)$  admits a unique stationary density  $f^\epsilon$  on  $\mathbb{SR}^n$  (see Sect. 5.3 for sufficient conditions).

The plan for this section is as follows. In Sect. 6.1 we carry out the main argument in the proof of Theorem C and Corollary D, which we reduce to checking Proposition 6.2 below. This proposition is checked in Sects. 6.2, 6.3, and 6.4.

### 6.1 Dichotomy: rigidity or positive Lyapunov exponents

Assumptions 1 and 2, as well the moment estimates needed for the application of Theorem A (see Proposition 3.2), are checked for Euler-like systems in Theorem B.1 in the Appendix. Theorem 2.4 ensures the existence of the top Lyapunov exponent  $\lambda_1$  and the sum Lyapunov exponent  $\lambda_\Sigma$ . Applying Theorem A, the Fisher information identity (3.1) reads as follows:

$$FI(f^\epsilon) = \frac{n\lambda_1^\epsilon}{\epsilon} - 2 \operatorname{tr} A \quad (6.1)$$

This is immediate from Proposition 3.2 on noting that  $\lambda_\Sigma = \epsilon \operatorname{tr} A$  by Theorem 2.4. Theorem B implies

$$\|f^\epsilon\|_{W^{s,1}(U \times \mathbb{S}^{n-1})}^2 \leq C (1 + FI(f^\epsilon)), \quad (6.2)$$

for any  $U \subset \mathbb{R}^n$  bounded, where  $s \in (0, 1)$  and  $C = C_U$  are constants independent of  $\epsilon$ .

In view of the form of (6.1) and (6.2) we see that if  $\epsilon^{-1}\lambda_1^\epsilon$  were to remain bounded as  $\epsilon \rightarrow 0$ , then  $f^\epsilon$  would be bounded in  $W^{s,1}$  uniformly in  $\epsilon$ . As  $W^{s,1}$  is locally compactly embedded in  $L^1$  (Lemma A.3), this observation leads naturally to the following alternative.

**Proposition 6.1** *At least one of the following holds:*

- (a)  $\lim_{\epsilon \rightarrow 0} \frac{\lambda_1^\epsilon}{\epsilon} = \infty$ ; or
- (b) the zero-noise flow  $(x_t^0, v_t^0)$  admits a stationary density  $f^0 \in L^1(\mathbb{R}^n \times \mathbb{S}^{n-1})$  (and moreover  $f^0 \in W_{loc}^{s,1}$ ).

*Proof* Suppose that (a) fails, i.e.

$$\liminf_{\epsilon \rightarrow 0} \frac{\lambda_1^\epsilon}{\epsilon} < \infty.$$

In this case, (6.1) implies that  $\liminf_{\epsilon \rightarrow 0} FI(f^\epsilon) < \infty$  and the hypoelliptic regularity estimate (6.2) implies  $\liminf_{\epsilon \rightarrow 0} \|f^\epsilon\|_{W^{s,1}(U)} < \infty$  for all bounded geodesic balls  $U$ . Combined with the uniform tightness of  $\{f^\epsilon\}_{\epsilon > 0}$  in (B.1) and the compactness criterion Lemma A.3, this yields precompactness in  $L^1$  of  $\{f^\epsilon\}_{\epsilon \in (0,1)}$  and so after extracting a subsequence  $\{\epsilon_j\}$ , we see that  $\exists f^0 \in L^1 \cap W_{loc}^{s,1}$  such that  $f^{\epsilon_j} \rightarrow f$  in  $L^1$ .

Let us now briefly check that  $f^0$  is an invariant density for the zero-noise flow  $(x_t^0, v_t^0)$ . For this, let  $\tilde{\mathcal{L}}^\epsilon$  denote the infinitesimal generator for  $w_t^\epsilon = (x_t^\epsilon, v_t^\epsilon)$  and let  $\phi \in C_c^\infty(\mathbb{S}\mathbb{R}^n)$  be a compactly supported test function. Starting with the Kolmogorov equation and pairing with  $\phi$  gives  $\int (\tilde{\mathcal{L}}^\epsilon \phi) f^\epsilon dq = 0$  for all  $\epsilon > 0$ , while taking  $\epsilon = \epsilon_j \rightarrow 0$  yields

$$\int (\tilde{\mathcal{L}}^0 \phi) f^0 dq = \int (\tilde{X}_0 \phi) f^0 dq = 0$$

using that  $f^{\epsilon_j} \rightarrow f^0$  in  $L^1$  while  $\tilde{\mathcal{L}}^\epsilon \phi \rightarrow \tilde{\mathcal{L}}^0 \phi$  in  $L^\infty$ . This last equality holds for all smooth compactly supported  $\phi$ , and so we conclude  $f^0$  is an invariant density for  $w_t^0 = (x_t^0, v_t^0)$ .  $\square$

In our setting, alternative (b) is ruled out by the following proposition, proved in the rest of Sect. 6.

**Proposition 6.2** *Assume that the bilinear mapping  $B$  is not identically 0. Let  $\nu$  be any invariant probability measure for  $\hat{\Phi}^t$  with the property that  $\nu(A \times \mathbb{S}^{n-1}) = \mu(A)$ , where  $\mu \ll \text{Leb}_{\mathbb{R}^n}$ . Then,  $\nu$  is singular with respect to Lebesgue measure  $\text{Leb}_{\mathbb{S}\mathbb{R}^n}$  on  $\mathbb{S}\mathbb{R}^n$ .*

Finally, we give the proof of Corollary D.

### Completing the proof of Corollary D

*Proof* To prove Corollary D we need only to verify the hypotheses of Theorem C for the L96 model (1.9). The projective hypoellipticity condition, Assumption 3, is verified in Sect. 5.2. The cancellation Assumption 4 is immediate using the canonical coordinate basis vectors of  $\mathbb{R}^J$ . Hence, the Assumptions 1 and 2 follow from Theorem B.1 in Appendix B. Therefore, all of the requirements to apply Theorem C are satisfied and the corollary follows.  $\square$

It remains to check Proposition 6.2, the proof of which occupies the remainder of this section.

## 6.2 Shearing between energy shells

In this subsection we begin the proof of Proposition 6.2. Here we show that the  $\epsilon = 0$  limit of (1.7), given by

$$\dot{x}_t = B(x_t, x_t), \quad (6.3)$$

must necessarily have some infinite-time growth in the gradient of the flow map. Some notation: let  $\Phi^t$  be the flow map for the (deterministic) ODE (6.3), for  $E > 0$  let us write  $S_E := \{x \in \mathbb{R}^n : |x|^2 = E\}$  for the “energy shells”, preserved by the flow  $\Phi^t$ , i.e.,  $\Phi^t(S_E) = S_E$  for all  $t \geq 0, E > 0$ . Write  $E(x) = |x|^2$ .

**Lemma 6.3** *Let  $x \in \mathbb{R}^n$  then the following identity holds*

$$D_x \Phi^t x = \Phi^t(x) + t B(\Phi^t(x), \Phi^t(x)). \quad (6.4)$$

Moreover, for each  $x \in \mathbb{R}^n$  and  $t \geq 0$ , we have that

$$|D_x \Phi^t| \geq t \frac{|B(\Phi^t(x), \Phi^t(x))|}{|x|}. \quad (6.5)$$

*Proof of Lemma* For a given  $\alpha > 0$ , note that the rescaled flow  $\alpha \Phi^{\alpha t}(x)$  also solves (6.3) with initial data  $\alpha x$ . Therefore by uniqueness, we have

$$\Phi^t(\alpha x) = \alpha \Phi^{\alpha t}(x) \quad (6.6)$$

Taking the derivative with respect to  $\alpha$  on both sides of (6.6) yields

$$D_{\alpha x} \Phi^t x = \Phi^{\alpha t}(x) + \alpha t B(\Phi^{\alpha t}(x), \Phi^{\alpha t}(x)).$$

Setting  $\alpha = 1$  gives (6.4). Inequality (6.5) follows from part (6.4) and the fact that  $\Phi^t(x) \cdot B(\Phi^t(x), \Phi^t(x)) = 0$  for all  $x$ , by assumption.  $\square$

### 6.3 Rigidity for invariant projective densities: review of general theory

We recall here an abstract result (Theorem 6.4) indicating that the presence of an invariant projective density as in Proposition 6.1 implies rigid properties of the corresponding flow.

In this section, we will state everything in the following *abstract linear cocycle* setting. Throughout,  $T : (X, \mathcal{B}, m) \curvearrowright$  is a (discrete-time) continuous transformation of a compact metric space  $X$ , with  $\mathcal{B}$  the Borel  $\sigma$ -algebra. Let  $A : (X, \mathcal{B}) \rightarrow SL_n(\mathbb{R})$ ,  $x \mapsto A_x$  be a measurable mapping<sup>13</sup>. This generates the *cocycle* of linear operators  $\mathcal{A} : X \times \mathbb{Z}_{\geq 0} \rightarrow SL_n(\mathbb{R})$  defined by

$$\mathcal{A}(n, x) = \mathcal{A}_x^n := A_{T^{n-1}x} A_{T^{n-2}x} \cdots A_{Tx} A_x.$$

Note that  $\mathcal{A}$  satisfies the *cocycle identity*<sup>14</sup>  $\mathcal{A}_x^{m+n} = \mathcal{A}_{T^m x}^m \mathcal{A}_x^n$  for all  $m, n \geq 0$ ,  $x \in X$ . Associated to  $T$ ,  $\mathcal{A}$  is the *projective action*  $\hat{T} : X \times \mathbb{S}^{n-1} \curvearrowright$  defined by

$$(x, v) \mapsto \left( T x, \frac{A_x v}{|A_x v|} \right), \quad x \in X, v \in \mathbb{S}^{n-1},$$

which we regard as a dynamical system on  $X \times \mathbb{S}^{n-1}$  in its own right.

Let  $\hat{m}$  be any  $\hat{T}$ -invariant measure on  $X \times \mathbb{S}^{n-1}$  projecting to  $m$  (i.e.,  $\hat{m}(K \times \mathbb{S}^{n-1}) = m(K)$  for all measurable  $K \subset X$ ), and consider its disintegration

$$d\hat{m}(u, v) = d\hat{m}_x(v) dm(x).$$

In this context, it is well-known [28, 82] that disintegrations  $(\hat{m}_x)_{x \in X}$  exist and are essentially unique (up to  $m$ -measure zero modifications) and  $x \mapsto \hat{m}_x$  is weak-\* measurably varying. Note that invariance of  $\hat{m}$  implies that

$$(A_x)_* \hat{m}_x = \hat{m}_{T x} \quad \text{for } m\text{-a.e. } x \in X,$$

where for a  $d \times d$  matrix  $A$  we write  $A_*$  for the action of  $A$  on probability measures on  $\mathbb{S}^{n-1}$ .

<sup>13</sup> Here,  $SL_n(\mathbb{R})$  is the group of  $d \times d$  real matrices of determinant 1.

<sup>14</sup> When  $T$  is a smooth mapping of a manifold and  $\mathcal{A}_x^n := D_x(T^n)$  is the so-called *derivative cocycle*, the cocycle identity is merely the chain rule for  $T^n$ .

The following result (more-or-less Theorem 3.23 of [7], up to a technical issue—see below) involves the rigidity of absolute continuity of the disintegration measures  $\hat{m}_x$  with respect to  $\text{Leb}_{\mathbb{S}^{n-1}}$ .

**Theorem 6.4** *Assume that  $\hat{m}_x \ll \text{Leb}_{\mathbb{S}^{n-1}}$  for  $m$ -almost every  $x \in X$ . Then, there exists a measurable family of inner products  $X \ni x \mapsto g_x(\cdot, \cdot)$  on  $\mathbb{R}^n$  and a  $T$ -invariant set  $\Gamma \subset X$  of full  $m$ -measure such that for all  $x \in \Gamma$  and  $v, w \in \mathbb{R}^n$ , we have that*

$$g_{Tx}(A_x v, A_x w) = g_x(v, w).$$

That is,  $A_x : (\mathbb{R}^n, g_x) \rightarrow (\mathbb{R}^n, g_{Tx})$  is an isometry.

This is slightly different from the form in Theorem 3.23 of [7]: there, it is supposed that  $\hat{m}_x \sim \text{Leb}_{\mathbb{S}^{n-1}}$ , whereas for our purposes we need the version with “ $\ll$ ”. For this reason, as well as for the sake of completeness, we sketch the proof of Theorem 6.4 here.

*Proof sketch* To start, let us assume for now that  $T : (X, \mathcal{B}, m) \curvearrowright$  is ergodic (note that we do not assume  $\hat{m}$  is ergodic). We require the following Lemma:

**Lemma 6.5** (Corollary 3.7 in [7]; Lemma 6.2 in [36]) *Assume  $(X, \mathcal{B}, m, T)$  is ergodic. Then, there is a full  $m$ -measure set of  $x_0 \in X$  with the following property: there exists a measurable mapping  $G : X \rightarrow \text{SL}_n(\mathbb{R})$ , depending on the choice of  $x_0$ , such that*

$$G(x)_* \hat{m}_{x_0} = \hat{m}_x \quad \text{for } m\text{-almost every } x \in X.$$

This version is slightly different from those appearing in [7, 36], so we briefly recall the proof below.

*Proof sketch of Lemma* Let  $\mathcal{P}(\mathbb{S}^{n-1})$  denote the space of probability measures on  $\mathbb{S}^{n-1}$  with the weak\* topology. Consider the quotient  $\mathcal{P}(\mathbb{S}^{n-1})/\text{SL}_n(\mathbb{R})$ , i.e., for  $\xi, \eta \in \mathcal{P}(\mathbb{S}^{n-1})$  we set  $\xi \sim \eta$  iff  $\exists B \in \text{SL}_n(\mathbb{R})$  so that  $B_* \xi = \eta$ . Writing  $[\eta]$  for the equivalence class of  $\eta \in \mathcal{P}(\mathbb{S}^{n-1})$ , note that  $[\hat{m}_x] = [\hat{m}_{T^k x}]$  for all  $k$ , i.e.,  $x \mapsto [\hat{m}_x]$  is constant along orbits. By Corollary 3.2.12 in [97], the Borel  $\sigma$ -algebra on the quotient space  $\mathcal{P}(\mathbb{S}^{n-1})/\text{SL}_n(\mathbb{R})$  is countably generated. Using this along with the fact that  $T : (X, m) \curvearrowright$  is ergodic, it follows from Proposition 2.1.11 in [97] that  $[\hat{m}_x]$  is constant  $m$ -almost surely. In particular, for  $m$ -a.e.  $x_0, x \in X$ , the measures  $\hat{m}_x$  and  $\hat{m}_{x_0}$  are related by the application of a matrix in  $\text{SL}_n(\mathbb{R})$ . It is now straightforward to construct a measurable selection  $G : X \rightarrow \text{SL}_n(\mathbb{R})$  as above.  $\square$

Fix  $x_0$  so that  $\hat{m}_{x_0} \ll \text{Leb}_{\mathbb{S}^{n-1}}$  and let  $G$  be as in Lemma 6.5. Observe that for any  $n \geq 0$  and  $m$ -a.e.  $x \in X$  we have that  $G(T^n x)^{-1} \mathcal{A}_x^n G(x) \in H_{x_0}$ ,

where

$$H_{x_0} := \{B \in SL_n(\mathbb{R}) : B_* \hat{m}_{x_0} = \hat{m}_{x_0}\}.$$

Observe that  $H_{x_0}$  is a subgroup of  $SL_n(\mathbb{R})$ , which we claim to be compact. If not, then a lemma of Furstenberg (see, e.g., Claim 4.8 in [17]) would imply the existence of proper subspaces  $V^1, V^2 \subset \mathbb{R}^n$  and a sequence  $\{B_n\} \subset H_{x_0}$  so that  $dist(B_n v, V^2) \rightarrow 0$  for all  $v \notin V^1$ , which would contradict  $\hat{m}_{x_0} \ll \text{Leb}_{\mathbb{S}^{n-1}}$ .

Since  $H_{x_0}$  is compact, there exists an inner product  $\langle \cdot, \cdot \rangle$  on  $\mathbb{R}^d$  with respect to which all members of  $H_{x_0}$  are isometries (Lemma 4.6 in [17]). The proof is complete on defining  $g_x$  through

$$g_x(v, w) = \langle G(x)^{-1}v, G(x)^{-1}w \rangle. \quad (6.7)$$

To handle the case when  $m$  is not ergodic, we use the ergodic decomposition [91]

$$m = \int_{\mathcal{E}_T(X)} \xi \, d\tau_m(\xi),$$

where  $\mathcal{E}_T(X)$  is the space of  $T$ -ergodic measures on  $X$  and  $\tau_m$  a Borel probability measure (w.r.t. the weak\* topology) on  $\mathcal{E}_T(X)$ . For each component  $\xi$ , we define  $\hat{\xi}$  through the formula

$$d\hat{\xi}(x, v) = d\hat{m}_x(v) d\xi(x),$$

noting that  $\hat{m}_x \ll \text{Leb}_{\mathbb{S}^{n-1}}$  for  $\xi$ -a.e.  $x \in X$  and  $\tau_m$ -a.e.  $\xi \in \mathcal{E}_T(X)$ . The proof now goes through the same as before, the only difference being that the measurable inner product (6.7) is defined along each  $\xi \in \mathcal{E}_T(X)$  one at a time.

□

## 6.4 Proof of Proposition 6.2

To start, let  $\nu$  be  $\hat{\Phi}^t$ -invariant, projecting to an absolutely continuous measure  $\mu$  on  $\mathbb{R}^n$ , and assume that

$$\nu = \nu^{ac} + \nu^\perp$$

where  $\nu^{ac} \ll \text{Leb}_{\mathbb{R}^n}$  is not identically zero (our contradiction hypothesis), while  $\nu^\perp$  is singular. Since  $\hat{\Phi}^t$  sends absolutely continuous measures to absolutely continuous measures and singular to singular, it follows that  $\nu^{ac}$  is  $\hat{\Phi}^t$ -invariant. Since  $\nu^{ac} \leq \nu$ , the measure  $\mu^{ac}(K) := \nu^{ac}(K \times \mathbb{S}^{n-1})$  satisfies  $\mu^{ac} \ll \mu \ll \text{Leb}_{\mathbb{R}^n}$  and is likewise  $\Phi^t$ -invariant. On replacing  $\nu$  with

the normalization of  $v^{ac}$ , going forward we may assume without loss that  $v \ll \text{Leb}_{\mathbb{S}^n}$ . Finally, since the energy shells  $S_E = \{|x|^2 = E\}$  are invariant, we may replace  $v$  with the normalization of its restriction to  $B(0, R) \times \mathbb{S}^{n-1}$  for some large, fixed  $R > 0$ .

Continuing, let  $\text{d}v(x, v) = \text{d}v_x(v)\text{d}\mu(x)$  denote the disintegration measures of  $v$  and note that  $v_x \ll \text{Leb}_{\mathbb{S}^{n-1}}$  for  $\mu$ -a.e.  $x$ . By Theorem 6.4, there exists a measurable family of inner products  $g_x, x \in \mathbb{R}^n$  so that

$$D_x \Phi^1 : (\mathbb{R}^n, g_x) \rightarrow (\mathbb{R}^n, g_{\Phi^1 x}) \quad (6.8)$$

is an isometry for  $\mu$ -a.e.  $x$ . By a standard procedure, we may assume that (6.8) holds for  $x \in \Gamma$ , where  $\Gamma \subset \mathbb{R}^n$  satisfies  $\mu(\Gamma) = 1$  and  $\Phi^1(\Gamma) = \Gamma$ .

For  $L > 0$ , define

$$\Gamma_L = \left\{ x \in \Gamma : L^{-1} \leq \frac{\sqrt{g_x(v, v)}}{|v|} \leq L \quad \text{for all } v \in \mathbb{S}^{n-1} \right\} \cap \left\{ x \in \Gamma : |B(x, x)| \geq L^{-1} \right\}.$$

and note that if  $x, \Phi^n x \in \Gamma_L$  for some  $n \geq 0$ , then  $|D_x \Phi^n| \leq L^2$  must hold by (6.8). Moreover, we have that  $\mu(\Gamma_L) \nearrow \mu(\Gamma) = 1$  as  $L \rightarrow \infty$ . Observe that this relies on the assumption that  $B$  is not identically 0, hence  $|B(x, x)| > 0$  Lebesgue-a.e. (here, we use the standard fact that  $\{B(x, x) = 0\}$  is a proper variety in  $\mathbb{R}^n$ , hence must have zero volume).

Fix  $L$  such that  $\mu(\Gamma_L) \geq 1/2 > 0$ . By the Poincaré Recurrence Theorem,  $\mu$ -a.e.  $x \in \Gamma_L$  visits  $\Gamma_L$  infinitely many times. Fix such an  $x \in \Gamma_L \setminus \{0\}$  and let  $0 := n_0 < n_1 < n_2 < \dots$ ,  $\lim_{\ell \rightarrow \infty} n_\ell = \infty$ , so that  $\Phi^{n_\ell}(x) \in \Gamma_L$  for all  $n_\ell$ , hence

$$|D_x \Phi^{n_\ell}| \leq L^2$$

for all such  $n_\ell$ . On the other hand, (6.5) implies

$$|D_x \Phi^{n_\ell}| \geq \frac{n_\ell}{L|x|}$$

as  $n_\ell \rightarrow \infty$ , a contradiction. This completes the proof of Proposition 6.2.

## A Basic geometric preliminaries and compactness criterion

In this appendix we summarize some basic lemmas surrounding Sobolev and Besov-type spaces on Riemannian manifolds, especially  $\mathbb{S}R^n := \mathbb{R}^n \times \mathbb{S}^{n-1}$ . None of the results here are new, but we could not locate proofs in the literature exactly matching the statements we use in the proof and so we have provided

sketches for the readers' convenience (though see [88] Chapter 7 for very similar results).

## A.1 Fractional Sobolev norms and Gagliardo–Nirenberg–Sobolev inequalities

First, we recall the definition of fractional Sobolev norms on  $(\mathcal{M}, g)$  a smooth, connected, geodesically complete,  $d$ -dimensional Riemannian manifold with bounded geometry and positive injectivity radius  $\delta_0 > 0$  (see [88] Chapter 7 for more discussion on the meaning and significance of these assumptions for function spaces). For any  $s \in (0, 1)$  and  $p \in [1, \infty)$ , define the  $W^{s,p}$  fractional Sobolev norm by

$$\|w\|_{W^{s,p}} = \|w\|_{L^p} + \left( \int_{\mathcal{M}} \int_{h \in T_x \mathcal{M}: |h| < \delta_0} \frac{|w(\exp_x h) - w(x)|^p}{|h|^{sp+d}} dh dq(x) \right)^{1/p} \quad (\text{A.1})$$

where  $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$  denotes the standard exponential map on  $\mathcal{M}$ . Note that if the manifold is geodesically complete but does not have bounded geometry and a globally positive injectivity radius, this definition can still be used locally (e.g. on closed, bounded, geodesic balls). In this work we will only need the spaces  $W^{s,1}$  and  $H^s := W^{s,2}$ .

By the aforementioned geometrical assumptions, if  $0 < \delta < \delta_0$ , then there exists a locally finite covering  $\{B_\delta(z_j)\}$  of geodesic balls and an associated smooth partition of unity  $\{\chi_j\}$ , where  $\text{supp} \chi_j \subset B_\delta(z_j)$  (see Proposition in 7.2.1 [88]). We have that the coordinate maps  $\mathbf{x}_j = \exp_{z_j} : B_\delta(0; \mathbb{R}^m) \rightarrow B_\delta(z_j; \mathcal{M})$  satisfy for each  $k \in \mathbb{Z}_+$

$$\sup_j \|\nabla^k \mathbf{x}_j\|_{L^\infty} + \sup_j \|\nabla^k \mathbf{x}_j^{-1}\|_{L^\infty} < \infty$$

and that in  $\mathbf{x}_j$  coordinates there exists a  $c > 1$  such that for each  $x \in B_\delta(0; \mathbb{R}^m)$ ,  $J_j(x) = \sqrt{|\det g|}$  satisfies  $c^{-1} \leq J_j(x) \leq c$ .

To begin, we record two Gagliardo–Nirenberg–Sobolev-type inequalities on  $\mathbb{SR}^d$  which are used in Appendix B to give qualitative estimates on stationary measures.

**Lemma A.1** *For all  $n \geq 1$ , all  $s \in (0, 1)$ , there exists a  $\theta \in (0, 1)$  such that the following holds  $\forall f \in C_c^\infty(\mathbb{SR}^n)$*

$$\|f\|_{L^2} \lesssim_{n,s} \|f\|_{L^1}^{1-\theta} \|f\|_{H^s}^\theta. \quad (\text{A.2})$$

For the homogeneous case, there exist  $\theta_1, \theta_2 \in (0, 1)$  such that the following holds for all  $f \in C_c^\infty(\mathbb{SR}^n)$

$$\|f\|_{L^2} \lesssim_n \|f\|_{L^1}^{1-\theta_1} \|(-\Delta_{x,v})^{1/2} f\|_{L^2}^{\theta_1} + \|f\|_{L^1}^{1-\theta_2} \|(-\Delta_{x,v})^{1/2} f\|_{L^2}^{\theta_2}, \quad (\text{A.3})$$

where  $\Delta_{x,v}$  denotes the (negative definite) Laplace-Beltrami operator on  $\mathbb{SR}^n$ .

*Proof* Let  $m = \dim \mathbb{SR}^n = 2n - 1$ . The inhomogeneous inequality (A.2) follows easily from the corresponding estimate on  $\mathbb{R}^m$  as the manifold  $\mathbb{SR}^n$  has uniformly bounded geometry and non-negative Ricci curvature (see e.g. [pg 301 inequality (15), [88]]). Let us briefly sketch the argument for the readers' convenience.

Define  $\tilde{f}_j := (\chi_j f) \circ \mathbf{x}_j^{-1}$ . Then, from the fact that (A.2) holds on  $\mathbb{R}^m$  and that, by Theorem 7.5.1 [88],  $\left(\sum_{j=1}^{\infty} \|\tilde{f}_j\|_{H^s(\mathbb{R}^m)}^2\right)^{1/2}$  is an equivalent norm for  $H^s(\mathcal{M})$ , we have

$$\begin{aligned} \|f\|_{L^2(\mathcal{M})}^2 &= \sum_{j=1}^{\infty} \|\tilde{f}_j\|_{L^2(\mathbb{R}^m)}^2 \lesssim \sum_{j=1}^{\infty} \|\tilde{f}_j\|_{L^1(\mathbb{R}^m)}^{2(1-\theta)} \|\tilde{f}_j\|_{H^s(\mathbb{R}^m)}^{2\theta} \\ &\leq \left( \sum_{j=1}^{\infty} \|\tilde{f}_j\|_{L^1(\mathbb{R}^m)}^2 \right)^{1-\theta} \left( \sum_{j=1}^{\infty} \|\tilde{f}_j\|_{H^s(\mathbb{R}^m)}^2 \right)^{\theta} \\ &\lesssim \|f\|_{L^1(\mathcal{M})}^{1-\theta} \left( \sum_{j=1}^{\infty} \|\tilde{f}_j\|_{L^1(\mathbb{R}^m)}^2 \right)^{1-\theta} \|f\|_{H^s(\mathcal{M})}^{2\theta} \\ &= \|f\|_{L^1(\mathcal{M})}^{2(1-\theta)} \|f\|_{H^s(\mathcal{M})}^{2\theta}, \end{aligned}$$

which is the desired result.

The homogeneous  $\|(-\Delta_{x,v})^{-1} f\|$  norms in (A.3) require a more intrinsic treatment. Thus, we use heat semigroup methods, commonly used to treat such inequalities on manifolds; see e.g. [57]. Denote  $e^{t\Delta}$  the heat semigroup associated to the Laplace-Beltrami operator (omitting the  $x, v$  subscript for simplicity). By the Li-Yau inequality [61] and the fact that  $\mathbb{SR}^n$  has non-negative Ricci curvature, the following pointwise upper bound holds for  $e^{t\Delta} f$

$$|e^{t\Delta} f(z)| \lesssim \int_{\mathbb{SR}^d} \frac{1}{\text{vol}B(z', \sqrt{t})} e^{-cd(z, z')^2/t} |f(z')| \, dq(z'),$$

where  $B(z', \sqrt{t})$  is the geodesic ball of radius  $\sqrt{t}$  centered at  $z$  and  $d(z, z')$  denotes geodesic distance and  $0 < c < 1/4$  is some constant. Note that for  $t <$

1 we have  $\text{vol}B(z', \sqrt{t}) \approx t^{-m/2}$  whereas for  $t > 1$  we have  $\text{vol}B(z', \sqrt{t}) \approx t^{-n/2}$ . From this, we obtain the following  $L^2$  regularization estimate

$$\|e^{t\Delta} f\|_{L^2} \lesssim \frac{1}{(t^{-n/2} + t^{-m/2})^{1/2}} \|f\|_{L^1} \approx \frac{1}{(t^{-n/4} + t^{-m/4})} \|f\|_{L^1}. \quad (\text{A.4})$$

Next, observe that

$$\|e^{t\Delta} f - f\|_{L^2} = \left\| \int_0^t \Delta e^{\tau\Delta} f d\tau \right\|_{L^2} \leq \int_0^t \|\Delta e^{\tau\Delta} f\|_{L^2} d\tau \lesssim t^{1/2} \|(-\Delta)^{1/2} f\|_{L^2}.$$

Let  $g = f / \|f\|_{L^1}$ . Using the heat semigroup to mollify  $g$ , we obtain from (A.4)

$$\|g\|_{L^2} \leq \|e^{t\Delta} g - g\|_{L^2} + \|e^{t\Delta} g\|_{L^2} \lesssim t^{1/2} \|(-\Delta)^{1/2} g\|_{L^2} + \frac{1}{(t^{-n/4} + t^{-m/4})}.$$

Choosing  $t = \|(-\Delta)^{1/2} g\|_{L^2}^{-\gamma}$  yields

$$\|g\|_{L^2} \leq \|(-\Delta)^{1/2} g\|_{L^2}^{1-\gamma/2} + \|(-\Delta)^{-1/2} g\|_{L^2}^{n\gamma/4} + \|(-\Delta)^{1/2} g\|_{H^s}^{m\gamma/4}.$$

As  $m > n$ , it makes sense to set  $1 - \gamma/2 = m\gamma/4 =: \theta_1$  to obtain  $\gamma = (\frac{1}{2} + \frac{m}{4})^{-1}$  so that  $\theta_1 \in (0, 1)$  and  $n\gamma/4 =: \theta_2 \in (0, 1)$  as well. Hence we have

$$\|g\|_{L^2} \lesssim \|(-\Delta)^{1/2} g\|_{L^2}^{\theta_1} + \|(-\Delta)^{1/2} g\|_{L^2}^{\theta_2}.$$

Inequality (A.2) follows on recalling that  $g = f / \|f\|_{L^1}$ .  $\square$

## A.2 Local embeddings and compactness in $L^1$

For Theorem 4.2, we need to prove the following embedding of the Besov-type norm  $\|\cdot\|_{\Lambda^s}$  (defined in (4.2)) into  $W^{s,1}$  locally. In the remainder of this section, we will assume that  $(\mathcal{M}, g)$  is a geodesically complete, connected smooth Riemannian manifold. Since we will be working locally, we do not need any assumptions about bounded geometry.

**Lemma A.2** *Let  $U \subset \mathcal{M}$  be any bounded, open geodesic ball and fix a suitable atlas of  $U$ ,  $\{\mathbf{x}_j\}_{j=1}^N$ , as above to define the norm (4.2). For all  $0 < s' < s < 1$  there holds the following  $\forall w \in C_c^\infty(U)$*

$$\|w\|_{W^{s',1}} \lesssim_U \|w\|_{\Lambda^s}.$$

*Proof* First, note that we can restrict the integrals and suprema in the norms to be over sufficiently small sets depending on the local geometry. Let  $B_R(z_0; \mathcal{M}) = U$  and consider the canonical basis for the tangent space  $T_{z_0} \mathcal{M}$ ,  $\{\partial_j\}$  defined by the pullback of the canonical directions in  $\mathbb{R}^d$  under the exponential map  $\exp_{z_0}$ . Using geodesics, this basis can be parallel transported to a full set of vector fields  $\{Z_k\}_{k=1}^d$ , over  $B(z_0, \delta)$  for some  $\delta > 0$  smaller than the injectivity radius and that this set of vector fields forms a basis for the tangent space at every point. Then we can write (for some  $\delta' > 0$  possibly smaller than  $\delta$ ),

$$\|w\|_{W^{s',1}} \lesssim \|w\|_{L^1} + \int_{\mathcal{M}} \int_{c \in \mathbb{R}^d : |c| \leq \delta'} \frac{|w(\exp_x \sum_j c_j Z_j(x)) - w(x)|}{|c|^{d+s'}} dc dq(x).$$

Note that since  $\sum_j |Z_j(x)| \approx 1$ , we have

$$d \left( \exp_x \sum_j c_j Z_j(x), x \right) = \left| \sum_j c_j Z_j(x) \right| \approx |c|,$$

where  $d(\cdot, \cdot)$  here refers to the geodesic distance on  $\mathcal{M}$ . The key point here is that we have re-written the integral over the unit ball in the tangent spaces  $T_z \mathcal{M}$  into a ball in  $\mathbb{R}^d$  which is the same over all  $x \in B(x_0, \delta)$  (more accurately we have estimated the previous integral from above by this quantity plus the  $L^1$  norm). Therefore, we can apply Fubini and conclude that for  $s' < s$ ,

$$\begin{aligned} \|w\|_{W^{s',1}} &\lesssim \|w\|_{L^1} + \int_{c \in \mathbb{R}^d : |c| \leq \delta'} \int_{\mathcal{M}} \frac{|w(\exp_x \sum_j c_j Z_j(x)) - w(x)|}{|c|^{d+s'}} dq(x) dc \\ &\lesssim \|w\|_{L^1} + \sup_{c \in \mathbb{R}^d : |c| \leq \delta'} \int_{\mathcal{M}} \frac{|w(\exp_x \sum_j c_j Z_j(x)) - w(x)|}{|c|^s} dq(x). \end{aligned}$$

It is not hard to see that by writing the integral in the coordinate parameterization used to define  $\Lambda^s$ , the right hand side is bounded above by the  $\Lambda^s$  norm.  $\square$

Next, we give a proof that local  $W^{s,1}$  Sobolev smoothness plus a tightness condition estimate yields precompactness in  $L^1$ . While this type of result is very standard in functional analysis, we could not find a reference that gives the proof in the form we need. A proof is provided for the reader's convenience using a standard compactness criterion in  $L^1$  on metric spaces.

**Lemma A.3** *Consider a bounded sequence  $\{f_n\}_{n=1}^{\infty} \subset L^1(\mathcal{M})$  that satisfies the following properties:*

1. (Local uniform regularity) For every geodesic ball  $B = B_R(x; \mathcal{M})$ , one has

$$\sup_n \|\chi_B f_n\|_{W^{s,1}} < \infty,$$

where  $\chi_B$  is a smooth cut-off function equal to 1 inside  $B_R(x)$  and compactly supported in  $B_{2R}(x)$ .

2. (Tightness) There exists an  $x_0 \in \mathcal{M}$  such that

$$\lim_{R \rightarrow \infty} \sup_n \|f_n\|_{L^1(\mathcal{M} \setminus B_R(x_0))} = 0.$$

Then the sequence  $\{f_n\}_{n=1}^\infty$  is strongly precompact in  $L^1(\mathcal{M})$  with limit points in  $L^1 \cap W_{loc}^{s,1}$ .

*Proof* To start, we will use a general Fréchet-Kolmogorov compactness criterion in  $L^1$  on bounded metric spaces due to Krotov [52]. We state it below for convenience.

**Theorem A.4** ([52] Theorem 5) *Let  $(X, d)$  be a bounded complete metric space with a finite measure  $\mu$  satisfying the doubling condition, that is, there exists a constant  $c_\mu > 0$  such that*

$$\mu(B_{2r}(x)) \leq c_\mu \mu(B_r(x)), \quad x \in X, \quad r > 0.$$

Let  $\mathcal{S}$  be a bounded subset of  $L^1(X)$ , then  $\mathcal{S}$  is precompact in  $L^1(X)$  if and only if the following condition is satisfied:

$$\lim_{r \rightarrow 0} \sup_{f \in \mathcal{S}} \int_X \left( \int_{B_r(x)} |f(x) - f(y)|^{\frac{1}{2}} d\mu(y) \right)^2 d\mu(x) = 0.$$

Note that the volume measure on any  $n$ -dimensional compact Riemannian manifold satisfies  $\text{vol } B_r(x) \approx r^n$  and is therefore a doubling measure on the complete metric space with metric given by the usual geodesic distance. In what follows fix  $x_0$  as above and let  $U_R := \overline{B_R(x_0)}$  be a closed geodesic ball for a fixed  $R > 0$ . By the Hopf-Rinow theorem (since  $\mathcal{M}$  is complete)  $U_R$  is itself a compact manifold and a bounded complete metric space with the geodesic metric  $d$  inherited from  $\mathcal{M}$ . We denote  $\hat{B}_r(x) = B_r(x) \cap U_R$  the metric ball on this space. It follows by Cauchy-Schwarz and the fact that for  $r$  small enough  $\text{vol } \hat{B}(x, r) \gtrsim r^d$  gives

$$\left( \int_{\hat{B}_r(x)} |f(x) - f(y)|^{\frac{1}{2}} dq(y) \right)^2 \lesssim \left( \int_{\hat{B}_r(x)} \frac{|f(x) - f(y)|}{d(x, y)^{d+s}} dq(y) \right) r^s.$$

Taking  $r > 0$  smaller than the injectivity radius of  $U_R$  and changing coordinates using the inverse exponential map  $\exp_x^{-1}$  and using that  $d(x, \exp_x h) = |h|$  gives

$$\int_{\hat{B}_r(x)} \frac{|f(x) - f(y)|}{d(x, y)^{d+s}} dq(y) = \int_{h \in T_x \mathcal{M}: |h| < r} \frac{|w(\exp_x h) - w(x)|}{|h|^{d+s}} J_x(h) dh,$$

where  $J_x$  is the Jacobian factor that describes the volume measure on  $U_R$ . Since  $U_R$  is compact  $J_x(h)$  is bounded and therefore we have

$$\int_{U_R} \left( \int_{\hat{B}_r(x)} |f(x) - f(y)|^{\frac{1}{2}} dq(y) \right)^2 dq(x) \lesssim (\|\chi_{U_{2R}} f\|_{W^{s,1}}) r^s.$$

This implies that for each fixed  $j$ , the set  $\{\chi_{U_{2j}} f_n\}_n$  is a precompact set in  $L^1(\mathcal{M})$  for all  $j$ . Diagonalization gives a limit  $f \in L^1$  and a subsequence  $\{f_{n_k}\}$  that converges in  $L^1$  on every geodesic ball. Now, for all  $\epsilon > 0$ ,

$$\|f_{n_k} - f\|_{L^1} \leq \|f_{n_k} - f\|_{L^1(U_R)} + \|f_{n_k}\|_{L^1(\mathcal{M} \setminus U_R)} + \|f\|_{L^1(\mathcal{M} \setminus U_R)}.$$

By tightness, we can choose the last  $R > 0$  sufficiently large such that the last two terms are together less than  $\epsilon/2$  and by the convergence on all compacts, choose  $n_k$  large enough so that the first term is  $\epsilon/2$ .

The fact that  $f \in W_{\text{loc}}^{s,1}$  follows from lower semi-continuity of the  $W_{\text{loc}}^{s,1}$  semi-norms for  $s \in (0, 1)$  with respect to  $L^1$  convergence (this can be easily be deduced by choosing a further sub-sequence so that  $f_{n_k} \rightarrow f$  almost surely and applying Fatou's lemma to the double integral of the difference  $|f_{n_k}(x) - f_{n_k}(y)|/d(x, y)^{s+d}$ ).  $\square$

## B Qualitative properties of the projective stationary measure

In this section we record basic properties of the SDE (1.7).

**Theorem B.1** *Consider the Euler-like model (1.7). Suppose that  $B$  satisfies Assumption 4 and that the vector fields  $\{\tilde{X}_0^\epsilon, \tilde{X}_1, \dots, \tilde{X}_r\}$  satisfies the uniform parabolic Hörmander condition on  $\mathbb{SR}^n$  as in Definition 4.1 (Assumption 3). Then  $\forall \epsilon > 0$ , the SDE (1.7) satisfies Assumptions 1 and 2. Moreover, the stationary measure of the  $(w_t)$  process  $f^\epsilon$  has a smooth density with respect to Lebesgue measure  $f^\epsilon \in L^1 \cap L^2 \cap C^\infty$  with  $f^\epsilon \log f^\epsilon \in L^1$ , and  $\exists C, \gamma > 0$  such that  $\forall \epsilon \in (0, 1]$ ,*

$$\int_{\mathbb{SR}^n} f^\epsilon e^{\gamma|x|^2} dq < C. \quad (\text{B.1})$$

and  $\forall N > 0$  the following moment bound holds  $\forall \epsilon \in (0, 1]$  (not uniformly in  $N$  or  $\epsilon$  of course)

$$\int_{\mathbb{S}\mathbb{R}^n} \langle x \rangle^N f^\epsilon \log f^\epsilon \, dq < \infty. \quad (\text{B.2})$$

Furthermore, the estimate in Assumption 1 (iii) holds for all  $\epsilon > 0$ .

*Proof of Theorem B.1* Claims (i) and (ii) of Assumption 1 are standard or proved in [18]. The proof of Assumption 1 (iii) follows by providing suitable moment estimates on  $\log |\det D\Phi^t|$  and  $\log |D\Phi^t|$  using the SDE derived in the proof of Proposition C.4. Indeed, denoting  $J_t = D\Phi^t(x)\xi$  we have

$$\frac{d}{dt} J_t = B(x, J_t) + B(J_t, x_t) + \epsilon A J_t,$$

and hence,  $\exists C > 0$ , such that

$$\|J_t\| \leq \|\xi\| \exp \left( C \int_0^t \|x_\tau\| \, d\tau \right).$$

By time-reversal of ODEs,  $V_t = J_t^{-1}$  satisfies the analogous estimate, possibly by adjusting  $C$  (the growth is due to the damping)

$$\|J_t^{-1}\| \leq \|\xi\| \exp \left( C \int_0^t (\epsilon + \|x_\tau\|) \, d\tau \right).$$

Therefore,

$$\begin{aligned} \mathbf{E} \int_{\mathbb{R}^n} [\log^+ |D_x \Phi^t| + \log^+ |(D_x \Phi^t)^{-1}|] \, d\mu(x) &\lesssim \mathbf{E} \int_{\mathbb{R}^n} \left[ \epsilon t + \int_0^t \|x_\tau\| \, d\tau \right] \, d\mu(x) \\ &= t \left( \epsilon + \int_{\mathbb{R}^n} \|x\| \, d\mu(x) \right), \end{aligned}$$

which is finite by standard moment estimates (see e.g. the drift condition (B.4) below). This completes the proof of Assumption 1.

The results of Assumption 2 follow from similar methods with the exception of the uniqueness of the stationary measure. By the parabolic Hörmander's condition on  $\{\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_r\}$  together with Hörmander's theorem [39, 41] and the Doob-Khasminskii theorem [30], it is sufficient to verify irreducibility as in Definition 2.3. This follows by Proposition 5.7. We are not aware of any existing results that directly imply  $f^\epsilon \in L^2$  or  $f^\epsilon \log f^\epsilon \in L^1$  in the literature and we therefore include the proof. For this we use some ideas that appear in

[18], in which similar estimates are proved for the density of the base process  $\rho^\epsilon$ . As in [18], a convenient method to justify many formal calculations begins by first regularizing the problem by adding elliptic Brownian motions. Recall that the generator  $\tilde{\mathcal{L}}$  for the projective process  $(w_t)$  is given by

$$\tilde{\mathcal{L}} = \tilde{X}_0 + \frac{1}{2} \sum_{k=1}^r \tilde{X}_k^2.$$

We then regularize this by the perturbing the generator

$$\tilde{\mathcal{L}}^\delta = \tilde{\mathcal{L}} + \delta \Delta_{x,v},$$

where  $\Delta_{x,v} = \Delta_x + \Delta_v$  with  $\Delta_x$  the usual Laplacian on  $\mathbb{R}^n$  and  $\Delta_v$  the Laplace-Beltrami operator on  $\mathbb{S}^{n-1}$ . This corresponds to perturbing the SDE (1.7) by a non-degenerate  $\sqrt{\delta}$  Brownian motion on  $\mathbb{S}\mathbb{R}^n$ . It is not hard to show that  $\tilde{\mathcal{L}}^\delta$  satisfies a drift condition<sup>15</sup>

$$\tilde{\mathcal{L}}^\delta e^{\gamma|x|^2} \leq -\epsilon \alpha e^{\gamma|x|^2} + \epsilon K,$$

for some  $\alpha \in (0, 1)$ ,  $K \geq 1$  (uniformly in  $\epsilon, \delta$ ). This gives rise to a globally defined Markov process  $(w_t^\delta)$ . Moreover for a given initial density  $f \in C_c^\infty(\mathbb{S}\mathbb{R}^d)$  with  $\int f dq = 1$  and  $f \geq 0$ , such that  $\text{Law}(w_0^\delta) = f$  we denote  $f_t = \text{Law}(w_t^\delta)$ , which solves the forward Kolmogorov equation

$$\partial_t f_t = (\tilde{\mathcal{L}}^\delta)^* f_t + \delta \Delta_{x,v} f_t. \quad (\text{B.3})$$

From the drift condition we have,  $\forall \gamma$  sufficiently small,  $\exists \alpha \in (0, 1)$  such that (uniformly in  $\epsilon, \delta$ ),

$$\int_{\mathbb{S}\mathbb{R}^d} f_t e^{\gamma|x|^2} dq \lesssim 1 + e^{-\alpha t} \int_{\mathbb{S}\mathbb{R}^d} f e^{\gamma|x|^2} dq. \quad (\text{B.4})$$

Let  $\bar{\chi} \in C_c^\infty(B(0, 1))$  with  $0 \leq \bar{\chi} \leq 1$ , and  $\bar{\chi} = 1$  for  $|x| \leq 1/2$  and define  $\chi(x) = \bar{\chi}(x/2) - \bar{\chi}(x)$ . Define  $\chi_j = \chi(2^{-j}x)$ , which defines the partition of unity  $1 = \bar{\chi} + \sum_{j=0}^\infty \chi_j(x)$ . From energy estimates on (B.3) we have the following,

<sup>15</sup> The lifted part of the vector fields vanish due to the lack of  $v$  dependence, and so this reduces to the same drift condition for the base process. This follows immediately noting that since  $x \cdot B = 0$  we have  $X_0 e^{\gamma|x|^2} = 0$ , that  $|\partial_{x_j x_j} e^{\gamma|x|^2}| \lesssim (\gamma^2 |x|^2 + \gamma) e^{\gamma|x|^2}$ , and by negative-definiteness,  $Ax \cdot \nabla e^{\gamma|x|^2} \lesssim -\gamma |x|^2 e^{\gamma|x|^2}$ . Hence, the condition follows for  $\gamma$  chosen sufficiently small.

$$\frac{d}{dt} \|f_t\|_{L^2}^2 + \delta \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^2 \lesssim \|(1+|x|) f_t\|_{L^2}^2 \lesssim \|\tilde{\chi} f_t\|_{L^2}^2 + \sum_{j=1}^{\infty} 2^{2j} \|\chi_j f_t\|_{L^2}^2,$$

(in order to justify such estimates one may apply smooth,  $v$ -independent radially symmetric cut-offs to the nonlinearity and pass to the limit). By the Gagliardo–Nirenberg–Sobolev-type inequality (A.3) (Lemma A.1), for  $\theta_1, \theta_2 \in (0, 1)$  as therein, there holds (recall that  $\|f_t\|_{L^1} = 1$ )

$$\begin{aligned} \frac{d}{dt} \|f_t\|_{L^2}^2 + \delta \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^2 &\lesssim \|\tilde{\chi} f_t\|_{L^2}^2 + \sum_{j=1}^{\infty} 2^{2j} \|\chi_j f_t\|_{L^2}^2 \\ &\lesssim \|(-\Delta_{x,v})^{1/2} \tilde{\chi} f_t\|_{L^2}^{2\theta_1} \|\tilde{\chi} f_t\|_{L^1}^{1-2\theta_1} + \sum_{j=1}^{\infty} 2^{2j} \|(-\Delta_{x,v})^{1/2} \chi_j f_t\|_{L^2}^{2\theta_1} \|\chi_j f_t\|_{L^1}^{1-2\theta_1} \\ &\quad + \|(-\Delta_{x,v})^{1/2} \tilde{\chi} f_t\|_{L^2}^{2\theta_2} \|\tilde{\chi} f_t\|_{L^1}^{1-2\theta_2} + \sum_{j=1}^{\infty} 2^{2j} \|(-\Delta_{x,v})^{1/2} \chi_j f_t\|_{L^2}^{2\theta_2} \|\chi_j f_t\|_{L^1}^{1-2\theta_2} \\ &\lesssim \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_1} + \|\nabla_x \tilde{\chi} f_t\|_{L^2}^{2\theta_1} + \sum_{j=1}^{\infty} 2^{2j} \left( \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_1} + \|\nabla_x \chi_j f_t\|_{L^2}^{2\theta_1} \right) \|\chi_j f_t\|_{L^1}^{1-2\theta_1} \\ &\quad + \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_2} + \|\nabla_x \tilde{\chi} f_t\|_{L^2}^{2\theta_2} + \sum_{j=1}^{\infty} 2^{2j} \left( \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_2} + \|\nabla_x \chi_j f_t\|_{L^2}^{2\theta_2} \right) \|\chi_j f_t\|_{L^1}^{1-2\theta_2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{d}{dt} \|f_t\|_{L^2}^2 + \delta \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^2 &\lesssim_{\delta} \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_1} \\ &\quad + \|(-\Delta_{x,v})^{1/2} f_t\|_{L^2}^{2\theta_2} + \int_{\mathbb{SR}^d} f_t e^{\gamma|x|^2} dq. \end{aligned}$$

Hence, from (B.4), there holds for some  $q > 2$ ,

$$\frac{1}{t} \int_0^t \|(-\Delta_{x,v})^{1/2} f_{\tau}\|_{L^2}^2 d\tau \lesssim \delta^{-q} \int_{\mathbb{SR}^d} f e^{\gamma|x|^2} dq.$$

Combined with the uniform drift condition, this allows to pass to the limit  $t \rightarrow \infty$  and conclude that the density of the unique stationary measure, denoted below as  $f^{\epsilon, \delta}$  is in  $H^1(\mathbb{SR}^d)$  (recall (A.2)); we note that  $f^{\epsilon, \delta}$  is a smooth solution of the Kolmogorov equation

$$(\tilde{\mathcal{L}}^* + \delta \Delta_{x,v}) f^{\epsilon, \delta} = 0. \quad (\text{B.5})$$

Next, we obtain an  $L^2$  estimate that is uniform in  $\delta$  in order to pass to the  $\delta \rightarrow 0$  limit. For this, we clearly need to depend on hypoelliptic regularity. Define the regularized Hörmander norm pair (see discussions in [4, 18, 41] for

motivations),

$$\begin{aligned}\|w\|_{\mathcal{H}_\delta} &:= \|w\|_{L^2} + \sum_{k=1}^r \|X_k w\|_{L^2} + \delta \|(-\Delta_{x,v})^{1/2} w\|_{L^2} \\ \|w\|_{\mathcal{H}_\delta^*} &:= \sup_{\varphi: \|\varphi\|_{\mathcal{H}_\delta} \leq 1} \left| \int_{\mathbb{SR}^n} (\tilde{X}_0 \varphi) w \, dq \right|\end{aligned}$$

The proof is similar to [Lemma 2.3; [18]] provided we have the following quantification of Hörmander's inequality.

**Lemma B.2** (Quantitative Hörmander inequality for the projective process)  
*Suppose that  $\{\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_r\}$  satisfies the uniform parabolic Hörmander condition on  $B(0, 2) \times \mathbb{S}^{n-1}$  as in Definition 4.1. There exists  $s > 0$  and  $q > 0$ , such that for any  $R \geq 1$ ,  $w \in C_c^\infty(B_R \times \mathbb{S}^{n-1})$  and  $\delta \in [0, 1]$  there holds*

$$\|w\|_{H^s} \lesssim R^q (\|w\|_{\mathcal{H}_\delta} + \|w\|_{\mathcal{H}_\delta^*}), \quad (\text{B.6})$$

where both  $s > 0$  and the implicit constant do not depend on  $\epsilon$ ,  $\delta$ , or  $R$ , and  $H^s = W^{s,2}$ .

*Proof* The proof begins with a re-scaling as in [Lemma 3.2; [18]]. Define  $h(x, v) = w(Rx, v)$  which solves a PDE of the following form for suitable vector fields  $N, V, Y$ ,

$$\epsilon \delta \Delta_{x,v} h + \frac{1}{2} \sum_{j=1}^r \epsilon (\tilde{X}_j^*)^2 h - Nh + R^{-1} V^* h - \frac{\epsilon}{R} Y^* h = 0.$$

where  $N(x) = B(x, x)$ ,  $Y(x) = Ax$  and  $V(x, v) = \Pi_v \nabla F(x)v$ , and their action on  $h$  is interpreted as a differential operator. We see that the proof here is more subtle than in the corresponding [Lemma 3.2; [18]] as  $R^{-1}V$  is required to span the directions in projective space. From Proposition 2.7, we see that the spanning in  $x$  and  $v$  can be considered essentially separately, first choosing brackets to span in  $x$  and then correcting by choosing suitable brackets in  $\mathfrak{m}_x(X_0; X_1, \dots, X_r)$  to span in  $v$ . Using this structure we see that given a vector field  $Z \in \mathfrak{X}(\mathbb{SR}^n)$  and  $q_0 \in B(0, 1) \times \mathbb{S}^{n-1}$ , there exists  $p_j < \dots < p_2 < p_1 \leq k$  (with  $k$  as in Definition 4.1) such that for  $q$  in a neighborhood of  $q_0$ , there are finitely many smooth coefficients  $c_j$  and vectors  $Z_j \in \mathcal{X}_k$  with

$$Z(q) = \sum_j R^{p_j} c_j(q) Z_j(q),$$

where if  $Z$  varies in a bounded set in  $C^m$ , then  $\{c_j\}_j$  varies in a similarly bounded set as well. A careful reading of [41] shows that this introduces powers of  $R$  matching the powers of  $t$  into all of the estimates in [Sects. 4 and 5; [41]], the maximal power arising being  $R^k$ . In particular, the error estimates come in the form  $\mathcal{O}(R^{k/\sigma})$ , provided that  $R^k t < 1$  and  $0 < \sigma < s^*$  as in [41]. This restriction on  $t$  in the estimates further introduces only polynomial dependence on  $R$ , as for any  $Z \in \mathfrak{X}(\mathbb{SR}^n)$ ,

$$\sup_{|t| \leq 1} |t|^{-\sigma} \|e^{tZ} g - g\|_{L^2} \lesssim R^{k\sigma} \|g\|_{L^2} + \sup_{|t| \leq R^{-k}} |t|^{-\sigma} \|e^{tZ} g - g\|_{L^2}.$$

Combining the above observations with those of [41] implies that the constant in (B.6) remains polynomial in  $R$  (exponential would also be sufficient for our purposes, as we only use that the constant is bounded above by  $e^{\eta R^2}$  for  $\eta < \gamma$ ).  $\square$

Once one has Lemma B.2, the proof of Theorem B.1 follows easily, given that we are not seeking  $\epsilon$ -independent bounds, as these such bounds will be false for all but the most degenerate models (see [Lemma 2.4; [18]] for the corresponding argument on  $\rho^\epsilon$ , which does yield  $\epsilon$ -independent estimates). Let  $\bar{\chi} \in C_c^\infty(B(0, 1))$  with  $0 \leq \bar{\chi} \leq 1$ , and  $\bar{\chi} = 1$  for  $x \leq 1/2$  and define  $\chi(x) = \chi(x/2) - \bar{\chi}(x)$ . Define  $\chi_j = \chi(2^{-j}x)$ , which defines the partition of unity  $1 = \bar{\chi} + \sum_{j=0}^\infty \chi_j(x)$ .

We now obtain a uniform-in- $\delta$   $L^2$  estimate. By Lemma B.2 and the Gagliardo–Nirenberg–Sobolev-type estimate (A.2) (recall  $f^{\epsilon, \delta}$  is a probability measure), there holds

$$\|f^{\epsilon, \delta}\|_{L^2} \lesssim \|\bar{\chi} f^{\epsilon, \delta}\|_{H_{\text{Hyp}, \delta}^1}^{1-\theta} + \sum_{j=1}^\infty 2^{jq(1-\theta)} \|\chi_j f^{\epsilon, \delta}\|_{L^1}^\theta \|\chi_j f^{\epsilon, \delta}\|_{H_{\text{Hyp}, \delta}^1}^{1-\theta},$$

where we have denoted  $\|\cdot\|_{H_{\text{Hyp}, \delta}^1} = \|\cdot\|_{\mathcal{H}_\delta} + \|\cdot\|_{\mathcal{H}_\delta^*}$ . Pairing (B.5) with  $\bar{\chi} f^{\epsilon, \delta}$  and  $\chi_j f^{\epsilon, \delta}$  followed by standard manipulations gives

$$\|\bar{\chi} f^{\epsilon, \delta}\|_{H_{\text{Hyp}, \delta}^1} + \sup_j \|\chi_j f^{\epsilon, \delta}\|_{H_{\text{Hyp}, \delta}^1} \lesssim_\epsilon \|f^{\epsilon, \delta}\|_{L^2}.$$

Therefore, we have

$$\|f^{\epsilon, \delta}\|_{L^2} \lesssim \|f^{\epsilon, \delta}\|_{L^2}^\theta + \sum_{j=1}^\infty 2^{jq(1-\theta)} \|\chi_j f^{\epsilon, \delta}\|_{L^1}^\theta \|f^{\epsilon, \delta}\|_{L^2}^\theta,$$

which implies that for  $\theta_1, \theta_2 \in (0, 1)$  we have a uniform-in- $\delta$  estimate on the  $L^2$  norm. Note that the estimate still depends badly on  $\epsilon$ . Passing to the  $\delta \rightarrow 0$  limit shows that  $f^\epsilon \in L^2$  for each  $\epsilon > 0$ .

Finally, observe that  $\langle x \rangle^N f^\epsilon \log f^\epsilon \in L^1$  for all  $N > 0$  indeed,

$$\begin{aligned} \int_{\mathbb{S}\mathbb{R}^n} \langle x \rangle^N f^\epsilon |\log f^\epsilon| \, dq &\lesssim \int_{\mathbb{S}\mathbb{R}^n} \langle x \rangle^N \left( \sqrt{f^\epsilon} + (f^\epsilon)^{3/2} \right) \, dq \\ &\lesssim \|f^\epsilon e^{\gamma|x|^2}\|_{L^1} + \|f^\epsilon e^{\gamma|x|^2}\|_{L^1}^{1/4} \|f^\epsilon\|_{L^2}^{3/4}. \end{aligned}$$

Note that in fact, one can obtain similar moment estimates also on  $L^2$ . This completes the proof of Theorem B.1.  $\square$

## C Properties of projective lifts and the Furstenberg–Khasminskii formula

### C.1 Sufficient conditions for projective spanning: Proof of Proposition 2.7

In this section, we give a proof of Proposition 2.7 that characterizes when a collection of lifted vector fields  $\{\tilde{X}_k\}_{k=0}^r$  satisfy the Hörmander condition on  $\mathbb{S}M$  in terms of transitivity of the matrix lie algebra  $\mathfrak{m}_x(X_0, \dots, X_r)$  (defined in (2.4)).

Before we prove Proposition 2.7, we will need some preliminary results. As we will be taking commutators of the above vector fields, it is important to record how projective vector fields behave under the Lie bracket.

**Lemma C.1** *Let  $A, B \in \mathfrak{sl}(\mathbb{R}^n)$ , then the following identity holds*

$$[V_A, V_B](v) = -V_{[A, B]}(v),$$

where  $[A, B] := AB - BA$  denotes the usual commutator on linear operators.

Proof. Let  $\nabla$  denote the Levi-Civita connection on  $\mathbb{S}^{n-1}$ , then since  $\nabla$  is torsion-free, we have the following formula for the Lie bracket in terms of the covariant derivative

$$[V_A, V_B] = \nabla_{V_A} V_B - \nabla_{V_B} V_A.$$

Recall from the proof of Lemma C.3 that using the embedding of  $\mathbb{S}^{n-1}$  into  $\mathbb{R}^n$ , we have the following formula for the total covariant derivative of  $V_A$  (viewed as a linear operator on  $T_v \mathbb{S}^{n-1}$ )

$$\nabla V_A(v) = \Pi_v A - \langle v, Av \rangle I.$$

It follows that

$$\begin{aligned}(v) &= \nabla V_B(v)V_A(v) - \nabla V_A(v)V_B(v) \\ &= \Pi_v B \Pi_v A v - \Pi_v A \Pi_v B v - \langle v, Bv \rangle V_A(v) + \langle v, Av \rangle V_B(v).\end{aligned}$$

Using the fact that  $\Pi_v u = u - \langle u, v \rangle v$  for  $u \in T_x M$ , we find

$$\Pi_v B \Pi_v A v + \langle v, Av \rangle V_B(v) = \Pi_v B A v,$$

and

$$\Pi_v A \Pi_v B v + \langle v, Bv \rangle V_A(v) = \Pi_v A B v,$$

hence

$$[V_A, V_B] = V_{BA} - V_{AB} = -V_{[A, B]}. \quad \square$$

Of fundamental importance is the following observation for the lifting operation  $X \mapsto \tilde{X}$ .

**Lemma C.2** *Any two vector fields  $X, Y \in \mathfrak{X}(M)$  satisfy the identity*

$$[\tilde{X}, \tilde{Y}] = [X, Y]^\sim.$$

*Thus the lifting operation  $X \mapsto \tilde{X}$  is a Lie algebra isomorphism onto  $\tilde{\mathfrak{X}}(M) = \{\tilde{X} : X \in \mathfrak{X}(M)\}$ .*

*Proof* Given a vector field  $X$  on  $M$ , its lift  $\tilde{X}$  on  $\mathbb{S}M$  we can always be split into horizontal  $\hat{X}$  and vertical  $\hat{V}_{\nabla X}$  fields as

$$\tilde{X} = \hat{X} + \hat{V}_{\nabla X},$$

according to the orthogonal splitting  $T_{(x, v)}\mathbb{S}M = T_x M \oplus T_v \mathbb{S}_x M$  induced by the Sasaki-Metric  $\tilde{g}$  and the associated lift of the Levi-Civita connection  $\tilde{\nabla}$  on  $\mathbb{S}M$ . Explicitly, the horizontal  $\hat{X}$  and vertical  $\hat{V}_{\nabla X}$  components of  $\tilde{X}$  are given by

$$\hat{X}(x, v) = (X(x), 0), \quad \hat{V}_{\nabla X}(x, v) = (0, V_{\nabla X(x)}(v)).$$

Let  $U(x, v) = (v, 0)$  be the ‘‘canonical’’ horizontal vector field on  $\mathbb{S}M$ . Note that  $U$  is parallel to  $\hat{X}$  in the sense that  $\tilde{\nabla}_{\hat{X}} U = 0$ . Additionally, define a linear mapping  $\hat{\Pi}_{(x, v)}$  on  $T_{(x, v)}\mathbb{S}M$  by

$$\hat{\Pi}_{(x, v)}(u_1, u_2) = (0, \Pi_{(x, v)}u_1),$$

so that we can express the vertical field  $\hat{V}_{\nabla X}$  as

$$\hat{V}_{\nabla X} = \hat{\Pi} \tilde{\nabla}_U \hat{X}.$$

Note that for any “horizontal” vector field  $\hat{X}$ ,  $\tilde{\nabla}_{\hat{X}} \hat{\Pi} = 0$  holds since  $\nabla$  preserves the metric  $g$ . In light of the fact that  $[\hat{X}, \hat{Y}] = [X, Y]$ , our proof will be complete if we show that

$$[\tilde{X}, \tilde{Y}] = [\hat{X}, \hat{Y}] + \hat{\Pi} \tilde{\nabla}_U [\hat{X}, \hat{Y}].$$

The Lie bracket of  $\tilde{X}$  and  $\tilde{Y}$  can be written as

$$[\tilde{X}, \tilde{Y}] = [\hat{X}, \hat{Y}] + [\hat{X}, \hat{V}_{\nabla Y}] - [\hat{Y}, \hat{V}_{\nabla X}] + [\hat{V}_{\nabla X}, \hat{V}_{\nabla Y}]. \quad (\text{C.1})$$

Likewise, a simple consequence of Lemma C.1 implies

$$[\hat{V}_{\nabla X}, \hat{V}_{\nabla Y}] = -\hat{\Pi} [\tilde{\nabla} \hat{X}, \tilde{\nabla} \hat{Y}] U = \hat{\Pi} \left( \tilde{\nabla}_{[U, \hat{X}]} \hat{Y} - \tilde{\nabla}_{[U, \hat{Y}]} \hat{X} \right),$$

where above  $[\tilde{\nabla} \hat{X}, \tilde{\nabla} \hat{Y}]$  denotes the commutator of  $\tilde{\nabla} \hat{X}, \tilde{\nabla} \hat{Y}$  viewed as linear endomorphisms on a fixed tangent space  $T_{(x, v)} \mathbb{S}M$ . The remaining terms in equation (C.1) can be computed as

$$[\hat{X}, \hat{V}_{\nabla Y}] - [\hat{Y}, \hat{V}_{\nabla X}] = \tilde{\nabla}_{\hat{X}} \hat{V}_{\nabla Y} - \tilde{\nabla}_{\hat{Y}} \hat{V}_{\nabla X} = \hat{\Pi} \left( \tilde{\nabla}_{\hat{X}} \tilde{\nabla}_U \hat{Y} - \tilde{\nabla}_{\hat{Y}} \tilde{\nabla}_U \hat{X} \right).$$

Therefore, putting everything together, we find

$$[\tilde{X}, \tilde{Y}] = [X, Y] \hat{\Pi} + \hat{\Pi} \left( \tilde{\nabla}_{\hat{X}} \tilde{\nabla}_U \hat{Y} - \tilde{\nabla}_{\hat{Y}} \tilde{\nabla}_U \hat{X} + \tilde{\nabla}_{[U, \hat{X}]} \hat{Y} - \tilde{\nabla}_{[U, \hat{Y}]} \hat{X} \right).$$

The proof will be complete once we show the identity

$$\tilde{\nabla}_{\hat{X}} \tilde{\nabla}_U \hat{Y} - \tilde{\nabla}_{\hat{Y}} \tilde{\nabla}_U \hat{X} + \tilde{\nabla}_{[U, \hat{X}]} \hat{Y} - \tilde{\nabla}_{[U, \hat{Y}]} \hat{X} = \tilde{\nabla}_U [\hat{X}, \hat{Y}]. \quad (\text{C.2})$$

For this, we can use the Riemann curvature tensor on  $\mathbb{S}M$

$$\tilde{R}(X, Y)Z := \tilde{\nabla}_X \tilde{\nabla}_Y Z - \tilde{\nabla}_Y \tilde{\nabla}_X Z - \tilde{\nabla}_{[X, Y]} Z$$

to change the order of covariant derivatives, giving

$$\begin{aligned} \tilde{\nabla}_{\hat{X}} \tilde{\nabla}_U \hat{Y} - \tilde{\nabla}_{\hat{Y}} \tilde{\nabla}_U \hat{X} + \tilde{\nabla}_{[U, \hat{X}]} \hat{Y} - \tilde{\nabla}_{[U, \hat{Y}]} \hat{X} &= \tilde{R}(\hat{X}, U) \hat{Y} - \tilde{R}(\hat{Y}, U) \hat{X} + \tilde{\nabla}_U \tilde{\nabla}_{\hat{X}} \hat{Y} - \tilde{\nabla}_U \tilde{\nabla}_{\hat{Y}} \hat{X} \\ &= \tilde{R}(\hat{X}, U) \hat{Y} + \tilde{R}(U, Y) \hat{X} + \tilde{\nabla}_U [\hat{X}, \hat{Y}]. \end{aligned}$$

The first Bianchi identity implies that

$$\tilde{R}(\hat{X}, U)\hat{Y} + \tilde{R}(U, \hat{Y})\hat{X} = \tilde{R}(\hat{X}, \hat{Y})U,$$

and therefore identity (C.2) follows from the fact that  $R(\hat{X}, \hat{Y})U = 0$  since, for any vector field  $Z \in \mathfrak{X}(M)$ , we have that  $\tilde{\nabla}_{\hat{Z}}U = 0$ .  $\square$

We are now ready to prove Proposition 2.7.

*Proof of Proposition 2.7* A simple consequence of Lemma C.2 that for any collection of vector fields  $\{X_k\}_{k=0}^r$  on  $M$  we have the following identification

$$\text{Lie}(\tilde{X}_0; \tilde{X}_1, \dots, \tilde{X}_r) = \{\tilde{X} : X \in \text{Lie}(X_0; X_1, \dots, X_r)\}.$$

Therefore the parabolic Hörmander condition for  $\{\tilde{X}_k\}_{k=0}^r$  is equivalent to

$$\{(X(x), V_{M_X(x)}(x, v)) : X \in \text{Lie}(X_0; X_1, \dots, X_r)\} = T_x M \oplus T_v \mathbb{S}_x M.$$

Clearly if the above condition is satisfied then  $\{X_k\}_{k=0}^r$  satisfies the parabolic Hörmander condition and 2.5 holds. The converse follows from the fact that (2.5) implies that for each  $X \in \text{Lie}(X_0; X_1, \dots, X_r)$ ,  $(x, v) \in \mathbb{S}M$  and  $h \in T_v \mathbb{S}_x M$  there exists a  $Y \in \text{Lie}(X_0; X_1, \dots, X_r)$  with  $Y(x) = 0$  such that

$$V_{M_Y(x)}(x, v) = h - V_{M_X(x)}(x, v)$$

and therefore  $Z = X + Y$  satisfies

$$\tilde{Z} = (X(x), h).$$

Combining this with the fact that

$$\{X(x) : X \in \text{Lie}(X_0; X_1, \dots, X_r)\} = T_x M,$$

concludes the proof.  $\square$

## C.2 Furstenberg–Khasminskii formula

The following identity is useful relating the divergence of a projective lift  $\tilde{X}$  to that of  $X$ ; it is well-known in the RDS community and the proof is simple, but as we could not locate a reference with the precise form we need, we provide a short proof here for the readers' convenience.

**Lemma C.3** *Let  $X \in \mathfrak{X}(M)$ . Then the following identity holds*

$$\text{div } \tilde{X}(x, v) = 2 \text{div } X(x) - n \langle v, \nabla X(x)v \rangle_x. \quad (\text{C.3})$$

Similarly, as we could not locate a statement that exactly matched the one we require (particularly the form of  $\tilde{Q}$  stated below), we include a proof of the Furstenberg-Khasminskii formula for (1.1) (see [13] for more details).

**Proposition C.4 (Furstenberg-Khasminskii)** *Define for each  $x \in M$*

$$Q(x) := \operatorname{div} X_0(x) + \frac{1}{2} \sum_{k=1}^r X_k \operatorname{div} X_k(x),$$

*and each  $w \in \mathbb{S}M$*

$$\tilde{Q}(w) := \operatorname{div} \tilde{X}_0(w) + \frac{1}{2} \sum_{k=1}^r \tilde{X}_k \operatorname{div} \tilde{X}_k(w).$$

*Suppose that  $(w_t)$  has a unique stationary probability measure  $\nu$  on  $\mathbb{S}M$  that projects to  $\mu$  on  $M$ , and that  $Q \in L^1(\mu)$  and  $\tilde{Q} \in L^1(\nu)$ , then the following formulas hold*

$$\lambda_\Sigma = \int_M Q \, d\mu, \tag{C.4}$$

$$n\lambda_1 - 2\lambda_\Sigma = - \int_{\mathbb{S}M} \tilde{Q} \, d\nu. \tag{C.5}$$

*Proof* We note that (C.4) is standard and can be found in a number of references (see for instance [13]). To prove (C.5), we see that a straight forward computation and formula (C.3) yields

$$\begin{aligned} d \log(|D\Phi^t(x)v|) &= \langle v_t, \nabla X_0(x_t)v_t \rangle dt + \sum_{k=1}^r \langle v_t, \nabla X_k(x_t)v_t \rangle \circ dW_t^k \\ &= \frac{1}{n} \left( 2 \operatorname{div} X_0(x_t) - \operatorname{div} \tilde{X}_0(w_t) \right) dt \\ &\quad + \frac{1}{n} \sum_{k=1}^r \left( 2 \operatorname{div} X_k(x_t) - \operatorname{div} \tilde{X}_k(w_t) \right) \circ dW_t^k. \end{aligned}$$

Converting to Itô and integrating in time gives

$$\frac{1}{t} \log(|D\Phi^t(x)v|) = \frac{1}{nt} \int_0^t 2Q(x_s)ds - \frac{1}{nt} \int_0^t \tilde{Q}(w_s)ds + \frac{1}{t} M_t,$$

with  $M_t$  is a mean-zero martingale arising from the Itô integral whose exact form is not important.

We now take  $t \rightarrow \infty$ , using a corollary of, e.g., Theorem III.1.2 in [49] to ensure the LHS converges to  $\lambda_1$ , while the first term on the RHS converges to  $\int \tilde{Q} d\nu$  by the ergodic theorem. In particular,  $\frac{1}{t} M_t$  must also converge, both pointwise and in  $L^1(\mathbf{P} \times \nu)$ , hence  $\frac{1}{t} M_t \rightarrow 0$  by the martingale law of large numbers.  $\square$

## References

1. Abedin, F., Tralli, G.: Harnack inequality for a class of Kolmogorov–Fokker–Planck equations in non-divergence form. *Arch. Ration. Mech. Anal.* **233**(2), 867–900 (2019)
2. Amari, S., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society (2000)
3. Anceschi, F., Polidoro, S., Ragusa, M.A.: Moser’s estimates for degenerate Kolmogorov equations with non-negative divergence lower order coefficients. *Nonlinear Anal.* **189**, 111568 (2019)
4. Armstrong, S., Mourrat, J.-C.: Variational methods for the kinetic Fokker–Planck equation, arXiv preprint [arXiv:1902.04037](https://arxiv.org/abs/1902.04037) (2019)
5. Arnold, L.: A formula connecting sample and moment stability of linear stochastic systems. *SIAM J. Appl. Math.* **44**(4), 793–802 (1984)
6. Arnold, L.: Random dynamical systems. *Dynamical systems*, pp. 1–43 (1995)
7. Arnold, L., Nguyen, D.C., Oseledets, V.: Jordan normal form for linear cocycles. *Random Oper. Stochastic Equ.* **7**(4), 303–358 (1999)
8. Arnold, L., Papanicolaou, G., Wihstutz, V.: Asymptotic analysis of the Lyapunov exponent and rotation number of the random oscillator and applications. *SIAM J. Appl. Math.* **46**(3), 427–450 (1986)
9. Auslender, E.I., Milstein, G.N.: Asymptotic expansion of Lyapunov exponent for linear stochastic systems with small noises. *Prikl. Mat. i Mekh.* **46**, 358–365 (1982). ((In Russ.))
10. Bakry, D., Émery, M.: Diffusions hypercontractives. *Séminaire de probabilités xix 1983*(84), 177–206 (1985)
11. Barreira, L., Pesin, Y.: Smooth ergodic theory and nonuniformly hyperbolic dynamics. *Handbook Dyn. Syst.* **1**, 57–263 (2006)
12. Baxendale, P.H.: Lyapunov exponents and relative entropy for a stochastic flow of diffeomorphisms. *Probab. Theory Relat. Fields* **81**(4), 521–554 (1989)
13. Baxendale, P.H.: Stability and equilibrium properties of stochastic flows of diffeomorphisms, *Diffusion processes and related problems in analysis, volume II. Stochastic flows*, pp. 3–35 (1992)
14. Baxendale, P.H.: Lyapunov exponents and stability for the stochastic Duffing–van der Pol oscillator. In: *Iutam Symposium on Nonlinear Stochastic Dynamics*, pp. 125–135 (2003)
15. Baxendale, P.H.: Stochastic averaging and asymptotic behavior of the stochastic duffing–van der pol equation. *Stoch. Process. Appl.* **113**(2), 235–272 (2004)
16. Baxendale, P.H., Goukasian, L.: Lyapunov exponents for small random perturbations of Hamiltonian systems. *Ann. Probab.*, pp. 101–134 (2002)
17. Bedrossian, J., Blumenthal, A., Punshon-Smith, S.: Lagrangian chaos and scalar advection in stochastic fluid mechanics. To appear in *J. Eur. Math. Soc.* (2018)
18. Bedrossian, J., Liss, K.: Quantitative spectral gaps and uniform lower bounds in the small noise limit for Markov semigroups generated by hypoelliptic stochastic differential equations. To appear in *Prob. Math. Phys.* (2020)
19. Bedrossian, J., Punshon-Smith, S.: Chaos in stochastic 2d Galerkin–Navier–Stokes. Preprint [arXiv:2106.13748](https://arxiv.org/abs/2106.13748) (2021)

20. Blumenthal, A., Xue, J., Young, L.-S.: Lyapunov exponents for random perturbations of some area-preserving maps including the standard map. *Ann. Math.*, pp. 285–310 (2017)
21. Bochi, J.: Genericity of zero Lyapunov exponents. *Ergodic Theory Dyn. Syst.* **22**(6), 1667–1696 (2002)
22. Bochi, J., Viana, M.: The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann. Math.*, pp. 1423–1485 (2005)
23. Boffetta, G., Cencini, M., Falcioni, M., Vulpiani, A.: Predictability: a way to characterize complexity. *Phys. Rep.* **356**(6), 367–474 (2002)
24. Bogachev, V.I.: Differentiable measures and the malliavin calculus. *J. Math. Sci.* **87**(4), 3577–3731 (1997)
25. Boothby, W.M., Wilson, E.N.: Determination of the transitivity of bilinear systems. *SIAM J. Control. Optim.* **17**(2), 212–221 (1979)
26. Carverhill, A.: A nonrandom Lyapunov spectrum for nonlinear stochastic dynamical systems. *Stoch.: Int. J. Probab. Stoch. Process.* **17**(4), 253–287 (1986)
27. Carverhill, A.: Furstenberg's theorem for nonlinear stochastic systems. *Probab. Theory Related Fields* **74**(4), 529–534 (1987)
28. Chang, J.T., Pollard, D.: Conditioning as disintegration. *Stat. Neerl.* **51**(3), 287–317 (1997)
29. Crovisier, S., Senti, S.: Un problème pour le xxi(i)ème siècle. *La Gazette des mathématiciens*, p. 156 (2018)
30. Da Prato, G., Zabczyk, J., Zabczyk, J.: Ergodicity for Infinite Dimensional Systems, vol. 229. Cambridge University Press, Cambridge (1996)
31. Doan, T. S., Engel, M., Lamb, J. S., Rasmussen, M.: Hopf bifurcation with additive noise, *Nonlinearity* **31** (2018), no. 10, 4567
32. Dolgopyat, D., Kaloshin, V., Koralov, L., et al.: Sample path properties of the stochastic flows. *Ann. Probab.* **32**(1A), 1–27 (2004)
33. Duarte, P.: Abundance of elliptic isles at conservative bifurcations. *Dyn. Stab. Syst.* **14**(4), 339–356 (1999)
34. Engel, M., Lamb, J.S., Rasmussen, M.: Bifurcation analysis of a stochastically driven limit cycle. *Commun. Math. Phys.* **365**(3), 935–942 (2019)
35. Furstenberg, H.: Noncommuting random products. *Trans. Am. Math. Soc.* **108**(3), 377–428 (1963)
36. Furstenberg, H.: Rigidity and cocycles for ergodic actions of semi-simple lie groups, *Séminaire bourbaki* vol. 1979/80 exposés 543–560, 1981, pp. 273–292
37. Glatt-Holtz, N.E., Herzog, D.P., Mattingly, J.C.: Scaling and saturation in infinite-dimensional control problems with applications to stochastic partial differential equations. *Ann. PDE* **4** (no. 2 (en)) (2018)
38. Golse, F., Imbert, C., Mouhot, C., Vasseur, A.: Harnack inequality for kinetic Fokker–Planck equations with rough coefficients and application to the landau equation, to appear in *Annali della Scuola Normale Superiore di Pisa* (2016)
39. Hairer, M.: On Malliavin's proof of Hörmander's theorem. *Bull. des Sci. Math.* **135**(6–7), 650–666 (2011)
40. Herzog, D. P. Mattingly, J. C: A practical criterion for positivity of transition densities. *Nonlinearity* **28** (2015), no. 8, 2823 (en)
41. Hörmander, L.: Hypoelliptic second order differential equations. *Acta Math.* **119**(1), 147–171 (1967)
42. Imkeller, P., Lederer, C.: An explicit description of the Lyapunov exponents of the noisy damped harmonic oscillator. *Dyn. Stab. Syst.* **14**(4), 385–405 (1999)
43. Jurdjevic, V.: Geometric Control Theory. Cambridge University Press, Cambridge (1997)
44. Jurdjevic, V.: Optimal Control and Geometry: Integrable Systems. Cambridge University Press, Cambridge (2016)
45. Jurdjevic, V., Kupka, I.: Control systems on semi-simple lie groups and their homogeneous spaces. *Ann. de l'institut fourier*, pp. 151–179 (1981)

46. Karimi, A., Paul, M. R.: Extensive chaos in the Lorenz-96 model. *Chaos: Interdiscip. J. Nonlinear Sci.* **20**(4), 043105 (2010)
47. Khasminskii, R.: *Stochastic Stability of Differential Equations*, Vol. 66, Springer, Berlin (2011)
48. Kifer, Y.: A note on integrability of  $C^r$ -norms of stochastic flows and applications. *Stochastic mechanics and stochastic processes*, pp. 125–131 (1988)
49. Kifer, Y.: *Ergodic Theory of Random Transformations*, vol. 10. Springer, Berlin (2012)
50. Kingman, J.F.C.: Subadditive ergodic theory. *Ann. Probab.* **1**(6), 883–899 (1973)
51. Kogoj, A. E., Polidoro, S.: Harnack inequality for hypoelliptic second order partial differential operators. *Potential Anal.* **45**(14), 545–555 (2016)
52. Krotov, V.G.: Criteria for compactness in  $l^p$ -spaces,  $p \geq 0$ . *Sb. Math.* **no. 7**, 1045 (en) (203) (2012)
53. Kuksin, S., Nersesyan, V., Shirikyan, A.: Exponential mixing for a class of dissipative pdes with bounded degenerate noise. *Geom. Funct. Anal.*, pp. 1–62 (2020)
54. Kuksin, S., Nersesyan, V., Shirikyan, A.: Mixing via controllability for randomly forced nonlinear dissipative pdes. *J. de l’École polytechnique-Mathématiques* **7**, 871–896 (2020)
55. Kunita, H.: *Stochastic Flows and Stochastic Differential Equations*, vol. 24. Cambridge University Press, Cambridge (1997)
56. Lanconelli, A., Pascucci, A., Polidoro, S.: Gaussian lower bounds for non-homogeneous kolmogorov equations with measurable coefficients. *J. Evol. Equ.*, pp. 1–19 (2020)
57. Ledoux, M.: On improved sobolev embedding theorems. *Math. Res. Lett.* **10**(5), 659–669 (2003)
58. Ledoux, M., Nourdin, I., Peccati, G.: Stein’s method, logarithmic Sobolev and transport inequalities. *Geom. Funct. Anal.* **25**(1), 256–306 (2015)
59. Ledrappier, F.: Quelques propriétés des exposants caractéristiques, École d’été de probabilités de saint-flour xii–1982, pp. 305–396 (1984)
60. Ledrappier, F.: Positivity of the exponent for stationary sequences of matrices. *Lyapunov exponents*, pp. 56–73 (1986)
61. Li, P., Yau, S.T.: On the parabolic kernel of the schrödinger operator. *Acta Math.* **156**, 153–201 (1986)
62. Lin, K.K., Young, L.-S.: Shear-induced chaos. *Nonlinearity* **21**(5), 899 (2008)
63. Liu, P.-D., Qian, M.: *Smooth Ergodic Theory of Random Dynamical Systems*. Springer, Berlin (2006)
64. Lorenz, E.N.: *The Nature and Theory of the General Circulation of the Atmosphere*, Vol. 218, World Meteorological Organization Geneva (1967)
65. Lorenz, E.N.: Predictability: a problem partly solved. In: *Proceedings of Seminar on Predictability* (1996)
66. Lyubich, M.: Almost every real quadratic map is either regular or stochastic. *Ann. Math.*, pp. 1–78 (2002)
67. Majda, A.J.: *Introduction to Turbulent Dynamical Systems in Complex Systems*. Springer, Berlin (2016)
68. Malicet, D.: Lyapunov exponent of random dynamical systems on the circle. *Ergodic Theory Dyn. Syst.*, pp. 1–28 (2020)
69. Meyn, S.P., Tweedie, R.L.: *Markov Chains and Stochastic Stability*. Springer, Berlin (2012)
70. Moshchuk, N., Khasminskii, R.: Moment Lyapunov exponent and stability index for linear conservative system with small random perturbation. *SIAM J. Appl. Math.* **58**(1), 245–256 (1998)
71. Mouhot, C.: De Giorgi-Nash-Moser and Hörmander theories: new interplays. In: *Proceedings of the International Congress of Mathematicians Rio de Janeiro*, pp. 2467–2493 (2018)
72. Newhouse, S.E.: Diffeomorphisms with infinitely many sinks. *Topology* **13**(1), 9–18 (1974)
73. Newhouse, S.E.: The abundance of wild hyperbolic sets and non-smooth stable sets for diffeomorphisms. *Publ. Mathé. de l’IHÉS*, **50**, 101–151 (1979)

74. Oseledets, V.I.: A multiplicative ergodic theorem. characteristic Ljapunov exponents of dynamical systems. *Trudy Moskovskogo Matematicheskogo Obshchestva* **19**, 179–210 (1968)
75. Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D., Yorke, J.A.: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus?: Dyn. Meteorol. Oceanogr.* **56**(5), 415–428 (2004)
76. Pardoux, E., Wihstutz, V.: Lyapunov exponent and rotation number of two-dimensional linear stochastic systems with small diffusion. *SIAM J. Appl. Math.* **48**(2), 442–457 (1988)
77. Pazó, D., Szendro, I.G., López, J.M., Rodríguez, M. A: structure of characteristic Lyapunov vectors in spatiotemporal chaos. *Phys. Rev. E* **78**(1), 016209 (2008)
78. Pesin, Y., Climenhaga, V.: Open problems in the theory of non-uniform hyperbolicity. *Discrete Contin. Dyn. Syst.* **27**(2), 589–607 (2010)
79. Pinsky, M.A., Wihstutz, V.: Lyapunov exponents of nilpotent Itô systems. *Stoch.: Int. J. Probab. Stoch. Process.* **25**(1), 43–57 (1988)
80. Raghunathan, M.S.: A proof of Oseledec's multiplicative ergodic theorem. *Israel J. Math.* **32**(4), 356–362 (1979)
81. Rezakhanlou, F., Villani, C., Golse, F.: Entropy methods for the Boltzmann equation: lectures from a special semester at the Centre Émile Borel. Institut H. Poincaré, Paris, 2001, Springer, Berlin (2008)
82. Rokhlin, V.A.: On the fundamental ideas of measure theory. *Matematicheskii Sbornik* **67**(1), 107–150 (1949)
83. Romito, M., Xu, L.: Ergodicity of the 3D stochastic Navier–Stokes equations driven by mildly degenerate noise. *Stoch. Process. Appl.* **121**(4), 673–700 (2011)
84. Royer, G.: Croissance exponentielle de produits Markoviens de matrices aléatoires. *Ann. de l'ihp probabilités et Stat.*, pp. 49–62 (1980)
85. Sasaki, S.: On the differential geometry of tangent bundles of riemannian manifolds, II. *TMJ Update* **14**(2), 146–155 (en) (1962)
86. 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, Calif., 1970/1971), Vol. III: Probability theory, pp. 333–359 (1972)
87. Toscani, G.: Entropy production and the rate of convergence to equilibrium for the Fokker–Planck equation. *Q. Appl. Math.* **57**(3), 521–541 (1999)
88. Triebel, H.: *Theory of Function Spaces II*. Birkhauser (1992)
89. Virtser, A.: On products of random matrices and operators. *Theory Probab. Appl.* **24**(2), 367–377 (1980)
90. Walters, P.: A dynamical proof of the multiplicative ergodic theorem. *Trans. Am. Math. Soc.* **335**(1), 245–257 (1993)
91. Walters, P.: *An Introduction to Ergodic Theory*, Vol. 79. Springer, Berlin (2000)
92. Wang, Q., Young, L.-S.: Toward a theory of rank one attractors. *Ann. Math.* **167**(2), 349–480 (2008)
93. Wilkinson, A.: What are Lyapunov exponents, and why are they interesting? *Bull. Am. Math. Soc.* **54**(1), 79–105 (2017)
94. Young, L.-S.: Ergodic theory of differentiable dynamical systems. *Real Complex Dyn. Syst.*, pp. 293–336 (1995)
95. Young, L.-S.: Mathematical theory of Lyapunov exponents. *J. Phys. A: Math. Theor.* **46**(25), 254001 (2013)
96. Zaslavsky, G.: The simplest case of a strange attractor. *Phys. Lett. A* **69**(3), 145–147 (1978)
97. Zimmer, R. J: *Ergodic Theory and Semisimple Groups*, Vol. 81. Springer, Berlin (2013)