

Long-time behaviour of degenerate diffusions: UFG-type SDEs and time-inhomogeneous hypoelliptic processes

Thomas Cass* Dan Crisan* Paul Dobson[†] Michela Ottobre[‡]

Abstract

We study the long time behaviour of a large class of diffusion processes on \mathbb{R}^N , generated by second order differential operators of (possibly) degenerate type. The operators that we consider need not satisfy the Hörmander Condition (HC). Instead, they satisfy the so-called UFG condition, introduced by Herman, Lobry and Sussman in the context of geometric control theory and later by Kusuoka and Stroock. We demonstrate the importance of the class of UFG processes in several respects: i) we show that UFG processes constitute a family of SDEs which exhibit, in general, multiple invariant measures (i.e. they are in general non-ergodic) and for which one is able to describe a systematic procedure to study the basin of attraction of each invariant measure (equilibrium state). ii) We use an explicit change of coordinates to prove that every UFG diffusion can be, at least locally, represented as a system consisting of an SDE coupled with an ODE, where the ODE evolves independently of the SDE part of the dynamics. iii) As a result, UFG diffusions are inherently "less smooth" than hypoelliptic SDEs; more precisely, we prove that UFG processes do not admit a density with respect to Lebesgue measure on the entire space, but only on suitable time-evolving submanifolds, which we describe. iv) We show that our results and techniques, which we devised for UFG processes, can be applied to the study of the long-time behaviour of non-autonomous hypoelliptic SDEs and therefore produce several results on this latter class of processes as well. v) Because processes that satisfy the (uniform) parabolic HC are UFG processes, this paper contains a wealth of results about the long time behaviour of (uniformly) hypoelliptic processes which are non-ergodic.

Keywords: diffusion semigroups; parabolic PDE; UFG condition; Hörmander condition; long time asymptotics; processes with multiple invariant measures; non-ergodic SDEs; distributions with non-constant rank; stochastic control theory.

*Department of Mathematics, Imperial College London, Huxley Building, 180 Queen's Gate, London SW7 2AZ, UK. E-mail: thomas.cass@imperial.ac.uk, d.crisan@imperial.ac.uk

[†]Delft Institute of Applied Mathematics, Delft University of Technology, 2628 XE Delft, The Netherlands. E-mail: p.dobson@tudelft.nl

[‡]Maxwell Institute for Mathematical Sciences, Department of Mathematics, Heriot-Watt University, Edinburgh EH14 4AS, UK. E-mail: m.ottobre@hw.ac.uk

MSC2020 subject classifications: 60H10; 35K10; 35B35; 35B65; 58J65; 49J55; 93E03; 37H10.

Submitted to EJP on February 2, 2020, final version accepted on December 23, 2020.

1 Introduction

1.1 Context and scope of the paper

Consider stochastic differential equations (SDEs) in \mathbb{R}^N of the form

$$X_t = X_0 + \int_0^t V_0(X_s) ds + \sum_{i=1}^d \int_0^t V_i(X_s) dB^i(s); \quad (1.1)$$

where V_0, \dots, V_d are smooth vector fields on \mathbb{R}^N , \int denotes Stratonovich integration and $B^1(t), \dots, B^d(t)$ are one dimensional independent standard Brownian motions. The Markov semigroup P_t associated with the SDE (1.1) is defined on the set $C_b(\mathbb{R}^N)$ of continuous and bounded functions as

$$P_t : C_b(\mathbb{R}^N) \rightarrow C_b(\mathbb{R}^N); \quad (P_t f)(x) := E[f(X_t) | X_0 = x]; \quad (1.2)$$

We recall that, given a vector field $V : \mathbb{R}^N \rightarrow \mathbb{R}^N$, we can interpret V both as a vector-valued function on \mathbb{R}^N and as a first order differential operator on \mathbb{R}^N :

$$V = (V^1(x); V^2(x); \dots; V^N(x)) \quad \text{or} \quad V = \sum_{j=1}^N V^j(x) \partial_j; \quad x \in \mathbb{R}^N; \quad \partial_j = \frac{\partial}{\partial x_j}; \quad (1.3)$$

With this notation, the Kolmogorov operator associated with the semigroup P_t is the second order differential operator given on smooth functions by

$$L = V_0 + \sum_{i=1}^d V_i^2; \quad (1.4)$$

The Markov diffusion X_t is called hypoelliptic (elliptic, respectively) when the operator L is hypoelliptic (elliptic, respectively) [4]. The study of diffusion processes of hypoelliptic type has by now produced a fully-fledged theory, involving several branches of mathematics: stochastic analysis, analysis of differential operators, (sub-)Riemannian geometry and control theory. One of the key steps in the development of such a theory has been the seminal paper of Hörmander [27] and a large body of work has been dedicated for over forty years to the study of diffusion processes under the Hörmander Condition (HC) (in one if its many forms), which is a sufficient condition for hypoellipticity. In particular, the ergodic theory for hypoelliptic SDEs is well developed, see [51, 23, 50, 19] and references therein – throughout the paper we define a process to be *ergodic* if it admits a unique invariant measure (stationary state).

To the best of our knowledge, this is the first paper that attempts to build a framework for the study of the long time asymptotics of solutions of SDEs which are non-necessarily hypoelliptic. We will work in the setting in which the vector fields V_0, \dots, V_d satisfy a weaker condition, the so-called *UFG condition*. The acronym UFG stands for *Uniformly Finitely Generated*. We give a precise statement of the UFG condition in Definition 3.1. For the moment let us just point out that, while the Parabolic Hörmander condition imposes the following

$$\left[\begin{array}{l} \text{span} \{L_j(x)g\} = \mathbb{R}^N \quad \text{for every } x \in \mathbb{R}^N; \end{array} \right. \quad \text{(PHC)}$$

where as customary the hierarchy of operators L_j is defined as $L_1(x) := fV_1(x); \dots; V_d(x)g$ and, for $j > 1$, $L_j(x) = L_{j-1}(x) [f[V_k; V_k]; V \geq L_{j-1}; k \geq f0; \dots; dgg$, under the UFG condition the vector space appearing in **(PHC)** is not required to have constant rank; roughly speaking, it is only required to be *finitely generated*. In particular, we emphasize that the UFG condition does not impose the vector space in **(PHC)** to be equal to \mathbb{R}^N for any $x \in \mathbb{R}^N$. Hence, in this sense, the UFG condition is weaker than the parabolic Hörmander condition. The UFG condition has been long known by the (geometric) control theory community, although perhaps under other names (see Section 3 for a more detailed account on the matter), and it is indeed well-studied in the works of Hermann, Lobry and Sussman [26, 39, 57]. It was then considered by Stroock and Kusuoka in the eighties [33, 34, 35, 36], though in a completely different context (which we briefly explain below). The purpose there was to study smoothing properties of the semigroup $fP_t g_{t=0}$ under the UFG condition. In this paper we combine the geometric viewpoint with the functional analytic and probabilistic one to introduce new results on the asymptotic behaviour of UFG diffusions. In broad terms, the two main achievements of this paper can be described as follows:

i) We study the diffusion process (1.1) in absence of the Hörmander condition. To this end, we establish explicit connections between the geometric theory of finitely generated Lie algebras and the related stochastic dynamics. Because every (uniformly) hypoelliptic process is a UFG process, our results cover a very large class of SDEs. In particular we show that our approach can be fruitfully employed to study the asymptotic behaviour of non-autonomous hypoelliptic diffusions.

ii) We argue that UFG processes constitute a class of SDEs which exhibit, in general, multiple equilibria and for which one is able, given an initial datum, to determine the invariant measure to which the dynamics will converge.

Let us further remark on the significance of the latter point: although a large body of work has been devoted to the study of ergodic processes, the development of a general framework to understand problems with multiple equilibria is at a very early stage. It is well known that ergodic processes will, under appropriate general conditions, converge to their unique equilibrium irrespective of the initial configuration, i.e. they will tend to lose memory of the initial datum. Clearly this cannot be the case, in general, for more complicated systems. When the invariant measure is not unique it is typically extremely difficult to determine the basin of attraction of each equilibrium measure and we are indeed not aware of any criteria developed to this effect. To be more precise, one can ask one of the two (complementary) questions: given an initial datum for the SDE, which equilibrium measure will the process converge to? Conversely, given an equilibrium measure μ , one may wish to describe the basin of attraction of such a measure, i.e. the set of initial data $x \in \mathbb{R}^N$ such that the process $X_t^{(x)}$ ¹ converges to μ . Beyond numerical simulations, no theoretical framework currently exists to tackle this kind of problems.

In this paper we introduce a systematic way to study long-time convergence for a large class of SDEs which will, in general, admit several stationary states. This methodology applies to UFG diffusions and hence, because processes that satisfy the (uniform) parabolic Hörmander condition are UFG processes, our results produce further understanding on non-ergodic Hörmander processes – we stress here, and we will emphasize it again in Section 4, that hypoelliptic processes need not be ergodic (see Section 4 for examples of hypoelliptic processes which are not ergodic).

The Markov diffusions studied here are *linear*, in the sense that their generators (1.4) are linear second order differential operators. As a point of comparison, another class of

¹We use the notation $X_t^{(x)}$ to emphasize the fact that the initial datum of the process is $X_0 = x$.

systems exhibiting multiple equilibria is the class of so-called *collective dynamics*: in this case the system is constituted by a large number of particles or agents that interact with each other. The underlying kinetic-PDEs for this type of models are *non-linear in the sense of McKean* and the existence of multiple stationary states here is due to such a nonlinearity. In our case, the nature of the phenomenon is completely different and in a way simpler, as multiple invariant measures arise as a result of the non-trivial control-theory implied by the UFG condition.

In the remainder of this introduction we comment on the implications and significance of the UFG condition first from an analytic perspective and then from a geometric and probabilistic viewpoint. In Subsection 1.2 we explain the main results of the paper and the reasons for studying UFG diffusions; we then conclude the introduction with Subsection 1.3, where we illustrate the organization of the paper.

As is well known, under the (parabolic) Hörmander condition, the transition probabilities of the semigroup $fP_t g_t$ have a smooth density; furthermore, $P_t f$ is differentiable in every direction and $u(t; x) := (P_t f)(x)$ is a classical solution of the Cauchy problem

$$\begin{aligned} @_t u(t; x) &= Lu(t; x) \\ u(0; x) &= f(x) \end{aligned}$$

In the present paper we will relax the hypoellipticity assumption and study the long-time behaviour of the dynamics (1.1) in absence of the Hörmander condition.

In a series of papers [33, 34, 35, 36, 15, 10, 12], Kusuoka and Stroock first and Crisan and collaborators later, have analyzed the smoothness properties of diffusion semigroups $fP_t g_t$ associated with the stochastic dynamics (1.1) when the vector fields $fV_i; i = 0; 1; \dots; dg$ satisfy the UFG condition. Such works showed that, as opposed to what happens under the PHC, under the UFG condition the semigroup P_t is no longer differentiable in every direction; in particular it is no longer differentiable in the direction V_0 , but it is still differentiable in the direction $V := @_t V_0$ when viewed as a function $(t; x) \mapsto u(t; x)$ over the product space $(0; \infty) \times \mathbb{R}^N$. This fact has been proved by means of Malliavin calculus and in this paper we give a geometric and analytic explanation of such a phenomenon. Because of differentiability in the direction V , a rigorous PDE analysis can still be built starting from the stochastic dynamics (1.1). In this case one can indeed prove that for every $f \in C_b$ (continuous and bounded), the function $u(t; x) := (P_t f)(x)$ is a classical solution² of the Cauchy problem

$$\begin{aligned} Vu(t; x) &= \sum_{i=1}^d V_i^2 u(t; x) \\ u(0; x) &= f(x) \end{aligned} \tag{1.5}$$

From a geometric and control-theoretical point of view, working with the UFG condition will imply dealing with distributions of non-constant rank.³ If the geometric understanding of the Hörmander condition is rooted in the classic Frobenius Theorem, which deals with distributions of constant rank, the geometry of the UFG condition is described in the works of Hermann, Lobry and Sussman [26, 39, 57]. In these works, the UFG condition was considered for geometric and control theoretical purposes, in particular for the study of reachability (i.e., roughly speaking, to answer questions regarding the set of points that can be reached by the integral curves of given vector fields). In this respect we should stress that the UFG condition is not optimal from a control-theoretical point of view (an optimal condition for reachability has been described by Sussmann [57]). However, it is the closest to being optimal, while still being easy to check in practice.

²The notion of classical solution for the PDE (1.5) and further background material can be found in [14, Appendix A].

³In this paper we use the word distribution only in geometric sense, see definition at the beginning of Section 1.2.

Finally, by a probabilistic standpoint, it is well known that the Parabolic Hörmander condition (PHC) is a sufficient (and almost necessary) condition for the law of the process (1.1) to have a density, see [24], and this fact has motivated the large literature on hypoelliptic SDEs. Again, the understanding of this matter relies on Frobenius Theorem, as Hörmander himself noted [27]. In his seminal paper [7], Bismut proved that, when the Hörmander condition is enforced in place of the PHC,⁴ the law of the process no longer admits a density on \mathbb{R}^N ; however, it admits a density on appropriate time-dependent submanifolds of \mathbb{R}^N . In this paper we prove that a similar statement holds, in more generality, for UFG processes, and in Section 8 we explicitly describe the time-dependent manifolds on which the process admits a law. Throughout the paper we will make several comparisons between the setting of [7] and the present setting.

1.2 Main results

The main results of this paper are the following: Proposition 4.7, Proposition 5.1 and Proposition 5.7 give a description of the global behaviour of the dynamics (1.1), under the sole assumption that the vector fields V_0, \dots, V_d satisfy the UFG condition; Theorem 6.6 and Theorem 7.9 describe the long time behaviour of non-autonomous hypoelliptic processes and of UFG processes, respectively, identifying invariant measures and characterizing their basin of attraction; finally in Theorem 8.9 we describe appropriate manifolds where the process X_t admits a density. Let us give a rather informal description of such results. Precise notation, assumptions and statements are deferred to the relevant sections.

A *distribution* on \mathbb{R}^N is a map that, to each point $x \in \mathbb{R}^N$, associates a linear subspace of the tangent space $T_x\mathbb{R}^N$. Given a set D of smooth vector fields on \mathbb{R}^N , the distribution generated by D , denoted by $\hat{\cdot}_D$, is the map $x \mapsto \text{span}fX(x) : X \in Dg$. Let us introduce two distributions, $\hat{\cdot}(x)$ and $\hat{\cdot}_0(x)$, that will play a fundamental role in this paper. To avoid having to set too much notation and nomenclature, we introduce them now informally but we will give precise definitions at the beginning of Section 4.⁵ The distribution $\hat{\cdot}$ is generated by the vector fields contained in the Lie algebra (PHC), i.e. the distribution

$$\hat{\cdot}(x) = \bigcup_{j=1}^d \text{span}fL_j(x)g \tag{1.6}$$

while

$$\hat{\cdot}_0(x) = \text{span}f\text{Lie}fV_0(x); V_1(x); \dots; V_d(x)gg \tag{1.7}$$

$$= \text{span}fV_0(x) \cap \hat{\cdot}(x)g: \tag{1.8}$$

Clearly, $\hat{\cdot}(x) \cap \hat{\cdot}_0(x)$ for every $x \in \mathbb{R}^N$ and the two distributions coincide at x if and only if $V_0(x)$ is a combination of the vectors contained in $\hat{\cdot}$. More precisely, we decompose the vector V_0 into a component which belongs to $\hat{\cdot}$, $V_0^{(\hat{\cdot})}$, and a component which is orthogonal to $\hat{\cdot}$, $V_0^{(?)}$:

$$V_0 = V_0^{(\hat{\cdot})} + V_0^{(?)}: \tag{1.9}$$

In other words, $V_0^{(?)}(x)$ is the projection of $V_0(x)$ on the orthogonal of the vector space $\hat{\cdot}(x)$, so the two distributions coincide if and only if $V_0^{(?)} = 0$. We will see that the vector $V_0^{(?)}$ plays an important role for the dynamics and, ultimately, it is the component of V_0 responsible for the lack of smoothness in the direction V_0 .⁶ Therefore, in a way, the

⁴The difference between the PHC and the HC will be clarified in Section 3.

⁵In that section we define them differently, but we then prove that the definition we give there is equivalent to the one we state in (1.6)–(1.7).

⁶Note that even when V_0 is smooth, $V_0^{(?)}$ need not be smooth, see Note 4.14 on this matter.

distribution $\hat{\nu}$ is the one containing all the directions along which the problem (1.5) is smooth. We will come back to this later.

Under the UFG condition the integral manifolds (see Section 3.3 for definition) of $\hat{\nu}_0$ form a partition of the state space \mathbb{R}^N . Let S be one such manifold.⁷ If $X_0 = x \in S$ then $X_t^{(x)} \in S$ for all $t \geq 0$. That is, if the process starts from one of the manifolds of the partition, then it remains in the closure of such a manifold; but, crucially, it may hit the (topological) boundary $\partial S := \overline{S} \setminus S$ of the manifold S . This is the content of Proposition 4.7. Such a statement is obtained by combining the known geometric theory of distributions with non-constant rank with the classical Stroock–Varadhan support theorem. We further prove that if X_t hits the boundary ∂S of the manifold S , then it never leaves it, see Proposition 5.1 and Note 5.2. Therefore: i) because the dimension of the boundary ∂S is smaller than the dimension of S , along the path of $X_t^{(x)}$ the rank of the distribution cannot increase; ii) if the solution of the SDE leaves the manifold S from where it started, then any invariant measure can only be supported on the boundary ∂S of such a manifold, see Proposition 5.7.

Further understanding of the dynamics relies on the results of Section 4.2: in this section we show that, after an appropriate change of coordinates, any N -dimensional SDE of UFG-type can be written, at least locally, as a system of the form

$$dZ_t = U_0(Z_t; \hat{\nu}_t)dt + \sum_{j=1}^{\infty} U_j(Z_t; \hat{\nu}_t) dB_t^j \tag{1.10}$$

$$d\hat{\nu}_t = \hat{W}_0(\hat{\nu}_t)dt; \tag{1.11}$$

where $\hat{\nu}_t$ solves an ordinary differential equation (ODE), $\hat{\nu}_t \in \mathbb{R}^{N \times n}$,⁸ $Z_t \in \mathbb{R}^n$, $\hat{W}_0 : \mathbb{R}^{N \times n} \rightarrow \mathbb{R}^{N \times n}$ and $U_j : \mathbb{R}^n \rightarrow \mathbb{R}^n$ for every $j \geq 0, \dots, dg$. Beyond details about the dimensionality of the ODE component, the important thing is that the solution of the ODE $\hat{\nu}_t$ evolves independently of the SDE part, while the coefficients of the SDE depend on the evolution of the ODE. We will informally refer to such a representation as being of the form “ODE+SDE”. In general, this representation is only local. This change of coordinates has been known for a long time in differential geometry, at least since Frobenius, see [28]; here we are simply expressing it in a way which is more congenial to our setting and purposes and we apply it to SDEs. While the change of coordinates itself is not new, to the best of our knowledge it has been used in SDE theory only by Bismut in [7, Section 5], to study the density of SDEs that satisfy the Hörmander Condition, but it has never been used to study the long-time behaviour of SDEs. To clearly compare our work with [7], let us emphasize that in the notation introduced so far, the HC is satisfied when the distribution $\hat{\nu}_0$ has rank equal to N at every point (see Section 3 for more details). In this paper we primarily exploit the representation (1.10)–(1.11) to study the long time behaviour of SDEs that satisfy the UFG condition (so, as we said already, the rank of $\hat{\nu}_0$ is not constant) but we also use it briefly in Section 8 to study the density of UFG processes.⁹ This local representation is both an important technical tool throughout the paper and a fundamental element in understanding the evolution of the dynamics. Referring to the PDE (1.5), we also note here that the change of coordinates gives a geometric interpretation of the (potential) lack of smoothness in the direction V_0 and of the reason why smoothness is instead maintained in the direction V , see Note 4.13 on this point.

⁷By definition of integral manifold, on each one of these manifolds the rank of the distribution $\hat{\nu}_0$ is constant and it is equal to the dimension of the manifold itself.

⁸To make a link with the more precise notation that we will use in Section 4.2, we are denoting here by $\hat{\nu}_t$ the components $(\nu_t; a_t)$ in (4.8)–(4.9), i.e. $\hat{\nu}_t = (\nu_t; a_t)$.

⁹Further comparisons between the setting of [7] and the setting of this paper can be found later in this section, in Note 8.1 and in Note 7.2.

In view of the discussed change of coordinates, it makes sense to start by studying UFG dynamics for which the representation (1.10)–(1.11) is global. For this reason in Section 6 we consider systems which are (globally) of the form (1.10)–(1.11), where the ODE is assumed to be one-dimensional and the SDE satisfies a form of Hörmander condition. More precisely, the dynamics studied in Section 6 are non-autonomous hypoelliptic SDEs; because the topic is somewhat of independent interest, this section has been written in such a way that it can be read independently of the rest of the paper. Non-autonomous SDEs and their associated two-parameter semigroup have been studied in [9], where a detailed analysis of the law of the process is carried out, in [13] where the associated semigroup is examined, and in [32, 16], where the authors introduce some interesting techniques to deal with the analysis of invariant measures and long-time behaviour of time-inhomogeneous processes. The work [9] assumes that the non-autonomous SDE is hypoelliptic, while in [32] a uniform ellipticity assumption is enforced. From a technical point of view, the results of Section 6 extend the approach of [32, Section 6.1] to the hypoelliptic setting. However the main difference between our results and the results in [32] is that here we highlight the fact that the process may admit several invariant measures and we characterize the basin of attraction of each of them. In this setting convergence to equilibrium is driven by the ODE component. We will indeed show that the basin of attraction of each invariant measure can be completely described by just looking at the behaviour of the solution of the ODE. Because the ODE is assumed to be one-dimensional and autonomous, it can only behave monotonically, so the analysis of the ODE and of the full problem is relatively intuitive in this setting (see Section 6 for details).

In Section 7 we consider the general case of UFG processes for which the representation (1.10)–(1.11) is only local. While this case is substantially richer than the previous one, the fact that, locally, we can always represent the SDE (1.1) as a system of the form ODE+SDE, still means that there is some deterministic behaviour which is intrinsic to UFG dynamics. It turns out that one is still able to single out the deterministic behaviour. Recalling the definition of the vector $V_0^{(?)}$, formula (1.9), we will show that the (N -dimensional) ODE

$$d_t = V_0^{(?)}(t)dt$$

plays, in this more general context, the same driving role that the ODE (1.11) had in the context of Section 6. Motivated by the above discussion, we introduce the process

$$Z_t := e^{tV_0^{(?)}}(X_t):$$

This process is non-autonomous and, as we will explain, it can be interpreted geometrically as being a projection of the process X_t on an appropriate integral manifold of the distribution $\hat{\cdot}$. We apply the techniques of Section 6 to the study of such a non-autonomous process, producing results on the long-time behaviour of Z_t . We then relate the asymptotic behaviour of Z_t to the asymptotic behaviour of X_t . Notice that the procedure that we have just described is somewhat the reverse of the one that is traditionally used (and it is, to the best of our knowledge, new): given a non-autonomous system, the established methodology consists of increasing the dimension of the state space by adding time as an auxiliary variable, thereby reducing the given non-autonomous system to a (larger) autonomous one. Here we do the converse: by projecting the process on an appropriate manifold, we reduce to a (lower-dimensional) non-autonomous one, Z_t , with the advantage that now the techniques of Section 6 can be adapted to prove statements on Z_t . Once the latter process has been understood, we deduce results about the autonomous process X_t from those shown for Z_t .

From a probabilistic point of view it is clear that, in absence of the Hörmander condition, we cannot expect the process X_t to have a density with respect to the

Lebesgue measure. This is made explicit by the local representation (1.10)–(1.11), which also clarifies that it is the ODE component to be responsible for the lack of smoothness. Notice also that, in the coordinates (1.10)–(1.11), the vector $V_0^{(?)}$ is given by $V_0^{(?)} = (0; \dots; 0; \dot{W}_0)$, i.e. it is precisely the vector driving the ODE behaviour (we have elaborated on this fact in Note 4.13). However in Section 8 we show that the law of the SDE (1.1) still has a density on an appropriate time-dependent submanifold, which can be explicitly described. In order to do so, we correct and then extend the results of [55].

One may also wish to point out that systems of the form “ODE+SDE” appear as diffusion limits of some Metropolis-Hastings type of algorithms, see e.g. [31]. It is noted in [31] that one may use the ODE as a way to monitor convergence. We believe that the lack of smoothness of UFG processes could be seen as a perk in the context of sampling. The authors intend to explore this fact in future work. Finally, we mention in passing that UFG processes play a fundamental role in the study of cubature methods, see [12] and references therein for a complete account on the matter.

1.3 Organisation of the paper

In Section 2 we introduce the standing notation for the remainder of the manuscript. To make the paper self-contained, in Section 3 we gather background definitions and notions. In particular Subsection 3.1 contains details of the UFG condition, while Subsection 3.3 covers basic definitions and standard results in differential geometry and (stochastic) control theory. In Section 4 we exploit the existing theory of distributions of non-constant rank to produce both global and local results about the SDE (1.1), under the UFG condition. In Section 4.1 we cover the *global* behaviour of the SDE, in Section 4.2 we study *local* properties. In Section 5 we introduce several results for UFG-diffusions. These results are quite general, in the sense that most of them valid under just the UFG condition. The following Section 6 can be read independently of the rest of the manuscript: in this section we describe the long-time behaviour of hypoelliptic SDEs of non-autonomous type. The class of SDEs considered in Section 6 is one for which the representation of the form “ODE+SDE” is global. This is the first section where we address the problem of studying the basin of attraction of different invariant measures.¹⁰ In Section 7 we instead study the long time behaviour of (1.1) in the general UFG case (in which the change of coordinates is only local). Section 8 is devoted to the study of the density of the process, via Malliavin calculus. Finally, for ease of reading, we chose to relegate almost all the proofs to the appendix. In particular, Appendix A contains some needed miscellaneous technical facts, Appendix B contains the proofs of all the statements contained in Section 3 to Section 8.

2 Notation

We will be interested in N -dimensional SDEs, of the form (1.1). The letter N will only be used to refer to the dimension of the state space. While examples of UFG diffusions can be found in any dimension, it is fair to say that the theory we develop in this paper is mostly interesting in dimension $N \geq 2$, so we will make this a standing assumption which will hold unless otherwise stated in specific examples.

If x is a point in \mathbb{R}^N , we denote the j th coordinate of x by x^j , i.e. $x = (x^1; \dots; x^N)$ (this is coherent with (1.3)). We will often use a local change of coordinates, presented in Section 4.2. The change of coordinates will be given by a local diffeomorphism $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^N$ and the new coordinates will typically be denoted by z , i.e. $z = \phi(x)$. In

¹⁰We clarify that throughout this paper by invariant measure we understand this to be an invariant probability measure, that is any invariant measure μ must have $\mu(\mathbb{R}^N) = 1$.

the new coordinate system it will be of particular importance to distinguish the role of the first n coordinates of \mathbf{z} from the others (n being an appropriate integer, $n < N$). In particular, if $N - n > 1$, we will use the following notation

$$\mathbf{z} = (\underbrace{z^1, \dots, z^n}_{z}; \underbrace{z^{n+1}, \dots, z^N}_{a}) \tag{2.1}$$

where $(z^1, \dots, z^n) = z \in \mathbb{R}^n$, $z^{n+1} = \dots = z^N = a \in \mathbb{R}^{N - (n+1)}$. The last block of coordinates plays a role which is different from the role of the first two blocks, as it will be explained (the coordinates in the last block should more be intended as parameters). If $N = n + 1$ then simply $\mathbf{z} = (z; a)$.

A similar reasoning holds for the vector fields appearing in (1.1): for any $j \in \{0, \dots, d\}$, $V_j = (V_j^1, \dots, V_j^N)$ and \mathcal{V}_j will denote the vector V_j , expressed in the new coordinate system $\mathbf{z} = (z; a)$. We will show that in the new coordinate system, one has

$$\mathcal{V}_j(\mathbf{z}) = (U_j(\mathbf{z}); 0, \dots, 0) \quad j = 1, \dots, d \tag{2.2}$$

$$\mathcal{V}_0 = (U_0(\mathbf{z}); W_0(z; a); 0, \dots, 0); \tag{2.3}$$

where $U_j : \mathbb{R}^N \rightarrow \mathbb{R}^n$ while W_0 is a real-valued function which depends only on the last two blocks of coordinates of \mathbf{z} , i.e. $W_0 : \mathbb{R}^{N - n} \rightarrow \mathbb{R}$.

Accordingly, if X_t is the solution at time t of the SDE (1.1), then X_t^j denotes the j th component of X_t . We will sometimes want to stress the dependence of the solution X_t on the initial datum; when this is the case, we will write $X_t^{(x)}$ if $X_0 = x$. Finally, given a probability measure \mathbb{P} and a function f which is integrable with respect to \mathbb{P} , we shall define $\mathbb{E}(f)$ by

$$\mathbb{E}(f) := \int f d\mathbb{P}$$

We shall use the following function spaces throughout the paper. For any $N \geq 1$ and closed set $E \subset \mathbb{R}^N$;

- We denote by $C_b(E)$ the space of all functions $f : E \rightarrow \mathbb{R}$ which are continuous and bounded; this space will be endowed with the supremum norm.
- We say that a real-valued function f is C^1 if it has continuous derivatives of all orders.
- We denote by $C_c^1(\mathbb{R}^N)$ the set of all functions $f : \mathbb{R}^N \rightarrow \mathbb{R}$ which are C^1 and with compact support.

Given a differentiable function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ we denote by $J_x f$ the Jacobian matrix of f , that is $(J_x f)^{ij}(x) = \partial_{x^j} f^i(x)$.

3 Preliminaries and assumptions

3.1 The UFG condition

Fix $d \in \mathbb{N}$ (here \mathbb{N} denotes the set of non-zero integers) and let A_0 be the set of all k -tuples, of any size $k \geq 1$, of integers of the following form

$$A_0 := \{ f = (k^1, \dots, k^d) : k \in \mathbb{N}; j \in \{0, 1, \dots, d\} \}$$

We endow A_0 with the product operation

$$f \cdot g := (k^1, \dots, k^h; l^1, \dots, l^d);$$

for any $\alpha = (\alpha_1, \dots, \alpha_h)$ and $\beta = (\beta_1, \dots, \beta_h)$ in A_0 . Set $A = A_0 \cap f(0)g$. We emphasise that all k -tuples of any length $k \geq 1$ are allowed in A , except the trivial one, $\alpha = (0)$ (however singletons $\alpha = (j)$ belongs to A if $j \in f(1, \dots, dg)$). If $\alpha \in A$, we define the *length* of α , denoted by $k(\alpha)$, to be the integer

$$k(\alpha) := h + \text{card}\{i : \alpha_i = 0\}; \quad \text{if } \alpha = (\alpha_1, \dots, \alpha_h):$$

For any $m \in \mathbb{N}; m \geq 1$, we then introduce the sets

$$A_m = \{ \alpha \in A : k(\alpha) \leq m \}$$

Given a vector field (or, equivalently, a first order differential operator) $V = (V^1(x), V^2(x), \dots, V^N(x))$ on \mathbb{R}^N , we refer to the functions $fV^j(x)g_{1 \leq j \leq N}$ as to the *components* or *coefficients* of the vector field. We say that a vector field is smooth or that it is C^1 if all the components $V^j(x), j = 1, \dots, N$, are C^1 functions. Given two differential operators V and W , the commutator between V and W is defined as

$$[V; W] := VW - WV:$$

Let now $fV_i : i = 0, \dots, dg$ be a collection of vector fields on \mathbb{R}^N and let us define the following ‘‘hierarchy’’ of operators:

$$\begin{aligned} V_{[i]} &:= V_i \quad i = 0, 1, \dots, d \\ V_{[i_1 \dots i_k]} &:= [V_{[i_1]}; V_{[i_2]}] \dots V_{[i_k]} \quad \alpha \in A; i = 0, 1, \dots, d: \end{aligned}$$

This hierarchy is completely analogous to the one constructed in Introduction, here we just need a more detailed notation. Note that if $k(\alpha) = h$ then $k(\alpha + ik) = h + 1$ if $i \in f(1, \dots, dg)$ and $k(\alpha + ik) = h + 2$ if $i = 0$. If $\alpha \in A$ is a multi-index of length h , with abuse of nomenclature we will say that $V_{[\alpha]}$ is a differential operator of length h . We can then define the space R_m to be the space containing all the operators of the above hierarchy, up to and including the operators of length m (but excluding V_0):

$$R_m := \{ V_{[\alpha]} : \alpha \in A_m \} \tag{3.1}$$

Let $C_V^1(\mathbb{R}^N)$ denote the set of bounded smooth functions, $f = f(x) : \mathbb{R}^N \rightarrow \mathbb{R}$, which have bounded derivatives of all orders and such that

$$\sup_{x \in \mathbb{R}^N} |V_{[\alpha_1]} \dots V_{[\alpha_k]} f(x)| < 1 \tag{3.2}$$

for all k and all $\alpha_1, \dots, \alpha_k \in A_m$.¹¹ With this notation in place we can now introduce the definition that will be central in this paper.

Definition 3.1 (UFG Condition) Let $fV_i : i = 0, \dots, dg$ be a collection of smooth vector fields on \mathbb{R}^N and assume that the coefficients of such vector fields have bounded partial derivatives (of any order). We say that the fields $fV_i : i = 0, \dots, dg$ satisfy the UFG condition if there exists $m \in \mathbb{N}$ such that for any $\alpha \in A$ of the form

$$\alpha = (0, \dots, i, \dots, 0) \in A_m; i \in f(0, \dots, dg);$$

one can find bounded smooth functions $f; g = f; (x) \in C_V^1(\mathbb{R}^N)$ such that

$$V_{[\alpha]}(x) = \sum_{\beta \in A_m} f; g(x) V_{[\beta]}(x) \tag{3.3}$$

¹¹The definition of the set $C_V^1(\mathbb{R}^N)$ depends on m as well, but we do not include this dependence in the notation for simplicity.

Again we emphasize that the set of vector fields appearing in the linear combination on the right hand side of the above identity does not include V_0 . It may be useful to compare the UFG condition with the Hörmander condition, the uniform parabolic Hörmander condition (UPHC) and the Parabolic Hörmander condition (PHC), which we recall. The HC is satisfied if

$$\text{span}(\text{Lie} fV_0(x); \dots; V_d(x)g) = \mathbb{R}^N \quad \text{for every } x \in \mathbb{R}^N : \quad \text{(HC)}$$

In other words, the HC is satisfied if $\hat{\wedge}_0(x) = \mathbb{R}^N$ for every $x \in \mathbb{R}^N$. The PHC has been recalled in Introduction, see **(PHC)**. We notice in passing that while the space R_m is in general different from the space L_m ,¹² it is the case that $[j-1]R_j = [j-1]L_j$. The UPHC (see [11]) is instead satisfied if

$$\rho_{j-1} > 0 : \sum_{V_{[j-1]}(x)} V_{[j-1]}(x) \cdot y^2 = \rho_{j-1} |y|^2 \quad \text{for every } x, y \in \mathbb{R}^N : \quad \text{(UPHC)}$$

In the above each term of the sum is the scalar product between the vector $V_{[j-1]}(x)$ and the vector $y \in \mathbb{R}^N$. Notice that the UPHC is the strongest of all these conditions, in the sense that

$$\begin{aligned} & \text{(UPHC)} \implies \text{(PHC)} \implies \text{(HC)} \\ & \text{(UPHC)} \implies \text{(UFG)} : \end{aligned}$$

However neither the HC nor the PHC imply the UFG condition (as one may, in general, need infinitely many fields to satisfy the PHC or the HC). We also note that while the various Hörmander conditions are imposed on an appropriate Lie Algebra, the UFG condition is rather a condition on the set of vectors $fV_{[j-1]} \in A_m g$, seen as a module over the ring C_V^j .

Example 3.2. Consider one-dimensional geometric Brownian motion, that is, the solution of the following SDE:

$$\begin{aligned} dX_t &= X_t dt + \rho_{-2} X_t dB_t \\ &= 2X_t dt + \rho_{-2} X_t dB_t \end{aligned}$$

Here $V_1 = x @_x; V_0 = 2V_1$ and $[V_1; V_0] = 0$. These vector fields satisfy the UFG condition with $m = 1$ however V_0 and V_1 vanish when $x = 0$ so the HC is not satisfied.

Example 3.3. Consider the following first order differential operators on \mathbb{R}^2

$$V_0 = \sin x @_y \quad V_1 = \sin x @_x :$$

Then $fV_0; V_1 g$ do not satisfy the Hörmander condition (e.g. there is always a degeneracy at $x = 0$) but they do satisfy the UFG condition with $m = 4$. If the role of the fields is exchanged, i.e. if we set

$$V_0 = \sin x @_x; \quad V_1 = \sin x @_y$$

then $fV_0; V_1 g$ still satisfy the UFG condition, this time with $m = 1$ (indeed, $[V_0; V_1] = \cos x V_1$). □

Note 3.4. Because the functions $f; g$ appearing in (3.3) belong to $C_V^j(\mathbb{R}^N)$, if the UFG condition holds for some $m \geq N$ then it also holds for any $m; k \geq N$. In other words, if the UFG condition holds for some m in \mathbb{N} then for any $V_{[k-1]}$ with $k > m$ one has

$$V_{[k-1]}(x) = \sum_{V_{[m]}(x)} V_{[m]}(x) V_{[k-1]}(x) ;$$

¹² $V_{[\alpha]} \in L_j$ if and only if $k - k - 1 + 2(j - 1) = 2j - 1$, hence $R_{2j-1} = [j_{k=1}^j L_j$.

for some functions $\psi_j \in C^1_V(\mathbb{R}^N)$. For this reason it is appropriate to remark that in the remainder of the paper, when we assume that “the UFG condition is satisfied for some m ”, we mean the smallest such m . \square

We will consider diffusion semigroups $\{P_t\}_{t \geq 0}$ of the form (1.2); that is, we consider Markov semigroups associated with the stochastic dynamics (1.1). In particular, we will be interested in studying the semigroup P_t when the vector fields $\{V_0, V_1, \dots, V_d\}$ satisfy the UFG condition.

As we have already mentioned, the UFG condition is strictly weaker than the uniform Parabolic Hörmander condition. However one can still prove that, when the UFG condition is satisfied by the vector fields $\{V_0, V_1, \dots, V_d\}$ appearing in the generator (1.4), the semigroup $\{P_t\}_{t \geq 0}$ still enjoys good smoothing properties: if $f(x)$ is continuous then $(P_t f)(x)$ is differentiable (infinitely many times) in all the directions spanned by the vector fields contained in R_m (we recall that the set R_m is defined in (3.1)). See Appendix A.2 for more details.

3.2 Obtuse angle condition

When the semigroup $\{P_t\}_{t \geq 0}$ is elliptic or hypoelliptic, several works have dealt with the study of the long and short time behaviour of the derivatives of the semigroup, for a review see [4, 40]. To the best of our knowledge, the only work addressing the study of the long-time behaviour of the derivatives of UFG semigroups is [14]. In [14] the authors identify a sufficient condition for exponential decay of the derivatives of the solution of (1.5). To be more precise, they proved the following: suppose the vector fields $\{V_0, V_1, \dots, V_d\}$ satisfy the UFG condition and assume there exists $\delta_0 > 0$ such that for all f sufficiently smooth and for every $\alpha \in A_m$ we have

$$(\nabla_{\alpha} f(x))(\nabla_{\beta} f(x)) \geq \delta_0 |\nabla_{\beta} f(x)|^2; \text{ for every } x \in \mathbb{R}^N; \tag{3.4}$$

It is understood in the above that δ_0 is independent of f . If δ_0 is sufficiently large then, for every $r > 0$ and $t_0 > 0$, we may find a constant $c_{t_0,r} > 0$ such that for any $f \in C_b(\mathbb{R})$, $t \geq t_0$ and $\alpha \in A_m$ we have

$$\sup_{x \in B(0;r)} |\nabla_{\alpha} (P_t f)(x)| \leq c_{r,t_0} \|f\|_1 e^{-\delta_0 t}; \tag{3.5}$$

for some $\delta_0 > 0$ (which depends on δ_0). In the above $B(0;r)$ is the centered ball (of \mathbb{R}^N) of radius r . Condition (3.4) was named the *Obtuse Angle Condition* (OAC) in [14]. Here we will need a second order version of such a result, as well.

Lemma 3.5. *Let $\{P_t\}_{t \geq 0}$ be the semigroup associated with the SDE (1.1) and assume that the vector fields $\{V_0, \dots, V_d\}$ satisfy the UFG condition. Suppose moreover that the following holds: there exists $\delta_0 > 0$ such that*

$$(\nabla_{\alpha} \nabla_{\beta} f)(x) \geq \delta_0 |\nabla_{\beta} f(x)|^2; \tag{3.6}$$

for every $x \in \mathbb{R}^N$ and for all $\alpha, \beta \in A_m$ such that $\alpha \neq \beta$ and $\alpha \perp \beta$. If $\delta_0 > 0$ is large enough then, for any $t_0 \in (0;1)$ and any $r > 0$ there exists a constant $c_{t_0,r} > 0$ such that, for some $\delta > 0$ (which depends, among other things, on δ_0 but independent of f), one has

$$\sup_{x \in B(0;r)} |\nabla_{\alpha} \nabla_{\beta} (P_t f)(x)| \leq c_{t_0,r} e^{-\delta t} \|f\|_1; \tag{3.7}$$

for all $\alpha, \beta \in A_m$, all $t > t_0$ and for every f continuous and bounded.

Proof of Lemma 3.5. The proof is simple and can be done by following the same procedures presented in [14], using the modifications outlined in [14, comments after Corollary 4.9]. We do not repeat all the details here \square

Example 3.6 (UFG condition and Obtuse Angle Condition for linear SDEs). Consider SDEs in \mathbb{R}^N of the form

$$dX_t = (AX_t + D)dt + \frac{\rho_- X_t^d}{2} \sum_{i=1}^d C_i dB_t^i; \tag{3.8}$$

where A is a constant $N \times N$ matrix, B_t^1, \dots, B_t^d are one-dimensional standard Brownian motions and $D; C_1, \dots, C_d \in \mathbb{R}^N$ are constant vectors. In this case $V_0(x) = Ax + D; V_i(x) = C_i$, and

$$V_{[i, 0]} = [V_i; V_0] = AC_i; \quad i \in \{1, \dots, d\};$$

Because $[V_i; V_j] = 0$ for every $i, j \in \{1, \dots, d\}$, the only relevant commutators are those of the form $V_{[(i, 0, \dots, 0)]}$, i.e. repeated commutators with V_0 . For simplicity, let \vec{i}, k be the $(k + 1)$ -tuple such that $i_{1;k} = i$ and $i_{j;k} = 0$ for $j > 1$; then

$$V_{[(i, 0, \dots, 0)]} = A^k C_i;$$

$\underbrace{\hspace{1.5cm}}_{k \text{ times}}$

It is now easy to show that, *irrespective of the choice of $A; D; C_1, \dots, C_d$ as above, the UFG condition is always satisfied by SDEs of the form (3.8)*. Indeed, by the Cayley Hamilton Theorem there is a polynomial p of degree at most $N - 1$ such that $A^N = p(A)$; so we can write any $V_{[\vec{i}, k]}$ as a linear combination of the vectors $V_{[\vec{i}, j]}$ with $j \leq N$. For comparison we recall that (3.8) is hypoelliptic if and only if the Kalman rank condition is satisfied, namely if

$$\text{rank}[Q; AQ; A^2Q; \dots; A^{N-1}Q] = N;$$

where Q is the overall diffusion matrix of (3.8), see e.g. [40]. As for the OAC (3.4), this is satisfied if and only if there exists some $\epsilon > 0$ such that for all $f \in C^1(\mathbb{R}^N)$ we have

$$(r f)^T A^{k+1} C_i C_i^T (A^k)^T r f \geq \epsilon_0 (r f)^T A^k C_i C_i^T (A^k)^T (r f);$$

for all $i \in \{1, \dots, d\}$ and $k \in \{0, 1, \dots, N - 1\}$. This holds if and only if

$$(A + \epsilon_0 I)G \geq 0$$

for some $\epsilon_0 > 0$, where $G = A^k C_i C_i^T (A^k)^T$ for all $i \in \{1, \dots, d\}$ and $k \in \{0, 1, \dots, N - 1\}$. □

3.3 Geometry

In this section we cover some basic notions from differential geometry and geometric control theory on which the rest of the paper relies. Further details can be found in the excellent references [56, 28, 57]. For the reader who is already familiar with this material, we point out that, among the results recalled in this section, Theorem 3.13 is possibly the one which will play the most important role in the remainder of the paper.

Given a vector field $V(x)$ on \mathbb{R}^N , we denote by $e^{tV}(x)$ the integral curve of V starting at $t = 0$ from x , i.e. the curve $\gamma(t) : \mathbb{R} \rightarrow \mathbb{R}^N$ such that $\gamma(0) = x$ and $\dot{\gamma}(t) = V(\gamma(t))$ for all $t \in \mathbb{R}$ such that the curve is defined. In general, integral curves exist only locally. In this paper we consider only smooth, globally defined and globally Lipschitz vector fields (see Hypothesis 3.16), so integral curves actually exist for every $t \in \mathbb{R}$. As already mentioned, a *distribution* on \mathbb{R}^N is a map that, to each point $x \in \mathbb{R}^N$, associates a linear subspace of the tangent space $T_x \mathbb{R}^N$. Given a set D of smooth vector fields on \mathbb{R}^N , the distribution generated by D , denoted by \mathcal{D} , is the map $x \mapsto \text{span}\{X(x) : X \in D\}$. Distributions generated by a set of smooth vector fields are usually referred to as *smooth distributions*. When we write \mathcal{D} instead of just \mathcal{D} it is understood that we are considering smooth distributions rather than general distributions. As customary, we say that the vector

field X on \mathbb{R}^N belongs to the distribution μ if $X(x) \perp \mu(x)$ for all $x \in \mathbb{R}^N$. The rank of μ at x is the dimension of the vector space $\mu(x)$. A *piecewise integral curve*, γ , of vector fields in the set D is a curve of the form

$$(t_1; \dots; t_h) = e^{t_1 X_1} e^{t_2 X_2} \dots e^{t_h X_h} \quad h \in \mathbb{N}; t_j \in \mathbb{R}; x \in \mathbb{R}^N;$$

where $X_1; \dots; X_h \in D$ (and they are not necessarily all distinct). A submanifold $M \subset \mathbb{R}^N$ is an *integral manifold* of μ if $T_x M \subset \mu(x)$ for every $x \in M$. A *maximal integral manifold* (MIM) of μ , M , is a connected integral manifold of μ which is maximal in the sense that every other connected integral manifold of μ that contains M coincides with M . Therefore, two MIMs either coincide or they are disjoint.

Definition 3.7. Let μ be a distribution on \mathbb{R}^N .

- μ is involutive if
$$X; Y \in \mu \Rightarrow [X; Y] \in \mu;$$
- μ is invariant under the vector field V if the Jacobian matrix $J_x(e^{tV} x)$ maps $\mu(x)$ into $\mu(e^{tV} x)$ for all x and for all t .¹³
- Suppose μ is generated by the collection of vector fields $D = \{X_1; \dots; X_k\}$, i.e. $\mu = \langle D \rangle$. Then two points $x; y \in \mathbb{R}^N$ belong to the same orbit of μ if there exists a curve $\gamma: [a; b] \rightarrow \mathbb{R}^N$ such that $\gamma(a) = x$, $\gamma(b) = y$ and γ is a piecewise integral curve of vectors in D .

In general, the integral manifolds of a given distribution are “smaller” than the orbits; we refer the reader to [57] for a detailed explanation of this matter, see in particular [57, Eqn. (3.1)]. Here we just illustrate this fact with a simple but important classical example.

Example 3.8. In \mathbb{R}^2 , consider the vector fields $X = \partial_x$ and $Y = (x) \partial_y$ where (x) is a smooth function vanishing on the half-plane $x = 0$. The orbit of the distribution generated by X and Y , $\mu_{X;Y}$, is the whole \mathbb{R}^2 . That is, given any two points in \mathbb{R}^2 there is a piecewise integral curve of $\{X; Y\}$ which joins the two points. However the integral manifolds through points $(x; y)$ with $x > 0$ are one dimensional. Notice that the distribution in this example is involutive but it satisfies neither the Hörmander condition nor the UFG condition. More precisely, in the sense that whether we take $X = V_0$ and $Y = V_1$ or viceversa, either way the UFG condition is not satisfied (more precisely, in the language of Definition 3.10 below, the set $\{X; Y\}$ is neither locally nor globally of finite type). The fact that $\{X; Y\}$ do not satisfy the UFG condition can be either seen as a consequence of Theorem 3.13 below (if it did, the orbits would have to coincide with the integral manifolds) or it can be shown with direct calculations (the problem arising on the line $x = 0$). For the reader’s convenience this calculation is contained in the Appendix, see Lemma A.2. \square

We say that a distribution μ on \mathbb{R}^N satisfies the *(maximal) integral manifolds property* if through every point of \mathbb{R}^N there passes a (maximal) integral manifold of μ . The following fundamental result, due to Sussman (see [57, Theorem 4.2]), completely characterizes the distributions enjoying the maximal integral manifolds property.

Theorem 3.9 (Sussman’s Orbit Theorem). *If $\mu = \langle D \rangle$ is a smooth distribution on \mathbb{R}^N , then the following statements are equivalent*

- (a) μ satisfies the maximal integral manifolds property;
- (b) μ satisfies the integral manifolds property;

¹³A useful criterion to check whether a distribution is invariant under a vector field will be given in Note 3.11.

- (c) the orbits of \mathcal{D} coincide with the integral manifolds of the distribution and the rank of \mathcal{D} at each point $x \in \mathbb{R}^N$ is equal to the dimension of the integral manifold of \mathcal{D} through x ;
- (d) \mathcal{D} coincides with the smallest distribution which contains the Lie algebra generated by D , $Lie Dg$, and is invariant under the vectors in D .

In view of the equivalence of (a) and (b) above, when either property hold we just say that the smooth distribution is *integrable*. It is clear that in this case every integral manifold is a maximal integral manifold. Some standard facts about integrable distributions which are useful to bear in mind and that follow (easily) from what we have said so far: if \mathcal{D} is integrable, then

- i) \mathcal{D} is involutive;
- ii) the state space \mathbb{R}^N is partitioned into orbits of \mathcal{D} ;
- iii) the rank of the distribution is constant along the orbits (of \mathcal{D} , which coincide with the integral manifolds of such a distribution).

The latter fact is a consequence of the fact that \mathcal{D} is invariant under the vectors in D together with the observation that the maps e^{tV} are diffeomorphisms for every fixed $t \in \mathbb{R}$ (hence the Jacobian matrix $J_x(e^{tV}x)$, which maps the tangent space at x into the tangent space at $e^{tV}x$, is always invertible).

Definition 3.10 ([57 page 185]) *Let D be a set of everywhere defined, smooth vector fields on \mathbb{R}^N and \mathcal{D} be the associated distribution. The set D (as well as the distribution \mathcal{D}) is locally of finite type or locally finitely generated (LFG) if for every $x \in \mathbb{R}^N$ there exist vector fields X_1, \dots, X_k belonging to D and such that*

- i) $\text{span}\{X_1(x), \dots, X_k(x)\} = \mathcal{D}(x)$
- ii) for every $X \in D$ there exists a neighbourhood U of x and C^1 functions $'_{ij}$ defined on U such that

$$[X; X_j] = \sum_{j=1}^k '_{ij}(x) X_j(x) \quad \text{for all } x \in U \text{ and every } i \in \{1, \dots, k\};$$

We emphasize that if \mathcal{D} is LFG then the rank of \mathcal{D} need not be constant. Moreover, the choice of k in point i) does depend on x .

Note 3.11. We recall the following useful criterion (see [28, Lemma 2.1.4]): if a distribution is either of constant rank or locally of finite type, then it is invariant under a vector field V if and only if $[V; \cdot] \in \mathcal{D}$ whenever $\cdot \in \mathcal{D}$. \square

Definition 3.12. *With the same notation and setting as in Definition 3.10, D is globally of finite type if point i) of Definition 3.10 holds with k independent of x and if for every $X \in D$ there exist C^1 functions $'_{ij}$ defined on \mathbb{R}^N such that*

$$[X; X_j] = \sum_{j=1}^k '_{ij}(x) X_j(x) \quad \text{for all } x \in \mathbb{R}^N \text{ and every } i \in \{1, \dots, k\};$$

The next theorem gives a sufficient condition for integrability, which is easy to check in practice.

Theorem 3.13 (Hermann, Lobry, Stephan and Sussman) *If D is locally of finite type then \mathcal{D} is integrable; in particular, the integral manifolds of \mathcal{D} coincide with the orbits of the vector fields of the set D .*

Note 3.14 (Comments on Theorem 3.13) Seen from a control-theoretical point of view, the above statement gives a global decomposition of the state space \mathbb{R}^N into sets reachable by piecewise integral curves of vector fields in D . To clarify this fact and provide some context, it is useful to compare it with the case where the HC holds. Start by noting that under the HC the Lie algebra generated by the vectors in D is required to have constant rank (and the rank is assumed to be precisely N at every point). The control-theoretical meaning of the HC is expressed by Chow's Theorem, see [5, 8], (and indeed in control theory the HC is known as *Chow's condition*). Chow's theorem states that if the vectors f, V_0, \dots, V_d satisfy (HC) then any two points in \mathbb{R}^N are *accessible* or *reachable* in finite time from each other along integral curves of the vectors in D . That is, given any two points $x, y \in \mathbb{R}^N$, there exists a piecewise integral curve of vectors in D , and a time $t > 0$ such that $\phi(0) = x$ and $\phi(t) = y$. This is not the case if we simply assume that D is a LFG set of vector fields. According to the above theorem, if D is LFG then, for every $x \in \mathbb{R}^N$, the set of states reachable from x in finite time coincides with the maximal integral manifold of D through x . Because the rank of the distribution is not constant, and in particular it need not be N at any point, this implies that, in general, the orbits of D will be proper subsets of \mathbb{R}^N (as we have mentioned, they form a partition of \mathbb{R}^N). \square

We conclude this subsection by recalling the following result, which will be used later on.

Lemma 3.15 ([28, Theorem 2.1.9]) *Let D be a smooth involutive distribution invariant under a vector field W . Suppose D is locally finitely generated. Let x_1, x_2 be two points belonging to the same maximal integral manifold of D . Then, for all $t \in \mathbb{R}$, the points $e^{tW}x_1$ and $e^{tW}x_2$ belong to the same maximal integral submanifold of D .*

To clarify the above statement: under the assumptions of the lemma, if x_1, x_2 belong to a given MIM of D , say \mathcal{M} then $e^{tW}x_1, e^{tW}x_2 \in \mathcal{M}$, where \mathcal{M} denotes another generic MIM of D . In general \mathcal{M} will be different from \mathcal{M} (unless W belongs to D). For example see Example 4.10.

3.4 Assumptions

Throughout the paper we will make the following standing assumptions.

Hypothesis 3.16. *Standing assumptions:*

[SA.1] *All the vector fields we consider in this paper are smooth, everywhere defined and globally Lipschitz.*

[SA.2] *In this paper we will consider partitions of \mathbb{R}^N into submanifolds; each one of such submanifolds is generically denoted by S , see definition after Proposition 4.3. Throughout, the manifold topology (on S) is assumed to be the Euclidean topology of S , seen as a subset of \mathbb{R}^N ; that is, the open sets of S in the manifold topology are sets of the form $O \cap S$, where O is a Euclidean open set of \mathbb{R}^N . In Appendix A.1 we motivate the choice of such a topology and give further details about this assumption.*

[SA.3] *When we say that the obtuse angle condition (3.4) (or its second order version (3.6), respectively) is satisfied, we mean that it is satisfied for some $\epsilon_0 > 0$ large enough so that the estimate (3.5) ((3.7), respectively) follows.*

Note 3.17. Assumption **[SA.1]** will be needed mostly to make sure that all the integral curves of the involved vector fields are well defined (and to guarantee well-posedness of the SDE (1.1)). However see Note 4.14 on this point. \square

4 Geometrical significance of the UFG condition and implications for the corresponding SDE

In this section we come to explain how the general results outlined in Section 3.3 apply to the study of the dynamics (1.1), assuming that the vector fields $V_0; \dots; V_d$ satisfy the UFG condition. For clarity, we will compare with the case in which $V_0; \dots; V_d$ satisfy the Hörmander condition. Subsection 4.1 contains *global* results, Subsection 4.2 is focussed on *local* results.

4.1 Global results

Recalling the notation and nomenclature of Section 3.1 and motivated by Theorem 3.9, we introduce two distributions associated with the vector fields $V_0; \dots; V_d$; such distributions will play a fundamental role in the analysis of the UFG-dynamics (1.1). Let

- $\hat{\cdot}_0$ be the smallest distribution which contains the space $\text{span}fV_0; V_1; \dots; V_dg$ and is invariant under the vector fields $fV_0; V_1; \dots; V_dg$;
- $\hat{\cdot}$ be the smallest distribution which contains the space $\text{span}fV_1; \dots; V_dg$ and is invariant under the vector fields $fV_0; V_1; \dots; V_dg$.

Let us denote by $n = n(x)$ the rank of the distribution $\hat{\cdot}(x)$. Notice that $n = n(x)$ is a function of the point $x \in \mathbb{R}^N$ and, as such its value can vary from point to point. As Lemma 4.1 below demonstrates, if at some point $x \in \mathbb{R}^N$ the rank of $\hat{\cdot}$ is n , then the rank of $\hat{\cdot}_0$ is *at most* $n + 1$. We will typically assume that $n < N$, where N is the dimension of the state space \mathbb{R}^N in which the vector fields $V_0; \dots; V_d$ live, see Note 4.12 on this point. We stress that $\hat{\cdot}$ may not contain the vector field V_0 itself (unless for example V_0 is a linear combination of $V_1; \dots; V_d$). Lemma 4.1 below gives a simpler equivalent description of the distributions $\hat{\cdot}$ and $\hat{\cdot}_0$ (which is the one we gave in the introduction).

Lemma 4.1. *Let $V_0; \dots; V_d$ be $d + 1$ vector fields on \mathbb{R}^N which satisfy the UFG condition for some m , see Note 3.4. Recall the decomposition (1.9), the definition of R_m , given in (3.1), and set $R_{m,0} := R_m [V_0$. Then*

$$\hat{\cdot} = \text{span}fR_mg \quad \text{and} \quad \hat{\cdot}_0 = \text{span}fR_{m,0}g: \tag{4.1}$$

In particular,

$$\hat{\cdot}_0(x) = \text{span}(\hat{\cdot}(x); V_0^{(?)}(x)):$$

Proof of Lemma 4.1. A proof of this lemma in a general setting can be found in [28, Lemma 1.8.7 and Remark 2.2.3]. For completeness (and to spare the reader from having to compare and match notations and setting with those in [28]), we have added a proof in Appendix B.1. □

Note 4.2. If the vector fields $V_0; V_1; \dots; V_d$ satisfy the UFG condition then the distributions $\hat{\cdot}$ and $\hat{\cdot}_0$ are locally of finite type because they are *globally* of finite type (see Definition 3.12). This can be checked by using Note 3.4 (and the fact that nested commutators can always be expressed as linear combinations of hierarchical commutators, see [8, page 11-12]). □

Since the UFG condition implies that the sets R_m and $R_{m,0}$ are locally of finite type, we can apply Theorem 3.13 to the distributions given by the span of R_m and $R_{m,0}$. By Lemma 4.1, the distributions $\hat{\cdot}$ and $\hat{\cdot}_0$ coincide with span of R_m and $R_{m,0}$ respectively. As a corollary, we have the following proposition.

Proposition 4.3. *If the vector fields $V_0; V_1; \dots; V_d$ satisfy the UFG condition, then both $\hat{\cdot}_0$ and $\hat{\cdot}$ enjoy the integral manifolds property. In particular the integral manifolds of $\hat{\cdot}_0$ coincide with the orbits of $\hat{\cdot}_0$ (and the same holds for the distribution $\hat{\cdot}$).*

We denote by S (S_x , respectively) a generic MIM of the distribution $\hat{\mu}$ ($\hat{\mu}_x$, respectively). Consistently, S_x (S_x , respectively) will denote the MIM of $\hat{\mu}$ ($\hat{\mu}_x$, respectively) through the point $x \in \mathbb{R}^N$. It is easy to see that for every $x \in \mathbb{R}^N$, $S_x \cap S_x = \emptyset$, so that S_x is a disjoint union of integral manifolds of $\hat{\mu}$. Notice that $n = n(x)$ is constant along the orbits S of $\hat{\mu}$.

It is important to observe that any deterministic dynamics started on a maximal integral manifold S of $\hat{\mu}$ and following the integral curves of the fields $V_0; \dots; V_d$, will remain in S for any positive time (see Note 3.14). On the other hand, if $X_0 = x$ is the initial datum of the stochastic dynamics (1.1) and $X_0 \in S_x$, then $X_t \in S_x$ for all $t \geq 0$. This is a consequence of the Stroock and Varadhan support theorem, which we recall below, see [6] for more details.

Theorem 4.4 (Stroock and Varadhan). *Let $X_t^{(x)}$ be the solution of the SDE (1.1). The support¹⁴ of the law of $fX_t^{(x)} g_{t \in [0; T]}$ in path space, coincides with the closure in $(C([0; T]; \mathbb{R}^k); k_1)$ of the set of paths $(p_t)_{t \in [0; T]}$ such that (p_t) satisfies the control problem:*

$$dp_t = V_0(p_t)dt + \sum_{i=1}^d \frac{p_i}{2} V_i(p_t) \alpha_i(t)dt; \quad p_0 = x; \tag{4.2}$$

for some $\alpha_1; \dots; \alpha_d : [0; T] \rightarrow \mathbb{R}$ piecewise constant functions.

Informally, Theorem 4.4 says that the stochastic dynamics (1.1) will access in time t the (closure of) the set reachable in time t by the control problem (4.2), as we vary the controls $\alpha_1; \dots; \alpha_d$ in a suitable set of functions.

Excursus 4.5. We would like to further elaborate on the comment started before Theorem 4.4. To this end, consider the following one-dimensional SDE:

$$dX_t = \sin(X_t)dt + \cos(X_t) dB_t. \tag{4.3}$$

Here $V_0 = \sin(x)@_x$; $V_1 = \cos(x)@_x$, and $[V_0; V_1] = @_x$, so that these fields satisfy both the HC and the PHC. According to Chow's theorem (see Note 3.14), if $V_0; V_1$ satisfy the HC then any two points in \mathbb{R} can be joined through integral curves of such fields. However, if we start the dynamics (4.3) at $x \in [-\pi/2; \pi/2]$ then the solution X_t never leaves the interval $[-\pi/2; \pi/2]$. This is not in contradiction to the statement of Chow's theorem. The behaviour of the stochastic dynamics (4.3) is related to the control problem (4.2). On the other hand, when we say that under the HC any two points in \mathbb{R}^N can be joined by integral curves of vectors in D , this is equivalent to saying that the set of points reachable from x by the control system

$$dp_t = V_0(p_t) \alpha_0(t)dt + \sum_{i=1}^d V_i(p_t) \alpha_i(t)dt; \quad p_0 = x; \tag{4.4}$$

is indeed the whole space \mathbb{R}^N (in the above the functions $\alpha_1; \dots; \alpha_d : [0; T] \rightarrow \mathbb{R}$ are say piecewise constant controls). Clearly, the set of points accessible by (4.2) is a subset of the set of points accessible by (4.4). In our example, the support of the law of the solution to SDE (4.3) is given by the (closure of the) set of points reachable by the control problem

$$dX_t = \sin(X_t)dt + \cos(X_t) \alpha_1(t)dt;$$

On the other hand, Chow's theorem applied to the vector fields $V_0; V_1$ refers to the problem

$$dX_t = \sin(X_t) \alpha_0(t)dt + \cos(X_t) \alpha_1(t)dt;$$

¹⁴The support of the law of a random variable X denotes the smallest closed set A such that $P(X \in A) = 1$.

¹⁵This is a known example, see for example [24].

Such a dynamics can indeed be stirred to access the whole real line, no matter where it is started. \square

The theory summarised in Subsection 3.3 describes completely the sets accessible by the control problem (4.4), which are precisely the orbits of the vector fields V_0, \dots, V_d . On the other hand, if we want to study the SDE (1.1) (under the UFG condition) then we are interested in understanding the behaviour of the control problem (4.2). Unfortunately, in full generality, one can only state the following (see [28, Section 2.2]).

Lemma 4.6. *With the notation and nomenclature introduced so far, let V_0, \dots, V_d be smooth vector fields on \mathbb{R}^N satisfying the UFG condition. Then the sets of points reachable from x by the control problem (4.2) is a subset of $\overline{S_x}$ and it contains at least a non-empty open subset of $\overline{S_x}$.*

Combining the above and Theorem 4.4 we obtain the following.

Proposition 4.7. *Consider the SDE (1.1) with initial datum $X_0 = x$ and assume that the vector fields V_0, \dots, V_d satisfy the UFG condition. Then $X_t \in \overline{S_x}$ for every $t \geq 0$.¹⁶*

Let us reiterate that Proposition 4.7 does not say that $X_t^{(x)}$ will explore the whole set $\overline{S_x}$ (that is, it does not imply irreducibility of the process on $\overline{S_x}$), it simply means that the process X_t will not leave such a set.

Example 4.8. Here we consider the Stochastic Geodesic Equation derived in [2]. The aim of [2] is to study solutions $u(t; x)$ of the stochastic wave geodesic equation on the unit sphere:

$$du = \left(u + (|r|^2 - |j|^2)u - \frac{1}{2}u \right) dt + u \cdot u dB_t; \quad |j| = 1; \quad u(0; x) = u(0; x)$$

Here u denotes the time derivative of u , i.e. $u(t; x) = \partial_t u(t; x)$. In [2] the authors were concerned with solutions that are independent of the space variables, i.e. $u(t; x) = u(t) = u_t$. By introducing an auxiliary process v_t which is \mathbb{R}^3 valued and defined by $v_t := u_t$ they find that (u_t, v_t) satisfies the following 6-dimensional Stratonovich SDE:

$$d \begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} v_t \\ |v_t|^2 u_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ u_t \cdot v_t \end{pmatrix} dB_t; \tag{4.5}$$

In our notation, $X_t = (u_t, v_t)$, $N = 6$; $d = 1$ and for every $(u, v) \in \mathbb{R}^3 \times \mathbb{R}^3$ we have

$$V_0(u, v) = \begin{pmatrix} v \\ |v|^2 u \end{pmatrix}; \quad V_1(u, v) = \begin{pmatrix} 0 \\ u \cdot v \end{pmatrix};$$

We then have the following commutator relationships:

$$[V_1; V_0] = \begin{pmatrix} u \\ 0 \end{pmatrix}; \quad [[V_1; V_0]; V_0] = |v|^2 V_1; \quad [[V_1; V_0]; V_1] = V_0;$$

From here we can see that we will generate no new directions by taking further commutators and the distribution $\hat{\mu}$ satisfies the LFG condition, see Definition 3.10; moreover we have

$$\hat{\mu}_0(u, v) = \hat{\mu}(u, v) = \text{span} (V_1(u, v); [V_1; V_0](u, v); [[V_1; V_0]; V_1](u, v));$$

Observe the dimension of $\hat{\mu}_0$ is 3 except at the point $(0; 0)$ where all the vector fields vanish. Define the 3-dimensional manifold

$$M_{r,R} = \{ (u, v) : |v| = r; |u \cdot v| = R; u \perp v \}$$

¹⁶We clarify again that the closure is intended to be in the Euclidean topology.

Note that the tangent space to $M_{r,R}$ at the point $(u; v)$ is $T_{(u;v)}M_{r,R}$ and $M_{r,R}$ is closed. By Proposition 4.7, $X_t \in M_{r,R}$ almost surely when $X_t = (u_t; v_t)$ and $X_0 \in M_{r,R}$. In [2] the authors consider the case $R = 1$, i.e. $|juj| = 1$. Notice that since $M_{r,R}$ is compact, and for any initial condition $x \in M_{r,R}$ the process remains on $M_{r,R}$ almost surely we obtain that there exists an invariant measure with the support in $M_{r,R}$, that is we recover the results of [2, Corollary 5.5 & 5.11]. In [2, Theorem 8.1] the authors show that the transition probabilities are absolutely continuous with respect to the normalised surface measure on $M_{r,R}$ we will see that this will follow from Theorem 8.9.

Observe that these vector fields are not globally Lipschitz, however as the solutions always remain in a compact set for fixed initial conditions our results still hold.¹⁷ \square

4.2 Local considerations: an important change of coordinates

Let $x \in \mathbb{R}^N$ be a *regular point* of a given distribution μ , i.e. suppose there exists a neighbourhood of x where the dimension of μ is constant, say equal to n . If this is the case then, locally, there exist n linearly independent vector fields, fX_1, \dots, X_n generating the distribution. Suppose furthermore that D_n is involutive and $n < N$ (see Note 4.12). For some small enough $\epsilon > 0$ we can define the map $\phi : (t_1, \dots, t_n) \in \mathbb{R}^n \rightarrow \mathbb{R}^N$ as follows:

$$\phi : (t_1, \dots, t_n) \in \mathbb{R}^n \rightarrow \mathbb{R}^N$$

$$\phi(t) := (t_1 X_1 + \dots + t_n X_n) \in \mathbb{R}^N$$

where X_1, \dots, X_n are as above and X_{n+1}, \dots, X_N are such that

$$\text{span}\{X_1, \dots, X_n, X_{n+1}, \dots, X_N\} = \mathbb{R}^N$$

(at least locally). The map ϕ is, at least locally, a diffeomorphism on its image, so it admits an inverse, which we denote by ψ . Differentiating the obvious identity $(\phi \circ \psi)(t) = t$, one obtains

$$(J_{\phi \circ \psi})(t) = Id_{\mathbb{R}^n}$$

Let us make the above notation more explicit. The map ϕ is a map from (opens sets of) \mathbb{R}^n to (opens sets of) \mathbb{R}^N , i.e.

$$\phi : (t_1, \dots, t_n) \in \mathbb{R}^n \rightarrow \mathbb{R}^N$$

where $J_{\phi} : \mathbb{R}^n \rightarrow \mathbb{R}^N$. Therefore the i th row of the matrix J_{ϕ} is the gradient ∇_{t_i} . On the other hand, the j th column of the matrix J_{ψ} is the vector $\frac{\partial \psi}{\partial t_j} := \frac{\partial \psi_1}{\partial t_j}, \dots, \frac{\partial \psi_N}{\partial t_j}$. The first n columns of the Jacobian matrix $(J_{\psi})(t)$ are linearly independent (because ϕ is a diffeomorphism) and, from the above, we have

$$\nabla_{t_i} \psi_j = 0 \quad \text{for all } j = 1, \dots, n; i = n + 1, \dots, N; \tag{4.6}$$

By the involutivity of D_n the vectors $\frac{\partial \psi}{\partial t_j}$ belong to D_n ; moreover because they are linearly independent, they span D_n . Therefore the vectors ∇_{t_i} are orthogonal to

¹⁷Indeed, fix some initial conditions $u; v$ with $|juj| = 1; |vj| = 1$ and construct globally Lipschitz vector fields $V_0; V_1$ with the properties: $V_0(x) = V_0(u; v)$ and $V_1(x) = V_1(u; v)$ for any $x \in \mathbb{R}^6$ with $|jxj| \leq 2$. Let $(u_t; v_t)$ be the solution of

$$d \begin{pmatrix} u_t \\ v_t \end{pmatrix} = V_0(u_t; v_t) dt + V_1(u_t; v_t) dB_t$$

Then we have that $(u_t; v_t)$ take values in $M_{r,R}$ almost surely and in particular, $|ju_tj| = 1; |jv_tj| = r$ for all $t \geq 0$. However since $V_0 = V_0$ and $V_1 = V_1$ on the set $M_{r,R}$ we have that $u_t = u$ and $v_t = v$ by pathwise uniqueness of solutions to SDEs (see [29, Theorem 5.2.5]). Hence $(u_t; v_t)$ must take values in $M_{r,R}$ almost surely also.

¹⁸See e.g. [28, item (ii) on page 25]

every vector of D_n , i.e.

$$r^i = 0 \text{ for every } z \in D_n \text{ and for every } i = n + 1; \dots; N:$$

Now notice that ψ is (locally) invertible so it can be used as a (local) change of coordinates $z = \psi(x)$. With these preliminaries in place, we have the following.

Proposition 4.9. *Let μ be a smooth involutive distribution on \mathbb{R}^N and x_0 a regular point of μ . In particular, assume that there exists a neighbourhood of x_0 where the dimension of μ is n . Then there exists a change of coordinates ψ (defined locally) such that*

- i)** *A vector field V on \mathbb{R}^N belongs to μ if and only if in the coordinates defined by ψ , the last $N - n$ components of V are zero;¹⁹*
- ii)** *if μ is invariant under a vector field W then, in the coordinates defined by ψ , the last $N - n$ components of W are functions independent of the first n coordinates. More explicitly, as per notation introduced in (2.1), let*

$$z = (z^1; \dots; z^n; z^{n+1}; \dots; z^N) = (z^1; \dots; z^n; \cdot; a) = (x^1; \dots; x^N)$$

and let \tilde{W} be the representation of W in the new coordinates. Then

$$\tilde{W}(z) = (\tilde{W}^1(z); \dots; \tilde{W}^n(z); \tilde{W}^{n+1}(\cdot; a); \dots; \tilde{W}^N(\cdot; a)):$$

Proof of Proposition 4.9. The proof is deferred to Appendix B.1. □

We now want to apply Proposition 4.9 to the vector fields appearing in the SDE (1.1). We assume that such vector fields on \mathbb{R}^N satisfy the UFG condition for some m . Let $\hat{\mu}_0$ and $\hat{\mu}$ be the distributions defined at the beginning of Section 4. We know that the rank of $\hat{\mu}_0$ is constant along the orbits of $\hat{\mu}_0$ (see comment before Definition 3.10). Let $x \in \mathbb{R}^N$ and consider the orbit of $\hat{\mu}_0$ through x . In view of Lemma 4.1, if we assume that $V_0^{(2)}(x) \notin 0$ then the rank of $\hat{\mu}_0$ at x is exactly $n + 1$. Recall that N is fixed and it is the dimension of the state space \mathbb{R}^N , while $n = n(x)$ is the dimension of the orbits of $\hat{\mu}$ and it is constant along each one of such orbits. Notice that $\hat{\mu}$ (and $\hat{\mu}_0$) is also involutive by construction, so we can use it to apply Proposition 4.9.

With this in mind, let us describe the coordinate change. This is obtained by combining the following two steps.

Step one: because $\hat{\mu}_0 = \text{span}(\mathcal{R}_m; V_0)$ is the tangent space of an $(n + 1)$ -dimensional submanifold of \mathbb{R}^N one can always *locally* express the vector fields $V_0; \dots; V_d$ as

$$\mathcal{V}_j = (\mathcal{V}_j^1; \dots; \mathcal{V}_j^{n+1}; 0; \dots; 0); \quad j = 0; 1; \dots; d;$$

i.e. the last $N - (n + 1)$ coordinates of the vectors \mathcal{V}_j are simply zero.

Step two: apply Proposition 4.9 using the distribution $\hat{\mu}$ (possibly only to the first $n + 1$ coordinates of the involved fields). Then, because $V_1; \dots; V_d$ belong to $\hat{\mu}$ and V_0 is invariant for $\hat{\mu}$, one obtains, in the new local coordinates, (and recalling the notation introduced in Section 2)

$$\begin{aligned} \mathcal{V}_0 &= (\mathcal{V}_0^1(z); \dots; \mathcal{V}_0^n(z); \mathcal{V}_0^{n+1}(\cdot; a); 0; \dots; 0) \\ \mathcal{V}_j &= (\mathcal{V}_j^1(z); \dots; \mathcal{V}_j^n(z); 0; \dots; 0); \quad j = 1; \dots; d; \end{aligned}$$

¹⁹If V is any vector, then the vector $\tilde{V}(z) = [(J_\psi^{-1})^{-1} V(\psi^{-1}(z))]_{\psi^{-1}(z)}$ is the representation of V after the change of coordinates ψ . Indeed, if $\tilde{\gamma}(t) = e^{tV} x$ and $\gamma(t) = \psi^{-1}(\tilde{\gamma}(t))$ then the tangent vector to γ is precisely \tilde{V} .

where we keep the same notation V_j for the new representation of the vector fields after this further change of coordinates. This shows that, in the new coordinates, the vector fields $V_0; \dots; V_d$ take the form (2.2)–(2.3).

We now want to express the SDE (1.1) in the new local coordinates. If X_t is the original process, Z_t is the process in the new coordinates. In particular

$$Z_t = (Z_t; \dots; a_t);$$

where $Z_t \in \mathbb{R}^n$ contains the first n coordinates of Z_t , a_t is the $(n + 1)$ th coordinate of the process and a contains the remaining $N - (n + 1)$ components (which do not change in time, see below). Putting everything together and using the convention (2.2)–(2.3), one obtains that, in the new coordinates, the SDE (1.1) with initial datum $Z_0 = (z_0; \dots; a_0)$ is simply

$$Z_t = z_0 + \int_0^t U_0(Z_s; \dots; a_0) ds + \sum_{j=1}^d \int_0^t U_j(Z_s; \dots; a_0) dB_s^j \tag{4.7}$$

$$a_t = a_0 + \int_0^t W_0(Z_s; \dots; a_0) ds \tag{4.8}$$

$$a_t = a_0. \tag{4.9}$$

Notice that from the above one can also deduce that, in the new coordinates, $V_0^{(\wedge)} = (U_0; 0; \dots; 0)$ while $V_0^{(?) } = (0; \dots; 0; W_0; 0; \dots; 0)$. Assuming for the moment that at the initial point $x = X_0$ the dimension of \wedge_0 is exactly $n + 1$, the fact that the last $N - (n + 1)$ components of the dynamics remain constant reflects the fact that, at least for a short enough time, the solution of the SDE remains in the integral submanifold of \wedge_0 from which it started, coherently with Lemma 4.6 and Proposition 4.7.

If at the initial point the rank of \wedge_0 is exactly N , i.e. $n + 1 = N$, then one simply has

$$Z_t = z_0 + \int_0^t U_0(Z_s; \dots; a_0) ds + \sum_{j=1}^d \int_0^t U_j(Z_s; \dots; a_0) dB_s^j \tag{4.10}$$

$$a_t = a_0 + \int_0^t W_0(Z_s; \dots; a_0) ds; \tag{4.11}$$

and this time $V_0^{(\wedge)} = (U_0; 0; \dots; 0)$ while $V_0^{(?) } = (0; \dots; 0; W_0)$. In this simpler case it is clearer that we have locally reduced the SDE (1.1) to an ODE component, a_t (which evolves independently of all the other components) and an $(N - 1)$ -dimensional SDE. We emphasize that, because the change of coordinates is local, such a representation will hold only for small enough t .

Example 4.10 (UFG-Heisenberg) . Consider the following dynamics in \mathbb{R}^3

$$\begin{aligned} dX_t &= X_t dt \\ dY_t &= Y_t dt + \frac{\rho_-}{2} dW_t^2 \\ dZ_t &= 2Z_t dt + \frac{\rho_-}{2} Y_t dW_t^1 + \frac{\rho_-}{2} X_t dW_t^2 \end{aligned}$$

Here $V_0 = (x; y; 2z)$, $V_1 = (0; 0; y)$, $V_2 = (0; 1; x)$. This example was introduced in [14] and named the *UFG-Heisenberg dynamics* (as it comes from a modification of the Heisenberg group). This is already globally in the form ODE+SDE. The ODE for the first coordinate can be solved explicitly, giving $X_t = x_0 e^{-t}$. Therefore, if we start the dynamics at $(x_0; y_0; z_0)$ with $x_0 > 0$ ($x_0 < 0$, respectively), then the system evolves (at

least for finite time) in the semispace with positive x -coordinates (negative, respectively). If the initial datum is on the plane $(0; y_0; z_0)$ then the dynamics remains confined to such a plane for all subsequent times. This is coherent with the following: for the above set of vector fields, one has $\hat{\Lambda}_0((x; y; z)) \subset \mathbb{R}^3$ if $x > 0$ or $x < 0$ and $\hat{\Lambda}_0((x; y; z)) \subset \mathbb{R}^2$ when $x = 0$. The distribution $\hat{\Lambda}_0$ has three orbits, namely the sets

$$S_+ = \{f(x; y; z) \in \mathbb{R}^3 : x > 0\}; \quad S_- = \{f(x; y; z) \in \mathbb{R}^3 : x < 0\};$$

$$S_0 = \{f(x; y; z) \in \mathbb{R}^3 : x = 0\};$$

As for the distribution $\hat{\Lambda}$, this spans \mathbb{R}^2 at every point. Moreover, the orbit of $\hat{\Lambda}$ through the point $(b; y; z)$ is the plane $x = b$. For this reason, when working on this example we will simply denote by S_b the orbit through the point $(b; y; z)$. In particular, notice that $S_0 = S_b$. □

Example 4.11 (Random Circles). Consider the SDE

$$dX_t = Y_t dt + \sqrt{2} X_t dB_t \tag{4.12}$$

$$dY_t = X_t dt + \sqrt{2} Y_t dB_t; \tag{4.13}$$

where B_t is a one-dimensional Brownian Motion. This system satisfies neither the HC nor the PHC, however the UFG condition is satisfied at level $m = 1$. Indeed we have

$$V_0(x; y) = \begin{pmatrix} y \\ x \end{pmatrix}; \quad V_1(x; y) = \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{and} \quad [V_1; V_0] = 0;$$

For every $(x; y) \in \mathbb{R}^2$, $\hat{\Lambda}(x; y) = \text{span}\{V_1(x; y)\}$; except for the origin, the orbits of $\hat{\Lambda}$ are radial half-lines. That is, $S_{(x; y)} = \{(0; 0)\}$ if $(x; y) = (0; 0)$ and $S_{(x; y)} = \{f(sx; sy) : s > 0\}$ otherwise. Indeed, $S_{(x; y)}$ coincides with the set of points accessible by the integral curves of V_1 , which can be found explicitly:

$$e^{tV_1}(x; y) = \begin{pmatrix} x e^t \\ y e^t \end{pmatrix}; \quad t \in \mathbb{R};$$

Moreover, V_0 is orthogonal to V_1 , so $V_0^{(\hat{\Lambda})} = 0$ and $V_0^{(?) } = V_0$; therefore $\hat{\Lambda}_0(0; 0) = \{f(0; 0)\}$, $\hat{\Lambda}_0(x; y) = \mathbb{R}^2$ outside the origin, $S_{(x; y)} = \mathbb{R}^2 \setminus \{f(0; 0)\}$ if $(x; y) \notin (0; 0)$ and $S_{(0; 0)} = \{f(0; 0)\}$. In this example the local change of coordinates in the neighbourhood of $(1; 0)$ is given by the diffeomorphism

$$(1; 0)(x; y) = \left(\arctan \frac{y}{x}; \frac{1}{2} \log(x^2 + y^2) \right);$$

After such a change of coordinates, the SDE (4.12)–(4.13) can be expressed, locally, as

$$d\theta_t = dt \tag{4.14}$$

$$dZ_t = \sqrt{2} dW_t \tag{4.15}$$

Let $C_t = (X_t; Y_t) \in \mathbb{R}^2$. In Figure 1 below we plot the evolution of C_t , i.e. the solution of (4.12)–(4.13). From the plots it should be clear that $(\theta_t; Z_t)$ are just the polar coordinates of the point C_t : θ_t represents the angle, which evolves deterministically with a simple anticlockwise motion, while Z_t (or, to be more precise, $\exp(2Z_t)$) is the radius, which changes randomly according to the SDE (4.15).

Note 4.12. If the dimension n of $\hat{\Lambda}$ was equal to N for every $x \in \mathbb{R}^N$, this would imply that $\hat{\Lambda}(x) = \hat{\Lambda}_0(x)$ for every $x \in \mathbb{R}^N$. In particular, the Parabolic Hörmander Condition

(a) A plot of $(X_t; Y_t)$ for $0 \leq t \leq 2$.

(b) A plot of $(X_t; Y_t)$ for $0 \leq t \leq 1$.

(c) A plot of $(X_t; Y_t)$ for $0 \leq t \leq \frac{3}{2}$.

(d) A plot of $(X_t; Y_t)$ for $0 \leq t \leq 2$.

Figure 1: The process $(X_t; Y_t)$ of Example 4.11, with initial condition $(X_0; Y_0) = (1; 0)$. The angle of rotation evolves deterministically in counterclockwise sense, while the radius changes randomly, according to (4.15).

(PHC) would hold. This case is well studied in the literature and we do not wish to consider it here. For this reason many of the statements of this section are made under the assumption that $n < N$. We need to emphasize that it may happen that the two distributions coincide on a manifold (see Example 4.10, where the two distributions coincide on the plane $x = 0$) and it may also happen that they both have full rank N on a manifold, while they differ on other manifolds (for example, in \mathbb{R}^2 take $V_0 = 1_A \otimes \delta_x$, $V_1 = 1_{A^c} \otimes \delta_x$; $V_2 = 1_{A^c} \otimes \delta_y$ where A is the set $A = \{ (x; y) \in \mathbb{R}^2 : x \in [1; 1] \}$). The case that is not interesting to our purposes is the one in which they coincide and have full rank on the whole of \mathbb{R}^N . Most of our theorems do cover that case as well (unless otherwise explicitly stated); but they are not really conceived in that framework. \square

Note 4.13. The change of coordinates illustrated in this section will be an important technical tool throughout. We would like to point out how such a change of coordinates gives a different (and complementary) perspective on the smoothness results of Kusuoka and Stroock and of Crisan et al [33, 34, 35, 36, 10] that we mentioned in Introduction. As recalled in Section 1.1, in these works the authors show that if f is a continuous and bounded function then, under the UFG condition, the function $(P_t f)(x)$ is not necessarily smooth in every direction (as it would be the case under the Hörmander condition), but it is in general only smooth in the directions V_{l+1}, \dots, V_m . In particular, it may not be differentiable in the direction V_0 . In view of the decomposition (1.9) and of the change of

coordinates presented in this section, this result is quite intuitive, as we explain. By (1.9), it is clear that if $V_0^{(?) = 0$ then $(P_t f)(x)$ is differentiable in the direction V_0 (as in this case V_0 is a combination of the vectors in R_m) and, as a consequence, it is differentiable in t as well. The loss of smoothness happens if and only if $V_0^{(?) \notin 0$. For simplicity (and without any loss of generality), let us restrict to a manifold where $n + 1 = N$, so that the local change of coordinates gives (4.10)–(4.11). As already observed, the representation of $V_0^{(?)$ in the new coordinates is given by $V_0^{(?) = (0; \dots; 0; W_0)$, where W_0 is the function driving the ODE component. Hence $V_0^{(?)$ is inherently linked to the deterministic part of the system, which clearly does not provide any smoothness. This also explains why, while there is no smoothness in the direction V_0 , the semigroup will always be smooth in the direction ∂_{V_0} (to be more precise, in the direction $\partial_{V_0^{(?)}$), as solutions of the ODE are constant in this direction. Finally, the deterministic part of the dynamics is responsible for the lack of density (i.e. for the fact that the law of the process does not admit a density on R^N). It is useful to the purposes of this discussion to point out that the one-dimensional transport equation is an extreme example of UFG condition; that is, consider the PDE $\partial_t u(t; x) = \partial_x u(t; x); (t; x) \in R_+ \times R^N$, with initial datum $u(0; x) = f(x)$. Here $V_0 = \partial_x$. As is well known, the solution to such a PDE is just $u(t; x) = f(x + t)$, hence no smoothing occurs in the space direction. However the solution is smooth in the direction $(\partial_x \partial_x) = \partial_x V_0$, as it is constant in such a direction. Therefore, UFG diffusions include a vast range of behaviours, from smooth elliptic diffusions to deterministic equations. □

Note 4.14. A final note on a technical point: as we have emphasized, to avoid having problems with the well-posedness of the integral curves, we work under the standing assumption [SA.1]. After the change of coordinates the coefficients of the vector fields (in the new coordinates) may grow more than linearly, but they will still be smooth. Hence, in the neighbourhood in which they are defined, the vector fields will still be locally Lipschitz. The situation is more delicate with the vector $V_0^{(?)$: if V_0 is smooth, this is not the case for $V_0^{(?)$ as well, see Example 7.11. Whenever this may cause issues, we will assume that $V_0^{(?)$ is at least such that the integral curve of $V_0^{(?)$ through a given point is unique and well defined (at least on given manifolds). □

We conclude this section by stating a couple of technical lemmata which will be useful in the following.

Lemma 4.15. Assume the vector fields $V_0; \dots; V_d$ satisfy the UFG condition. Let S be a maximal integral manifold of Λ_0 and S be an integral submanifold of Λ such that $S \subset S$. Then $\partial S = S \cap S$ is contained within $\mathcal{S} := S \cap S$.²⁰

Proof of Lemma 4.15. The proof is deferred to Appendix B.1. □

The statement of Lemma 4.15 would clearly not be true if S and S were two arbitrary sets, it only holds because of the particular structure of the integral manifolds of Λ and Λ_0 . As a side remark, notice that while $S \subset S$ implies $\partial S \subset \mathcal{S}$, it is not the case, in general, that the boundary of S is the union of boundaries of orbits of Λ , see Example 4.10.

Lemma 4.16. With the notation introduced so far, assume the vector fields $V_0; \dots; V_d$ satisfy the UFG condition. Let $x_0 \in R^N$ and recall that x_0 belongs to exactly one integral manifold of Λ_0 , the manifold S_{x_0} . Consider the vector field $V_0^{(?)$ (defined in (1.9)) and assume such a vector field is smooth. Then either $V_0^{(?) (y) = 0$ for every $y \in S_{x_0}$ or $V_0^{(?) (y) \notin 0$ for every $y \in S_{x_0}$.

²⁰Closures are meant in the Euclidean topology, see Appendix A.1.

Proof of Lemma 4.16. The proof is deferred to Appendix B.1. □

5 Qualitative results on UFG diffusions

In this section we study the behaviour of the diffusion X_t (1.1) under the sole assumption that the vector fields $V_0; \dots; V_d$ appearing in (1.1) satisfy the UFG condition. As observed also in [14, Note 4.3], under the sole UFG condition one cannot expect to make any quantitative deductions on the behaviour of the process X_t . Neither can one expect the UFG condition itself to imply any results about existence or uniqueness of invariant measures, as there are many elliptic diffusions that do not have an invariant measure (the simplest example being Brownian motion on \mathbb{R}). In order to study invariant measures and decay to equilibrium we will have to make further assumptions. Nonetheless, the geometric considerations made in the previous sections allow us to prove several qualitative statements on the behaviour of the diffusion. The main results of this section are Proposition 5.1, Proposition 5.3 and Proposition 5.7. Collectively, these three results impart a lot of intuition about UFG dynamics and contain a lot of useful information. After each one of these three statements we have inserted a note to comment on the meaning of these propositions, see Note 5.2, Note 5.4 and Note 5.8. The results of Section 6 and Section 7 heavily rely on the statements of this section.

Recall that we denote by S (S_x , respectively) a generic integral manifold of the distribution Δ (Δ_x , respectively). Consistently, S_x (S_x , respectively) denote the integral manifold of Δ (Δ_x , respectively) through the point $x \in \mathbb{R}^N$.

Proposition 5.1. Assume that the vector fields $V_0; V_1; \dots; V_d$ satisfy the UFG condition and let X_t be the solution of the SDE (1.1). Let S be a maximal integral manifold of Δ and let ∂S be the boundary of S , i.e. $\partial S := S \cap S^c$. Then the following holds:

- i) If ∂S is not empty, it is a union of integral submanifolds of Δ_0 ;
- ii) If $X_0 = x \in \partial S$ then $X_t \in \partial S$ for all $t > 0$ (almost surely).

Proof of Proposition 5.1. The proof is deferred to Appendix B.1. □

Note 5.2. Let us explain the meaning and consequences of Proposition 5.1. Suppose we start the SDE (1.1) at $x \in \mathbb{R}^N$. Because the integral manifolds of Δ_0 partition \mathbb{R}^N , x belongs to one of such integral manifolds, the one which we denote by S_x . As a consequence of Proposition 4.7 we know that the process will never leave the closure of S_x ; however, if it started in the interior, it could in principle hit the boundary (which is a manifold whose dimension is lower than the dimension of S_x) and then come back to the interior. What we prove here is that this is not possible. Furthermore, because the boundary of S_x is itself a union of integral manifolds of Δ_0 , one could repeat the previous reasoning once the process enters the boundary (if this is the case). As a result of iterating this line of thought, we have that, along the path of $X_t^{(x)}$, the rank of the distribution Δ_0 can only decrease (or stay the same). In other words, we have shown that for every $x \in \mathbb{R}^N$ and $t \geq 0$, one has

$$\text{rank}(\Delta_0(X_u^{(x)})) \leq \text{rank}(\Delta_0(X_t^{(x)})) \quad \square$$

Before stating the next result we recall that the vector $V_0^{(?)}$ has been defined in (1.9). We also recall our assumption (see Note 4.14) that $e^{tV_0^{(?)}}(x)$ is well defined for all $x \in \mathbb{R}^N$ and $t \geq 0$.

Proposition 5.3. Let X_t be the solution of the SDE (1.1) with initial condition $x_0 \in \mathbb{R}^N$. If the vector fields $V_0; V_1; \dots; V_d$ appearing in (1.1) satisfy the UFG condition then for all $t \geq 0$

$$X_t^{(x_0)} \in \bar{S}_{e^{tV_0^{(?)}}(x_0)}; \text{ almost surely } ;$$

We clarify that $\overline{S_{e^{tV_0^{(?)}}(x_0)}}$ is the closure (in the Euclidean topology) of the integral manifold of $\hat{\Lambda}$ through the point $e^{tV_0^{(?)}}(x_0) \in \mathbb{R}^N$.

Proof of Proposition 5.3. If $V_0^{(?)}(x_0) = 0$ then the result follows immediately from Proposition 4.7 and Lemma 4.1. Indeed, by Proposition 4.7 we know that $X_t^{(x_0)} \in S_{x_0}$ and by Lemma 4.1 (and Lemma 4.16) we have $S_{x_0} = \overline{S_{x_0}}$. So we only need to treat the case $V_0^{(?)}(x_0) \neq 0$. This will be done by considering the control problem associated with the SDE (1.1) and by using Stroock and Varadhan Support Theorem. We postpone this part of the proof to Appendix B.1. □

Note 5.4. Proposition 5.3 clarifies the pivotal role of the vector $V_0^{(?)}$. To convey more intuition about the role of $V_0^{(?)}$, let us assume that $V_0^{(?)}(x) \neq 0$ for every $x \in S_{x_0}$, x_0 being the starting point of the SDE (1.1). We already know by Proposition 4.7 that $X_t^{(x_0)}$ will not leave S_{x_0} , so that we can consider S_{x_0} to be the state space of the dynamics. As already observed before Proposition 4.3, every $x \in S_{x_0}$, belongs to exactly one orbit S of $\hat{\Lambda}$ and, moreover, the union of the manifolds $\bigcup_{x \in S_{x_0}} S_x$ gives precisely S_{x_0} . In other words, the orbits of $\hat{\Lambda}$ that belong to S_{x_0} partition S_{x_0} . Furthermore, because $V_0^{(?)}(x) \neq 0$ on S_{x_0} and the rank of $\hat{\Lambda}_0$ is constant on S_{x_0} , one has (see Lemma 4.1) that if S_{x_0} has rank $n + 1$ then every orbit $S_x; x \in S_{x_0}$, will be a manifold of dimension n . In particular, there is no $x \in S_{x_0}$ such that $S_x = S_{x_0}$ (so that the partition of S_{x_0} into orbits of the distribution $\hat{\Lambda}$ is not the trivial one). With this premise, it makes sense to ask the following question: while we know that the process will not leave $\overline{S_{x_0}}$ for every $t \geq 0$, if we fix an arbitrary positive time $t > 0$, can we tell more precisely where, within S_{x_0} , $X_t^{(x_0)}$ is? In particular, can we determine which submanifold S it belongs to, i.e. which element of the partition of S_{x_0} is visited at time $t \geq 0$? The answer, given by Proposition 5.3, is the following: let $y = e^{tV_0^{(?)}}(x_0)$. Then, while $x_0 \in S_{x_0}$, $X_t \in \overline{S_y}$. In other words, the vector $V_0^{(?)}$ will make the SDE move from one submanifold of the partition (of S_{x_0}) to another. Another question is whether it is possible that X_t will visit one of such submanifolds twice or whether it is the case that, once one of these submanifolds has been visited, it will never be hit again. Example 5.6 below shows that the submanifolds of the partition can be visited an arbitrary number of times. □

Example 5.5. Recall the UFG-Heisenberg SDE introduced in Example 4.10. In this case $V_0^{(?)}(x_0) = (x_0; 0; 0)$ and, as we have already mentioned, $S_{(x_0; y_0; z_0)}$ is the plane $S_{(x_0; y_0; z_0)} = \{f(x; y; z) : x = x_0\}$. If $V_0^{(?)}(x_0) = (x_0; 0; 0)$ then the integral curve of $V_0^{(?)}$ through $(x_0; y_0; z_0)$ is $e^{tV_0^{(?)}}(x_0; y_0; z_0) = (e^{tx_0}; y_0; z_0)$ so that

$$S_{e^{tV_0^{(?)}}(x_0; y_0; z_0)} = \{f(x; y; z) \in \mathbb{R}^3 : x = e^{tx_0}\}$$

It is therefore clear that if $(X_0; Y_0; Z_0) = (x_0; y_0; z_0)$ then $(X_t; Y_t; Z_t) = (x_0 e^{t}; Y_t; Z_t) \in \overline{S_{(e^{t}x_0; y_0; z_0)}}$.

Example 5.6 (Random Circles, Example 4.11, continued). Let us go back to Example 4.11. Consider the integral curve of $V_0^{(?)}$, namely

$$e^{tV_0^{(?)}}(x; y) = e^{tV_0^{(?)}}(x; y) = \begin{pmatrix} x \cos(t) - y \sin(t) \\ x \sin(t) + y \cos(t) \end{pmatrix} \quad (5.1)$$

To fix ideas, let $(x_0; y_0) = (1; 0)$ be the initial condition of the SDE; then the integral curve of $V_0^{(?)}$ through $(x_0; y_0) = (1; 0)$ is the unit circle:

$$e^{tV_0^{(?)}}(1; 0) = (\cos(t); \sin(t))$$

and $S_{e^{iv_0}(\cdot)}(1;0) = S_{(\cos(t); \sin(t))}$ is the (open half) radial line at an angle t from the x -axis; that is, it is the (open half) radial line that intersects the unit circle at the point $(\cos(t); \sin(t))$. On the other hand the solution of the SDE with initial datum $(x_0; y_0) = (1; 0)$ is given by

$$X_t^{(x_0; y_0)} = \cos(t) e^{p \sqrt{2} B_t}; \tag{5.2}$$

$$Y_t^{(x_0; y_0)} = \sin(t) e^{p \sqrt{2} B_t}; \tag{5.3}$$

Therefore one can again explicitly verify that for every $t > 0$, $(X_t^{(x_0; y_0)}; Y_t^{(x_0; y_0)})$ belongs to $\bar{S}_{(\cos(t); \sin(t))}$. □

Proposition 5.7. With the notation introduced so far, assume the vector fields $V_0; \dots; V_d$ satisfy the UFG condition and let

$$E_t := \{x \in \mathbb{R}^N : P_x(X_t^{(x)} \not\subseteq S_x) > 0\}$$

and

$$E := \{x \in \mathbb{R}^N : P_x(X_t^{(x)} \not\subseteq S_x) > 0 \text{ for some } t > 0\}.$$

Then, for any invariant measure of the SDE (1.1) (should at least one exist), we have $(E_t) = 0$ for every $t > 0$. As a consequence, $(E) = 0$ as well.

Proof of Proposition 5.7. The proof is deferred to Appendix B.1. □

Note 5.8. Informally, Proposition 5.7 says that any invariant measure (should at least one exist) gives zero weight to the set of points that, under the action of the dynamics prescribed by the SDE (1.1), leave in finite time the submanifold from which they start. That is, the set of points x such that $X_t^{(x)} \not\subseteq S_x$ for some time $t > 0$, has μ -measure zero. In view of Proposition 5.1 this result is intuitive: in general, if the dynamics leaves a set it can return in finite many times to that set (when this happens the set is said to be recurrent). Because along the trajectories of $X_t^{(x)}$ the rank of the distribution can only decrease, if the process $X_t^{(x)}$ leaves the integral manifold S_x from which it started, it will never return to it. The dynamics will therefore spend an infinite amount of time outside the manifold S_x , so that the invariant measure, if it exists, can only be supported outside such a manifold. In other words, the theorem says that an integral submanifold S is a recurrent set if and only if the process never leaves it (once it enters it). This argument constitutes an informal proof of the theorem. Notice also that this theorem does not say anything about say geometric Brownian motion (see Example 3.2) or the UFG-Heisenberg process of Example 4.10, as such dynamics only leave the initial submanifold in finite time; for any finite time they stay in the submanifold from which they started. □

Lemma 5.9. Assume the vector fields $V_0; \dots; V_d$ appearing in (1.1) satisfy the UFG condition and that the Obtuse Angle Condition (3.4) holds. Let P_t be the semigroup defined in (1.2). Then, given a maximal integral submanifold S of $\hat{\Lambda}$, we have

$$\lim_{t \downarrow 0} \int P_t f(x) P_t f(y) dx dy = 0; \tag{5.4}$$

for all $f \in C_b(\mathbb{R}^N)$ and $x, y \in S$.

Proof of Lemma 5.9. The proof is deferred to Appendix B.1 □

Proposition 5.10. Consider the assumptions and setting of the previous lemma and let S be a maximal integral manifold of $\hat{\Lambda}$. Then, among all the invariant measures of (1.1) (assuming at least one such measure exists), there exists at most one such that $(S) = 1$.

Moreover, if such a measure exists, then it is ergodic (in the sense that $P_t 1_E = 1_E$ for some Borel set E , implies that $\mu(E) = 1$ or 0) and for every $x \in S$ and $f \in C_b(\mathbb{R}^N)$ we have

$$P_t f(x) = \int_S f(y) \mu(dy) \quad \text{as } t \rightarrow \infty \quad (5.5)$$

Proof of Proposition 5.10. The proof is deferred to Appendix B.1. □

6 Long-time behaviour of UFG processes: the case of “non-autonomous hypoelliptic diffusions”

In this section we set $N = n + 1$ and study stochastic dynamics in $\mathbb{R}^N = \mathbb{R}^{n+1}$ of the form

$$dZ_t = U_0(Z_t; t)dt + \sum_{j=1}^n U_j(Z_t; t) dB_t^j; \quad (6.1)$$

$$d\tau = W_0(\tau)dt \quad (6.2)$$

$$Z_0 = z; \quad \tau_0 = \tau; \quad (6.3)$$

In other words, we consider systems for which the representation of the form “ODE + SDE” (4.10)–(4.11) is global.²¹ The above system consists of an n -dimensional process, $Z_t \in \mathbb{R}^n$, satisfying an SDE, equation (6.1), which is coupled with a one-dimensional autonomous ODE, (6.2). As in previous sections, $U_j : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n; j = 1, \dots, n$ and $W_0 : \mathbb{R} \rightarrow \mathbb{R}$. The evolution of Z_t depends on the evolution of τ_t , but the ODE solution τ_t evolves independently of the SDE. For the purposes of this paper, we do not think of τ_t as representing time, but rather as representing an additional space-coordinate. However notice that if $W_0 = 1$ and $\tau(0) = 0$ then $\tau_t = t$ and we recover a standard time-inhomogeneous setting, i.e. in this case (6.1) becomes a general time-inhomogeneous SDE, namely

$$dZ_t = U_0(Z_t; t)dt + \sum_{j=1}^n U_j(Z_t; t) dB_t^j; \quad (6.4)$$

Going back to the representation of the form “ODE + SDE” (6.1)–(6.2) under consideration, if we denote by X_t the process $\mathbb{R}^N \ni X_t = (Z_t; \tau_t)$, then X_t is the solution of an autonomous SDE. The one-parameter semigroup associated to X_t is, as usual, given by

$$(P_t f)(x) := E[f(X_t) | X_0 = x] = E[f(Z_t; \tau_t) | (Z_0; \tau_0) = (z; \tau)]; \quad x = (z; \tau) \in \mathbb{R}^{n+1};$$

for any $f \in C_b(\mathbb{R}^{n+1}; \mathbb{R})$. On the other hand one could consider the two-parameter semigroup associated with the non-autonomous process Z_t alone. Indeed, if we solve the ODE for τ_t and substitute the solution back into the SDE for Z_t , then we can simply consider equation (6.1) rather than the whole system. To be more precise, let us denote by τ_t the solution at time t of (6.2) with initial datum $\tau(0) = \tau$. That is, $\tau_t = e^{tW_0}$. Let also $Z_t^{s; z}$ be the solution of the following SDE:

$$Z_t = z + \int_s^t U_0(Z_u; u)du + \sum_{j=1}^n \int_s^t U_j(Z_u; u) dW_u^j;$$

The two-parameter semigroup associated with the above non-autonomous SDE is given by

$$(Q_{s;t} g)(z) := E[g(Z_t^{s; z})]; \quad z \in \mathbb{R}^n; s \leq t;$$

²¹We are not claiming that this representation necessarily results from the change of coordinates presented in Section 4.2.

We emphasize that this two-parameter semigroup depends on z_0 , i.e. on the initial datum of the ODE. When we do not wish to stress this dependence we may just write $Q_{s,t}$. With this notation, one can equivalently rewrite the definition of P_t as

$$(P_t f)(x) = E^h \left[f(Z_t^{0;z}; \cdot) \right]; \quad x = (z; \cdot) \in \mathbb{R}^{n+1}; \quad (6.5)$$

To make explicit the relation between the two-parameter semigroup $Q_{s,t}$ and the one-parameter semigroup P_t , $x \in \mathbb{R}$ and let $\hat{z} = e^{sW_0}$. Notice that

$$Z_t^{0;z; \hat{z}} = Z_{t+s}^{s;z; \hat{z}}$$

where the equality is intended in law. Therefore, for every $f \in C_b(\mathbb{R}^{n+1}; \mathbb{R})$, and $z \in \mathbb{R}^n$, we have

$$\begin{aligned} (P_t f)(z; \hat{z}) &= E^h \left[f(Z_t^{0;z; \hat{z}}; \cdot) \right] = E^h \left[f(Z_t^{0;z; \hat{z}}; \cdot) \right] \\ &= E^h \left[f(Z_{t+s}^{s;z; \hat{z}}; \cdot) \right]; \end{aligned}$$

Hence,

$$(P_t f)(z; \hat{z}) = (Q_{s;t+s} f(\cdot; \hat{z}))(\hat{z}); \quad (6.6)$$

On the right hand side of the above we mean to say that the semigroup Q is acting on the function $f(\cdot; a)$ obtained by freezing the value of the last coordinate of the argument.

From now on, unless otherwise specified, we write Z_t for $Z_t^{0;z; z_0}$. With this set up in place, we can start commenting on the long-time behaviour. Heuristically, if the solution of the ODE (6.2) is unbounded, then one can not expect the process X_t to have an invariant measure (see Proposition 6.8) – though the process Z_t may still admit an invariant measure. So we restrict to the case in which the solution of the ODE is bounded. However, because (6.2) is a one-dimensional time-homogeneous ODE, if Z_t is bounded then it can only either increase or decrease towards stable stationary points of the dynamics (a stationary point of the ODE (6.2) is a point $z \in \mathbb{R}$ such that $W_0(z) = 0$). We emphasise that there may be many such points. For these reasons, we work under the assumption that Z_t admits a finite limit, i.e. we assume that the initial datum $z_0 \in \mathbb{R}$ is such that there exists a point $z^* \in \mathbb{R}$ such that

$$Z_t \xrightarrow{t \rightarrow \infty} z^* \quad \text{as } t \rightarrow \infty; \quad (6.7)$$

As customary, the notation $z^* = z^*(z_0)$ is to emphasise the fact that the limit point will depend on the initial datum (when we do not wish to stress such a dependence we just denote a stationary point of the ODE by z^*). The dynamics (6.1)–(6.2) will, in general, admit several invariant measures. As pointed out in the introduction, when this is the case, it is typically extremely difficult to determine the basin of attraction of each invariant measure. However in the setting of this section the basin of attraction of a given invariant measure will only depend on the behaviour of the ODE. (In the next section we will show that, despite the fact that the representation of the form “ODE + SDE” is only local for generic UFG processes, it is still the case that we can relate in a simple way the initial datum to the invariant measure to which the process is converging). Given an initial datum z_0 for (6.2), let $z^* = z^*(z_0)$ be the corresponding limit point of the ODE dynamics, as in (6.7). Consider the SDE

$$dZ_t = U_0(Z_t; \cdot) dt + \sum_{j=1}^d U_j(Z_t; \cdot) dB_t^j; \quad Z_0 = z; \quad (6.8)$$

with associated semigroup

$$(Q_t g)(z) := E[g(Z_t) | Z_0 = z]; \quad z \in \mathbb{R}^n; g \in C_b(\mathbb{R}^n) :$$

We will assume that the dynamics (6.8) is hypoelliptic, see Hypothesis [H.1] below for a more precise statement of assumptions. Moreover, under Hypothesis [H.2], the semigroup Q_t admits a unique invariant measure, $\mu = (\cdot; \rho)$ (see Lemma 6.4). We emphasise that the asymptotic behaviour of Z_t is independent of the initial datum z , see Lemma 6.4.

In view of (6.7), it is reasonable to guess that the asymptotic behaviour of $Z_t = Z_t^{0; z; \rho}$ is the same as the asymptotic behaviour of Z_t which is the solution of (6.8). This is the content of Theorem 6.5 below. Theorem 6.5 and Theorem 6.6 are the main results of this section; the former is concerned with the asymptotic behaviour of the semigroup $Q_{s;t}$, the latter describes the related asymptotic behaviour of the semigroup P_t . We set ρ the assumptions used in the rest of this section and we comment on their significance in Note 6.2.

Hypothesis 6.1. With the notation introduced so far, we will consider the following assumptions:

[H.1] The vector fields $V_0 = (U_0; W_0); V_1 = (U_1; 0); \dots; V_d = (U_d; 0)$ satisfy the UFG condition for some $m \geq 1$; moreover,

$$\text{span} \{ R_m g = \text{span} \{ V_i \}_1(x) : \exists A_m g' \in \mathbb{R}^n; \text{ for every } x \in \mathbb{R}^{n+1} : \}$$

[H.2] Define the measures $q_t^{s;z}$ by $q_t^{s;z}(A) := Q_{s;t} 1_A(z)$, for any Borel measurable $A \subset \mathbb{R}^n$. Then we require that for each $z \in \mathbb{R}^n$ the family of measures $\{ q_t^{0;z} : t \geq 0 \}$ on \mathbb{R}^n is tight.

[H.3] The Obtuse angle condition (3.4) is satisfied (with the understanding of Hypothesis 3.16 [SA.3]).

[H.4] The ODE (6.2) has at least one stationary point and the initial datum $\rho \in \mathbb{R}$ of (6.2) is such that (6.7) holds, for some limit point $\rho = (\rho)$.

Note 6.2. Some comments on the above assumptions, in particular on Hypothesis [H.1].

- We start by remarking on the obvious fact that if $X_t = (Z_t; \rho)$, where $Z_t; \rho$ are as in (6.1)–(6.2), then X_t solves an SDE of the form (1.1), with $V_0 = (U_0; W_0); V_1 = (U_1; 0); \dots; V_d = (U_d; 0)$.
- With the notation of Section 4 and Section 5, assumption [H.1] implies that the distribution $\hat{\mu}(x)$ is n -dimensional for every $x \in \mathbb{R}^N$, with $n = N - 1$. In the setting of this section, this is the maximum rank that the distribution $\hat{\mu}$ can have (as $V_0 = (U_0; W_0)$ is not contained in \mathbb{R}_m when $W_0 \notin 0$). In other words, for every $x \in \mathbb{R}^N$, the integral manifolds S_x of $\hat{\mu}(x)$ are $(N - 1)$ -dimensional manifolds. Because of the particularly simple structure of the SDE, such manifolds are just hyperplanes: for $x = (z; \rho)$, $S_x = S_{(z; \rho)} = \{ u \in \mathbb{R}^{n+1} : u = (z; \rho); \rho = \rho; z \in \mathbb{R}^n \}$. In this explicit setting Proposition 5.3 is easy to check.
- To reconcile the present work with the framework of [9] and further elaborate on the meaning of Hypothesis [H.1], let us assume for the moment that $W_0 = 1$ and that $\rho = 0$, so that (6.1) becomes a standard time inhomogeneous SDE of the form (6.4). In this case the vector fields $U_0; \dots; U_d$ are \mathbb{R}^n -valued maps whose coefficients depend on time, i.e. $(z; t) \mapsto U_j(z; t) \in \mathbb{R}^n$. For simplicity, let also $n = 1$.

Then V_0 acts both on space and time, while $V_1; \dots; V_d$ act on the space coordinate z only. That is, $V_0 = U_0(z; t) \partial_t + \sum_{j=1}^d U_j(z; t) \partial_{z_j}$ while $V_j = U_j(z; t) \partial_{z_j}$ for $j = 1; \dots; d$, so that

$$[V_0; V_j] = [U_0; U_j] + (\partial_t U_j(z; t)) \partial_{z_j} \quad (6.9)$$

One can then rephrase Hypothesis [H.1] just in terms of the vector fields $U_0; \dots; U_d$; from (6.9) it is then clear that Hypothesis [H.1] is equivalent to assuming that the Lie algebra

$$\left[\sum_{k=1}^d \text{span} L_k^U(z; t) \right]$$

where $L_1^U(z; t) := f[U_1(z; t); \dots; U_d(z; t)]g$ and, for $k > 1$, $L_k^U(z; t) := f[U; U_j; U_2; \dots; U_{k-1}; U_{k+1}; \dots; U_d; U_0 + \partial_t U_j; U_2; \dots; U_{k-1}]g$, should be nilpotently generated and span \mathbb{R}_z^n , for every $(z; t) \in \mathbb{R}^n \times \mathbb{R}$.

Let us now go back to the general representation of the form “ODE + SDE” (6.1) – (6.2), without assuming $W_0 = 1$. Recall that in this context the vector fields U_j are \mathbb{R}^n -valued functions of $n + 1$ variables; that is, we view them as maps $\mathbb{R}^n \times \mathbb{R} \ni (z; t) \mapsto U_j(z; t) \in \mathbb{R}^n$. Set again $n = 1$ just for simplicity (everything we write in this comment would be true anyway). Then, as differential operators, $U_0; \dots; U_j$ only act on the variable z , while W_0 only acts on the variable t , i.e. we have the correspondence

$$U_j(z; t) \partial_{z_j} \quad \text{for } j = 0; \dots; d \quad \text{and} \quad W_0(t) \partial_t = W_0(t) \partial_t$$

One has for all $1 \leq j \leq d$

$$[V_0; V_j] = [U_0; U_j] + [W_0; U_j] = [U_0; U_j] + W_0(t) (\partial_t U_j) \partial_{z_j}$$

If we calculate the second term on the RHS of the above along a solution z_t of the ODE, we obtain

$$W_0(z_t) (\partial_t U_j(z; t)) \partial_{z_j} = \partial_t (U_j(z; t)) \partial_{z_j}$$

This suggests that we may evaluate the vector fields along the solution of the ODE and then think of them as functions of z and time t , rather than as functions of z and t , i.e. $\mathbb{R}_z^n \times \mathbb{R} \ni (z; t) \mapsto U_j(z; t) \in \mathbb{R}^n; j = 0; \dots; d$. If we do so, then Hypothesis [H.1] can be equivalently rephrased as follows: the Lie algebra

$$\left[\sum_{k=1}^d \text{span} L_k^U(z; t) \right]$$

is nilpotently generated and spans \mathbb{R}_z^n for every $z \in \mathbb{R}^n$ and along any solutions z_t of the ODE (6.2).²²

- As is well known, Hypothesis [H.2] is implied by a Lyapunov-type condition; namely, if there exists some non-negative function $V \in C^2(\mathbb{R}^n)$ with compact level sets and such that

$$L_t V(z) \leq -C_1 + C_2 V(z); \quad \text{for every } z \in \mathbb{R}^n; t \geq 0; \quad (6.10)$$

then [H.2] is satisfied. Here L_t is the operator

$$L_t V(z) = U_0(z; t) \cdot \nabla V(z) + \sum_{i=1}^d U_i(z; t) \cdot \nabla (U_i(z; t) \cdot \nabla V(z));$$

where $\nabla = (\partial_{z_1}; \dots; \partial_{z_n})$.

²² Given an initial datum, the solution of the ODE is unique. When we say that this should hold along any solutions, we mean along all the solutions that one can obtain by starting from different initial data.

- The Obtuse angle condition does not imply tightness; in Example 6.9 we show that [H.2] does not imply [H.3] and [H.3] does not imply [H.2]. Viceversa, we also show that [H.3] does not imply the existence of a Lyapunov function, condition (6.10). □

Note 6.3. As already pointed out, if $X_t = (Z_t; \tau_t)$, where $Z_t; \tau_t$ are given by a representation of the form “ODE + SDE” (6.1)–(6.2), then X_t solves an SDE of the form (1.1), with $V_0 = (U_0; W_0); V_1 = (U_1; 0); \dots; V_d = (U_d; 0)$. Hence $V_0^{(?) } = (0; \dots; 0; W_0)$ (see definition (1.9)). We note in passing that in this case one has

$$Z_t := e^{tV_0^{(?)}} X_t = e^{tV_0^{(?)}} (Z_t; \tau_t) = e^{tV_0^{(?)}} (Z_t; e^{tV_0^{(?)}} \tau_0) = (Z_t; \tau_0):$$

(This is not of much use at the moment, but it will help at the beginning of Section 7 to make a link between the setting of this section and the setting of the next). Therefore, while X_t belongs to the hyperplane $H_t := \{x \in \mathbb{R}^{n+1} : x = (z; \tau); z \in \mathbb{R}^n\}$ for each $t \geq 0$, Z_t remains, for every $t \geq 0$, on the same hyperplane, namely the hyperplane $H_0 := \{x \in \mathbb{R}^{n+1} : x = (z; 0); z \in \mathbb{R}^n\}$ (which is precisely the manifold $S_{x_0} = S_{(z_0; 0)}$, see second bullet point in Note 6.2) for every $t \geq 0$. □

Lemma 6.4. Let Hypothesis 6.1 hold. Then the SDE (6.8) admits a unique invariant measure, μ . Moreover,

$$(\mathbb{Q}_t g)(z) = \int_{\mathbb{R}^n} g(z) \mu(dz); \text{ for every } z \in \mathbb{R}^n \text{ and every } g \in C_b(\mathbb{R}^n):$$

Proof of Lemma 6.4. This is completely standard and we omit it. See for example [17]. We just point out that the existence of the invariant measure comes from assumption [H.2] and the uniqueness is a consequence of Hypothesis [H.3] and Proposition 5.10. □

Theorem 6.5. Let Hypothesis 6.1 hold. In particular, let $x_0 = (z_0)$ be a stationary point for the ODE (6.2) and μ be the invariant measure of the process (6.8). Then, for every $s \geq 0$,

$$\lim_{t \rightarrow \infty} (\mathbb{Q}_{s;t}^0 g)(z) = \int_{\mathbb{R}^n} g(z) \mu(dz); \text{ for every } z \in \mathbb{R}^n \text{ and every } g \in C_b(\mathbb{R}^n):$$

The proof of this theorem can be found after the statement of Theorem 6.6. Theorem 6.5 describes the asymptotic behaviour of the process Z_t . However, in this paper we are interested in the process X_t . The long-time behaviour of the process X_t is described by Theorem 6.6 below, which is just a straightforward consequence of Theorem 6.5. In order to state Theorem 6.6, we clarify the following: while Z_t is a process in \mathbb{R}^n with invariant measure(s) $\mu = (\mu; 0)$ supported on \mathbb{R}^n , X_t is a process in \mathbb{R}^{n+1} ; so, strictly speaking, any invariant measure of X_t is a probability measure on \mathbb{R}^{n+1} . However such a measure is supported on the n -dimensional hyperplane

$$H := \{x \in \mathbb{R}^{n+1} : x = (z; \tau); z \in \mathbb{R}^n\}$$

and it is just a trivial extension of the measure μ . That is, let $\nu = (\mu; 0)$ be the measure on \mathbb{R}^{n+1} such that

$$(\nu(A)) = (\mu(A \cap H)) \text{ for every Borel set } A \subset \mathbb{R}^{n+1}: \tag{6.11}$$

In particular, $\nu(A) = (\mu(A))$ if $A \subset H$ and $\nu(A) = 0$ if $A \cap H = \emptyset$. Let $I_0(\cdot) = \int_0^\cdot f_0 \, d\mathbb{R} : \tau^0 \rightarrow \tau$ as $t \rightarrow \infty$. Let also $I_0 = I_0(\cdot) := \{x_0 \in \mathbb{R}^{n+1} : x_0 = (z_0; 0); z_0 \in \mathbb{R}^n\}$.

Theorem 6.6. Consider the process $X_t = (Z_t; \tau_t) \in \mathbb{R}^{n+1}$ satisfying a representation the form of "SDE + ODE" (6.1)–(6.2) with initial condition (6.3) and associated semigroup P_t , defined in (6.5). Let Hypothesis 6.1 hold. In particular, according to Hypothesis 6.1 [H.4], let μ be a (any) stationary point of the ODE (6.2) and ν be the invariant measure of the corresponding process (6.8); let also $\mu_0 = (\mu; \nu_0)$ be the measure on \mathbb{R}^{n+1} defined in (6.11) and supported on the hyperplane H . Then, for every $x \in \mathbb{R}^{n+1}$, we have

$$\lim_{t \rightarrow \infty} (P_t f)(x) = \int_{\mathbb{R}^{n+1}} f(u) \, d\mu(u) = \int_H f(u) \, d\nu(u);$$

for every $f \in C_b(\mathbb{R}^{n+1})$. The limit in the above does not hold if $x \notin H$; that is, μ_0 is the whole basin of attraction of the measure $\mu_0 = (\mu; \nu_0)$.

We now introduce some definitions that will be needed for the proof of Theorem 6.5.

A family $\{g_t\}_{t \geq 0}$ of probability measures on \mathbb{R}^n is said to be an evolution system of measures for the two-parameter semigroup $\{Q_{s,t}\}_{0 \leq s \leq t}$ if for all $0 \leq s \leq t$ and $g \in C_b(\mathbb{R}^n)$ we have

$$\int_{\mathbb{R}^n} Q_{s,t} g(z) \, g_0(z) \, dz = \int_{\mathbb{R}^n} g(z) \, g_t(z) \, dz; \quad (6.12)$$

Let $Q_{s,t}^*$ denote the adjoint of $Q_{s,t}$ over the space $C_b(\mathbb{R}^n)$, that is

$$(Q_{s,t}^* g)(A) := \int_{\mathbb{R}^n} Q_{s,t} 1_A(z) \, g(z) \, dz; \quad \text{for any Borel measurable } A \subset \mathbb{R}^n;$$

Then we can write (6.12) as

$$Q_{s,t}^* g_0 = g_t; \quad \text{for all } 0 \leq s \leq t;$$

Further background on evolution system of measures can be found in [32, 16].

Proof of Theorem 6.5. The proof is in three steps.

Step 1: We first construct a tight evolution system of measures, $\{g_t\}_{t \geq 0}$, for the semigroup $\{Q_{s,t}\}_{0 \leq s \leq t}$. To this end, take any point $z_0 \in \mathbb{R}^n$, define $\mu_0 = \delta_{z_0}$ and then let $g_t := Q_{0,t} \mu_0$. Now $\{g_t\}_{t \geq 0}$ is an evolution system of measures; indeed,

$$Q_{s,t} g_s = Q_{s,t} Q_{0,s} \mu_0 = Q_{0,t} \mu_0 = g_t;$$

(A more general construction of the evolution system is given in [32, Section 5]). To see that $\{g_t\}_{t \geq 0}$ is tight, $\epsilon > 0$; by Hypothesis [H.2] we may take a compact set $K \subset \mathbb{R}^n$ such that $Q_t^{0,z_0}(K^c) < \epsilon$. By definition of g_t we then have

$$g_t(K^c) = (Q_{0,t} \mu_0)(K^c) = Q_{0,t} 1_{K^c}(z_0) < \epsilon;$$

Step 2: $Q_{s,t} g(z) - g_t(z)$ converges to zero as t tends to ∞ for all $s \geq 0; z \in \mathbb{R}^n; g \in C_b(\mathbb{R}^n)$. We defer the proof of this fact to Lemma B.1. Since $\{g_t\}_{t \geq 0}$ is tight, by Prokhorov's theorem there exists a diverging sequence $\{t_k\}$ such that g_{t_k} converges weakly to some probability measure μ_0 , as t_k tends to ∞ .

Step 3: Show that $\mu_0 = \nu$. We defer the proof of this equality to Lemma B.4. If $\mu_0 = \nu$, then g_t converges weakly to ν and the claim of the theorem follows; indeed,

$$|Q_{s,t} g(z) - g_t(z)| \leq |Q_{s,t} g(z) - g_t(z)| + |g_t(z) - \nu(z)|;$$

The first term converges to zero by Step 2 and the second term vanishes in the limit since g_t converges weakly to ν as $t \rightarrow \infty$. □

Note 6.7. The statements and proofs of Lemma B.1 and Lemma B.4 are the core of the proof of Theorem 6.5. The arguments used in the proofs of such lemmata are analogous in structure to those presented in [1, Section 6]. The main differences arise when dealing with the regularity of the semigroup, as [1] assumes uniform ellipticity. Lemma B.2 (needed to prove Lemma B.4) is the main place where we take care of the relaxed regularity assumptions. □

Let μ_t^x denote the measure defined by

$$\mu_t^x(A) = P_t 1_A(x); \quad \text{for all Borel sets } A \subset \mathbb{R}^{n+1} :$$

Proposition 6.8. If $\lambda < 1$ then the family of measures $\int \mu_t^{(z; \cdot)} g_t$ is not tight for any $z \in \mathbb{R}^n$ (hence, by Prokhorov's theorem, there is no probability measure μ such that $P_t f(z; \cdot) \rightarrow \mu(f)$, for all $f \in C_b(\mathbb{R}^{n+1})$).

Proof of Proposition 6.8. Fix $z \in \mathbb{R}^n$ and let $x = (z; \cdot) \in \mathbb{R}^{n+1}$. Assume by contradiction that $\int \mu_t^x g_t$ is tight. Then, for any $\epsilon > 0$ there exists a compact set $K \subset \mathbb{R}^{n+1}$ such that $\mu_t^x(K) > 1 - \epsilon$ for all $t \geq 0$. Since K is compact we may take R sufficiently large such that $K \subset \mathbb{R}^n \times [R; R]$; then one has

$$P_x \int_{t \geq R} \mu_t^x(K) > 1 - \epsilon; \quad \text{for all } t \geq 0: \tag{6.13}$$

However $\lambda < 1$ so we may take t sufficiently large that $\int_{t \geq R} > R$. This contradicts (6.13), hence μ_t^x is not tight. □

Example 6.9 (UFG-Grušin Plane). We give here a simple example of a process that satisfies the Obtuse Angle Condition but is not tight. Let $d = 1, N = 2$ and

$$V_0 = k \partial_x; \quad V_1 = \partial_z; \quad k \in \mathbb{R} :$$

This corresponds to the SDE

$$\begin{aligned} dz_t &= k dt \\ dx_t &= \sqrt{2} z_t dB_t; \end{aligned}$$

where $\int B_t g_t$ is a one-dimensional Brownian motion. Because $[V_1; V_0] = kV_1$, we have

$$([V_1; V_0]f)(V_1 f) = k(V_1 f)^2$$

therefore the Obtuse Angle Condition, (3.4) is satisfied if and only if $k > 0$ (it is also shown in [14, Example 4.4] that, if $k > 0$, then $V_1(P_t f)(\cdot)$ decays exponentially fast with rate $2k$). On the other hand, if $k < 0$ the process is not tight. Indeed, Hypothesis 6.1 ([H.2]) is satisfied if and only if $k < 0$, as we come to show. To this end, we first solve the SDE, and find

$$\begin{aligned} z_t &= e^{kt} \\ x_t &= x_0 + \sqrt{2} \int_0^t e^{ks} dB_s; \end{aligned}$$

As a consequence of Proposition 6.8, the whole process $(Z_t; x_t)$ is not tight if $k > 0$. However in this case also the process Z_t , seen as a non-autonomous one dimensional SDE, is not tight when $k > 0$. Indeed suppose by contradiction that ([H.2]) holds; then for any $\epsilon > 0$ there exists $R > 0$ such that

$$Q_{0;t} 1_{[R; R]}(z) > 1 - \epsilon; \quad \text{for all } t \geq 0: \tag{6.14}$$

However if $Z_0 = z$ then Z_t is normally distributed with mean z and variance $2(e^{2kt} - 1)/k$, so we may write

$$Z_t = z + \frac{\sqrt{2(e^{2kt} - 1)}}{k} \sigma \quad (6.15)$$

where σ is a one-dimensional standard normal random variable. Then we have

$$Q_{0,t} 1_{[R,R]}(z) = E 1_{[R,R]}(Z_t^{0;z}) = P \left(z + \frac{\sqrt{2(e^{2kt} - 1)}}{k} \sigma \in [R,R] \right); \quad (6.16)$$

and the right hand side of (6.16) converges to zero as t tends to ∞ which contradicts (6.14). Note that if $k = 0$ then $Z_t = \sqrt{2t} B_t$ which is not tight by a similar argument. However if $k < 0$ then the process Z_t is tight. Indeed, assume that $k = -\lambda < 0$; to see that $\{Q_t^{0;z}\}_{t \geq 0}$ is a tight family of measures, it is sufficient to apply a Lyapunov criterion and show that the function $L_t(z) = z^2$ satisfies $\sup_t Q_{0,t} L_t'(y) < 1$ (when $k = -\lambda < 0$). To prove the latter fact, observe that if $(Z_s; 0) = (z; 0)$ then, by (6.15), we get

$$Q_{s,t} L_t'(z) = E[Z_t^2 | Z_s = z] = z^2 + \frac{2e^{-2\lambda s}}{\lambda} (1 - e^{-2\lambda(t-s)}) \left(z^2 + \frac{e^{-2\lambda s}}{\lambda} \right);$$

If $k < 0$ we see that $X_t = (Z_t; t)$ converges in law. □

Example 6.10. We conclude this section with an example which satisfies all the points in Hypothesis 6.1 in a non-trivial way, in the sense that it exhibits many invariant measures. Take $k > 1$ and consider the following SDE

$$\begin{aligned} dX_t &= \sin(X_t) dt \\ dZ_t &= kZ_t dt + \sqrt{2} dB_t; \end{aligned}$$

In this case

$$V_0 = \sin(X) @ kZ @; \quad V_1 = @; \quad U_0 = kZ @; \quad U_1 = @;$$

Then we have

$$[V_1; V_0] = [@; \sin(X) @ kZ @] = k @ + \sin(X) @ = k + \frac{\sin(X)}{k} V_1;$$

Note that the function $h(X) = \sin(X)$ is bounded and smooth, when extended to the origin with the value $h(0) = 1$, so the UFG condition is satisfied at level $m = 1$. Moreover,

$$([V_1; V_0]f)(V_1 f) = \left(k + \frac{\sin(X)}{k}\right) jV_1 f j^2 - (k - 1) jV_1 f j^2$$

and hence (3.4) is satisfied. To apply the results of Section 6 we must show that Hypothesis 6.1 holds. Note that the vector field V_1 is non-zero except when $X = 0$ therefore Hypothesis 6.1 [H.1] is satisfied everywhere that $X \neq 0$. To show that Hypothesis 6.1 [H.2] holds we consider a function $L_t \in C^2(\mathbb{R})$ such that $L_t'(z) = |z|$ for $|z| > 1$. Then, for $|z| > 1$, one has

$$L_t'(z) = kz^0(z) + \frac{2}{t} z^0(z) = kz \text{sign}(z) = k'(z);$$

Therefore L_t' is a Lyapunov function so by Note 6.2 we have that the measures $\{Q_t^{0;z}\}_{t \geq 0}$ are tight for any $z \in \mathbb{R}$ and Hypothesis 6.1 [H.2] is satisfied. We also have that L_t' converges for any $z \in \mathbb{R}$ and the limit L' is given by

$$L' = \begin{cases} \frac{8}{2n} & \text{for } z \in ((2n-1); (2n+1)) \text{ for some } n \in \mathbb{Z} \setminus \{0\} \\ \frac{2}{2n+1} & \text{for } z = (2n+1) \text{ for some } n \in \mathbb{Z} \\ 0 & \text{for } z \in (-\infty; -1) \cup (1; \infty); \end{cases}$$

Hence for $\mathbb{Q}(\cdot; \cdot)$ we may apply Theorem 6.6 to obtain that $X_t = (Z_t; t)$ converges in law to a random variable which whose law corresponds to the unique invariant measure defined on the line $\mathbb{R} \setminus \{0\}$. Moreover, for $t = n$ for some $n \in \mathbb{Z}^+$ we see that $Z_t = Z_n$ and Z_t satisfies the Ornstein-Uhlenbeck SDE

$$dZ_t = -kZ_t dt + \sqrt{2k} dB_t$$

In particular, in this case Z_t has a unique invariant measure and this is given by a Gaussian measure with mean 0 and variance $2/k$. Therefore for any $n \in \mathbb{Z}^+$ and $t \in ((2n-1)/k; (2n+1)/k)$ we have that X_t converges in law to $(\sqrt{2/k} Z_n; 2n)$, where Z_n is a one-dimensional standard normal random variable. □

7 Long-time behaviour of UFG diffusions: general case

In the previous section we investigated the case in which the representation of the form “ODE + SDE” is global. In this section we study the general UFG-case, in which such a representation is, in general, only local. That is, we finally address the full problem of analysing the asymptotic behaviour of (1.1), assuming that the vector fields $V_0; \dots; V_d$ satisfy the UFG condition (see Definition 3.1). This case is substantially richer than the one considered in Section 6; however the fact that, locally, we can always represent the SDE (1.1) as a system of the form “ODE + SDE”, still means that we should be able to identify a suitable ODE which drives the dynamics. We will demonstrate that this is indeed the case and that such an ODE is the integral curve of the vector field $V_0^{(?)}$; that is, the curve

$$x_t = e^{tV_0^{(?)}} x_0; \tag{7.1}$$

where x_0 is the initial datum of the SDE (1.1), i.e. $X_0 = x_0$. This should not be a surprise in view of Proposition 5.3. Nevertheless, to understand why this is the case, it is useful to build an analogy with the setting of the previous section: if the SDE is of the form (6.1)–(6.2), then $V_0^{(?)} = (0; \dots; 0; W_0)$. Therefore in the simplified setting (6.1)–(6.2), the ODE (7.1) substantially reduces to (6.2). The previous sentence is correct for less than observing that (7.1) is an N -dimensional ODE, while (6.2) is a one-dimensional curve. We keep using the notation x_t for both curves only to emphasize the analogy; however, while the one-dimensional autonomous nature of the ODE (6.2) implies that its solution has a limit, the zoology of possible behaviours for the curve (7.1) is much more varied. In this paper we only analyse the case in which the curve (7.1) converges to a limit and in future work we will treat more general cases. ²⁴ However, roughly speaking, in Theorem 7.12, we prove that a necessary condition for the SDE (1.1) to have an invariant measure is that the ODE (7.1) should admit one as well (notice that if the curve (7.1) converges to a limit point x , then it admits the Dirac measure δ_x as invariant measure).

As anticipated in the introduction, the above discussion motivates introducing the process

$$Z_t := e^{tV_0^{(?)}} (X_t^{(x_0)}); \tag{7.2}$$

Clearly $Z_0 = x_0$, so Z_t and $X_t^{(x_0)}$ start from the same point. This process is time-inhomogeneous (as we show at the beginning of Section 7.2) and it will have a central role in what follows, hence further comments on the definition (7.2) are in order:

²³ Since Z_t satisfies a non-autonomous Ornstein-Uhlenbeck equation one can also study its asymptotic behaviour more directly, see e.g. [22].

²⁴ For example the curve (7.1) could be periodic, ergodic or chaotic (this is a non-exhaustive list of possibilities).

- To continue drawing the useful parallel with Section 6, notice that this process plays in this context an analogous role to the one that Z_t (solution of (6.1)) has in Section 6, see Note 6.3.
- Let us recall that if $X_0 \in S_{x_0}$ then $X_t \in \overline{S_{x_0}}$ for every $t \geq 0$ (see Proposition 4.7); more precisely, for every $t \geq 0$ X_t belongs to the integral submanifold $\overline{S_{e^{tV_0^{(?)}}(x_0)}}$ almost surely (see Proposition 5.3). We will make assumptions to guarantee that X_t hits neither the boundary of $S_{e^{tV_0^{(?)}}(x_0)}$ nor the boundary of S_{x_0} in finite time (see Hypothesis 7.5 [A.4], Lemma 4.15 and Note 7.6 for more precise comments on this). Therefore Z_t lives on the manifold S_{x_0} , for every $t \geq 0$. So, in the end, while X_t takes values in $\overline{S_{x_0}}$, Z_t takes values in $S_{x_0} \subset \overline{S_{x_0}}$. One can informally think of Z_t as being a “projection” of X_t on the submanifold $S_{x_0} \subset \overline{S_{x_0}}$, see again Note 6.3.
- Finally, on a small technical point, as we have already observed in Note 4.14, $V_0^{(?)}$ may not be uniformly Lipschitz. However, to avoid problems of well posedness and uniqueness, throughout this section we assume that $V_0^{(?)}$ is indeed Lipschitz.

We will show that the time-inhomogeneous process $fZ_t g_{t \geq 0}$ can be studied by means of slight modifications of the approach used in Section 6 to study the process (6.1). Therefore the strategy (and one of the main novelties) of this section is to use the auxiliary time-inhomogeneous process $fZ_t g_{t \geq 0}$ in order to make deductions on the behaviour of the time-homogeneous process X_t . We carry out this programme in Section 7.2 below. Before moving on, we give a simple example which demonstrates that $Z_t \in S_{x_0}$ for every $t \geq 0$ and, in Section 7.1, we gather further preliminary results on the process Z_t .

Example 7.1 (Random Circles continued). Consider again Example 4.11, in the case in which the initial datum is $(x_0; y_0) = (1; 0)$. Using (5.1) and (5.2)–(5.3), we have

$$\begin{aligned} Z_t := e^{tV_0^{(?)}}(X_t; Y_t) &= \begin{pmatrix} X_t \cos(t) - Y_t \sin(t) \\ X_t \sin(t) + Y_t \cos(t) \end{pmatrix} \\ &= \begin{pmatrix} e^{\frac{p}{2} B_t} \cos(t) \cos(t) + e^{\frac{p}{2} B_t} \sin(t) \sin(t) \\ e^{\frac{p}{2} B_t} \cos(t) \sin(t) + e^{\frac{p}{2} B_t} \sin(t) \cos(t) \end{pmatrix} \\ &= e^{\frac{p}{2} B_t} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{aligned}$$

In particular, Z_t takes values in the positive half-line, which is precisely $S_{(1;0)} = S_{(x_0; y_0)}$. □

Note 7.2. In [7] the author considers the curve $e^{tV_0}(X_t^{(x)})$ while here we consider the curve $e^{tV_0^{(?)}}(X_t^{(x)})$. This is due to two things: i) in [7] the author was not concerned with the dynamics of the process but was instead interested in studying the density of the SDE; ii) the work [7] assumes that the HC is satisfied, so that $\hat{\Lambda}_0$ has constant rank (constant and equal to N) and the decomposition of state space into integral manifolds of $\hat{\Lambda}_0$ becomes in that case trivial – there is only one submanifold, which is the whole \mathbb{R}^N .

When studying the density of the law of the process there is no advantage in using $V_0^{(?)}$ over V_0 – indeed even though V_0 is smooth, $V_0^{(?)}$ need not be continuous, see Note 4.14. On the other hand, if one is interested in the dynamics of the process (and the HC is not enforced), then $V_0^{(?)}$ becomes an indicator of whether the path has left the integral manifold of $\hat{\Lambda}_0$ from which it started or not. In other words for us it is important to know the behaviour of the manifolds $S_{e^{tV_0^{(?)}}(x)}$ as t varies. In particular, we wish to distinguish between the case when $S_{e^{tV_0^{(?)}}(x)}$ is constant in t which corresponds to the situation in which $V_0^{(?)} = 0$ and when $S_{e^{tV_0^{(?)}}(x)}$ varies in time, see Section 5 and Lemma 4.16.

7.1 The auxiliary process Z_t and its associated two-parameter semigroup

By differentiating (7.2) we see that Z_t satisfies the following SDE

$$\begin{aligned} dZ_t &= V_0^{(?)}(e^{tV_0^{(?)}}(X_t^{(x)}))dt + J_x e^{tV_0^{(?)}}(X_t^{(x)})V_0(X_t^{(x)})dt \\ &\quad + \frac{p}{2} \sum_{i=1}^d J_x e^{tV_0^{(?)}}(X_t^{(x)})V_i(X_t^{(x)}) dB_t^i \\ &= V_0^{(?)}(Z_t)dt + \text{Ad}_{tV_0^{(?)}} V_0(Z_t)dt + \frac{p}{2} \sum_{i=1}^d \text{Ad}_{tV_0^{(?)}} V_i(Z_t) dB_t^i; \end{aligned}$$

where, as customary, we have set $(\text{Ad}_{tV} Y)(x) := (J_x e^{tV})(e^{tV}(x)) Y(e^{tV}(x))$, for any two smooth vector fields V and Y . By using (1.9), the elementary property $\text{Ad}_{tV} V = V$ and introducing the notation

$$\begin{aligned} V_{0;t} &:= \text{Ad}_{tV_0^{(?)}} V_0^{(\hat{)}} \\ V_{j;t} &:= \text{Ad}_{tV_0^{(?)}} V_j; \quad j = 1, \dots, d; \end{aligned} \tag{7.3}$$

we conclude that Z_t satisfies the following SDE with time-dependent coefficients:

$$\begin{aligned} dZ_t &= \text{Ad}_{tV_0^{(?)}} V_0^{(\hat{)}}(Z_t)dt + \frac{p}{2} \sum_{i=1}^d \text{Ad}_{tV_0^{(?)}} V_i(Z_t) dB_t^i \\ &= V_{0;t}(Z_t)dt + \frac{p}{2} \sum_{i=1}^d V_{i;t}(Z_t) dB_t^i; \end{aligned} \tag{7.4}$$

As usual, we denote by $fP_t g_{t_0}$ the one parameter semigroup associated with $fX_t g_{t_0}$; the two-parameter semigroup associated with $fZ_t g_{t_0}$ is instead given by

$$Q_{s;t} f(z) = E[f(Z_t) | Z_s = z]; \quad z \in S_{x_0}; s \leq t; f \in C_b(\mathbb{R}^N);$$

The semigroups $fQ_{s;t} g_{0 \leq s \leq t}$ and $fP_t g_{t_0}$ are related as follows:

$$\begin{aligned} P_t f(x) &= Q_{s;s+t}(f e^{(s+t)V_0^{(?)}})(e^{sV_0^{(?)}}(x)); \quad x \in S_{e^{sV_0^{(?)}}(x_0)}; s \in \mathbb{R}; t \geq 0; f \in C_b(\mathbb{R}^N); \\ Q_{s;t} g(z) &= P_{t-s}(g e^{tV_0^{(?)}})(e^{sV_0^{(?)}}(z)); \quad z \in S_{x_0}; 0 \leq s \leq t; g \in C_b(\bar{S}_{x_0}); \end{aligned} \tag{7.5}$$

We stress that $fQ_{s;t} g_{0 \leq s \leq t}$ is defined on S_{x_0} (as per Hypothesis 7.5 below). In (7.5) we consider functions g which are continuous up to and including the boundary of S_{x_0} for purely technical reasons (see proof of Proposition 7.4).

In Proposition 7.4 we make some clarifications on the smoothing properties of the semigroup $fQ_{s;t} g_{0 \leq s \leq t}$. To state such a lemma, we need to properly formulate some preliminary facts. Consider the following ‘‘hierarchy’’ of operators:

$$\begin{aligned} V_{[i];t} &:= V_{i;t} \quad i = 0; 1; \dots; d \text{ (defined as in (7.3))} \\ V_{[0];t} &:= [V_{[1];t}; V_{[0];t} + \text{Ad}_{tV_0^{(?)}} V_0]; \quad \text{Ad}_{tV_0^{(?)}} V_0 \in \mathcal{A}; \\ V_{[i-1];t} &:= [V_{[i];t}; V_{[i-1];t}]; \quad \text{Ad}_{tV_0^{(?)}} V_i \in \mathcal{A}; \quad i = 1; \dots; d; \end{aligned}$$

For each $\mathcal{A} \in \mathcal{A}$ we can view the vector field $(z; t) \mapsto V_{[i];t}(z)$ as a vector field on \mathbb{R}^N , the coefficients of which depend on time or as a vector field on $\mathbb{R}^N \times \mathbb{R}$. We can define the UFG condition for vector fields in $\mathbb{R}^N \times \mathbb{R}$ in an analogous way to Definition 3.1. In Proposition 7.3 we prove that the set of vector fields $\{V_{[0];t} + \text{Ad}_{tV_0^{(?)}} V_0; V_{[1];t}; \dots; V_{[d];t}\} g$ satisfy the UFG condition on $\mathbb{R}^N \times \mathbb{R}$ provided the vector fields $\{V_0; V_1; \dots; V_d\} g$ satisfy the UFG condition on \mathbb{R}^N .

Proposition 7.3. Assume that the vector fields f, V_0, V_1, \dots, V_d on \mathbb{R}^N satisfy the UFG condition at level m ; then the vector fields $f \otimes + V_{[0];t}, V_{[1];t}, \dots, V_{[d];t}$ satisfy the UFG condition at level m when viewed as vector fields on $\mathbb{R}^N \times \mathbb{R}$. Moreover, for any $z \in A_m$,

$$V_{[j];t} = \text{Ad}_{tV_0^{(?)}} V_{[j]} \quad (7.6)$$

Proof of Proposition 7.3. The proof is deferred to Appendix B.3. □

Recall from Section 3 that the map $z \in S_{x_0} \rightarrow P_t f(z)$ is smooth (along the directions $V_{[j]}, z \in A_m$) for any $f \in C_b(\mathbb{R}^N)$. In Proposition 7.4 we show that for each fixed $s < t$ the map $z \in S_{x_0} \rightarrow Q_{s,t} g(z)$ is also smooth in the directions $V_{[j];s}$ for any $g \in C_b(\overline{S_{x_0}})$ and $z \in A_m$. A key observation to understand the statement of Proposition 7.4 is the following one:

$$V \in \mathbb{R}^N \text{ and } \mathbb{R}^N \text{ is invariant under the vector field } W \Rightarrow \text{Ad}_{tW} V \in \mathbb{R}^N \quad (7.7)$$

In particular, $V_{j;t} \in \mathbb{R}^N$ for every $j \in \{0, \dots, d\}$.

Proposition 7.4. Assume the vector fields f, V_0, \dots, V_d satisfy the UFG condition and that the vector $V_0^{(?)}$ is uniformly Lipschitz. Then, for any $g \in C_b(\overline{S_{x_0}})$, the map $(z; s) \rightarrow Q_{s,t} g(z)$ is differentiable in the time variable s and in the spatial directions $V_{[j];s}$ for any $z \in \overline{S_{x_0}}; t > s; z \in A_m$. Moreover $Q_{s,t} g(z)$ satisfies the equation

$$\partial_s Q_{s,t} g(z) = L_s Q_{s,t} g(z); \quad \text{for any } z \in S_{x_0}; s < t \quad (7.8)$$

Here L_s is the differential operator defined as

$$L_s(z) = V_{0;s}(z) + \sum_{i=1}^d V_{i;s}^2(z);$$

for $z \in \overline{S_{x_0}} \rightarrow \mathbb{R}$ sufficiently smooth.

Proof of Proposition 7.4. The proof is deferred to Appendix B.3. □

7.2 Convergence to equilibria

We now turn to the asymptotic behaviour of the process $\{Z_t\}_{t \geq 0}$. As we have already stated, we will concentrate on the case in which the solution of the ODE (7.1) converges. Let us define the map

$$W^1 : \text{Dom}(W^1) \subset \mathbb{R}^N \rightarrow \mathbb{R}^N$$

$$x \mapsto \lim_{t \rightarrow \infty} e^{tV_0^{(?)}}(x):$$

Here $\text{Dom}(W^1)$ is the set of all points $x \in \mathbb{R}^N$ such that the integral curve $e^{tV_0^{(?)}}(x)$ converges to a finite limit as t tends to ∞ .

Hypothesis 7.5. Assume the following:

[A.1] The vector fields f, V_0, V_1, \dots, V_d satisfy the UFG condition.

[A.2] The vector field $V_0^{(?)}$ is uniformly Lipschitz.

[A.3] Define the measures μ_t^x by $\mu_t^x(A) = P_t 1_A(x)$ for any Borel measurable $A \subset \mathbb{R}^N$. The family $\{\mu_t^x : t \geq 0\}$ is tight for all $x \in \mathbb{R}^N$.

²⁵ Indeed, by the definition of invariance (see Definition 3.7), we have that $\int_x e^{tW}(x)$ maps $\hat{\mu}(x)$ to $\hat{\mu}(e^{tW}(x))$. Therefore $\int_x e^{tW}(e^{-tW}(x))$ maps $\hat{\mu}(e^{-tW}(x))$ to $\hat{\mu}(x)$. Now $V \in \mathbb{R}^N$, so $V(e^{-tW}(x)) \in \mathbb{R}^N$ and we have that $\text{Ad}_{tW} V(x) = \int_x e^{tW}(x) V(e^{-tW}(x)) \in \mathbb{R}^N$. That is, $\text{Ad}_{tW} V \in \mathbb{R}^N$.

- [A.4] Define the measures $q_t^{s;z}$ by $q_t^{s;z}(A) = Q_{s;t} 1_A(z)$ for any Borel measurable $A \subset S_{x_0}$. Then we require that for each $z \in S_{x_0}$ the measures $\{q_t^{0;z} : t \geq 0\}$ are tight on S_{x_0} ; that is, for all $\epsilon > 0$ there exists a compact set $K_\epsilon \subset S_{x_0}$ such that $q_t^{0;z}(K_\epsilon) \geq 1 - \epsilon$ for all $t \geq 0$.
- [A.5] The Obtuse Angle Conditions (3.4) and (3.6) are satisfied (with the understanding of Hypothesis 3.16 [SA.3]).
- [A.6] The initial datum x_0 of the SDE (1.1) is such that the curve (7.1) started at x_0 , admits a limit, i.e. there exists $x \in \mathbb{R}^N$ such that $V_0^{(?)}(x) = 0$ and $e^{tV_0^{(?)}}(x_0) \rightarrow x$ as $t \rightarrow \infty$.
- [A.7] Assumptions on the map W^1 : the domain of W^1 contains the whole manifold \bar{S}_{x_0} and the image of S_{x_0} through W^1 is all contained in a submanifold $\hat{\Lambda}$. More explicitly, there exists an integral submanifold $\hat{\Lambda}, S_{\bar{x}}$, such that $V_0^{(?)} = 0$ on $S_{\bar{x}}$ and the image of S_{x_0} through W^1 is all contained in $S_{\bar{x}}, W^1(S_{x_0}) \subset S_{\bar{x}}$. Furthermore we assume that W^1 is a continuous map from $\bar{S}_{x_0} \setminus S_{x_0} \subset S_{\bar{x}}$ into \mathbb{R}^N .

Note 7.6. Some comments on the above assumptions, in the order in which they are stated.

- As a general premise, observe that, for every fixed $t \geq 0, X_t \in S_{x_0}$ if and only if $Z_t \in S_{x_0}$. Indeed, $X_t = e^{tV_0^{(?)}} Z_t$ so if Z_t is in S_{x_0} then in particular it is in S_{x_0} and X_t is just obtained by moving along an integral curve of $V_0^{(?)}$; hence, by construction of the manifold S_{x_0} , X_t is still in S_{x_0} . The validity of the reverse implication can be argued similarly (using Lemma 4.15 and Proposition 5.3 as well). As a consequence, if Z_t does not hit the boundary of S_{x_0} in finite time then X_t does not hit the boundary of S_{x_0} in finite time.
- Hypothesis 7.5 [A.4] implies that $Z_t \in S_{x_0}$ almost surely, for every $t \geq 0$, i.e. it implies that Z_t does not hit the boundary of S_{x_0} in finite time. Indeed, assume by contradiction that there exists $t_0 > 0$ such that $P(Z_{t_0} \in @S_0) =: \epsilon > 0$. Recall $@S_0 := \bar{S}_{x_0} \cap S_{x_0}$. By the previous bullet point if Z_{t_0} belongs to $@S_0$ then $X_{t_0} \in @S_{x_0}$. By Proposition 5.1 we then have that X_t is in the boundary of S_{x_0} for any $t > t_0$. That is, for any $t > t_0$

$$P(X_t \in @S_{x_0}) = P(X_{t_0} \in @S_{x_0}) = P(Z_{t_0} \in @S_0) = \epsilon > 0^{26}$$

We know from [A.3] that, given ϵ as in the above, there exists a compact set $K_{\epsilon=2} \subset S_{x_0}$ such that $P(Z_t \in K_{\epsilon=2}) = q_t(K_{\epsilon=2}) \geq 1 - \epsilon=2$ for every $t \geq 0$. Now using that S_{x_0} and $@S_{x_0} = \bar{S}_{x_0} \cap S_{x_0}$ are disjoint, for every $t > t_0$ we have

$$\begin{aligned} 1 &= P(X_t \in @S_{x_0}) + P(X_t \in S_{x_0}) \\ &= P(Z_{t_0} \in @S_0) + P(Z_t \in K_{\epsilon=2}) + P(Z_t \in (K_{\epsilon=2})^C) \\ &= P(Z_{t_0} \in @S_0) + P(Z_t \in K_{\epsilon=2}) + \epsilon + 1 - \epsilon=2 = 1 + \epsilon=2; \end{aligned}$$

where in the first inequality we have used the observation in the first bullet point of this note and $(K_{\epsilon=2})^C$ denotes complement in S_{x_0} . Hence $\epsilon = 0$, i.e. Z_t belongs to S_{x_0} almost surely.

- Hypothesis 7.5 [A.6] is the analogous of Hypothesis 6.1. [H.4].

²⁶The second inequality is an inequality rather than an equality because of Lemma 4.15.

- Hypothesis 7.5 [A.7] is slightly more complicated to explain, so we observe that it is satisfied in the representation of the form “ODE + SDE” (6.1)–(6.3) of the previous section, if $X_t = e^{tW_0} x_0$ converges to some x . Indeed in that case if $x_0 = (z_0; \theta_0)$ then $S_{x_0} = H_{\theta_0}$ and $S_{\bar{x}} = H_{\theta}$ (both of these manifolds are n -dimensional hyperplanes in \mathbb{R}^{n+1} , hence they are closed). Moreover, for every $x = (z; \theta) \in S_{x_0}$, $W^1(x) = W^1((z; \theta)) = (z; \theta)$, hence the map W^1 is continuous on S_{x_0} . If $x = (z; \theta) \in S_{\bar{x}}$ then $W^1(x) = x$, so W^1 is continuous on $S_{\bar{x}}$ as well. Because in this case the map W^1 is just a projection on the plane $S_{\bar{x}}$, W^1 is continuous on $S_{x_0} \cup S_{x_0} \cup S_{\bar{x}} = S_{x_0} \cup S_{\bar{x}}$ (the equality holding because $S_{x_0} \cap S_{\bar{x}} = \emptyset$).
- By [A.6] $V_0^{(?)}(x) = 0$; using Lemma 4.1, this implies that $S_x = S_x$. Hence, by Lemma 4.16, $V_0^{(?)}(x) = 0$ for every $x \in S_x$. So in reality [A.6] implies that part of [A.7] where we require $V_0^{(?)}$ to vanish on the whole S_x .
- If we do not make any assumptions on the map W^1 , when we look at the set $W^1(S_{x_0})$, it may occur that this is not a connected set and, even if it were connected, it may be contained in more than one submanifold of Λ (see Example 6.9). If we assume that W^1 is continuous, because S_{x_0} is connected then also $W^1(S_{x_0})$ is; for simplicity, we are also explicitly assuming that $W^1(S_{x_0})$ is contained in just one submanifold of Λ , the manifold $S_{\bar{x}}$. It could also occur that on the limit manifold $W^1(S_{x_0})$ we have that $V_0^{(?)}(x) \neq 0$ for every $x \in W^1(S_{x_0})$, see for instance Example 6.10. If this is the case, then one can take such a manifold as starting manifold and apply the theory that we explain here by taking starting points on this manifold; i.e. one can sort of “repeat the procedure” illustrated here by starting the dynamics again on that manifold. So, in conclusion one just needs to study the case in which $V_0^{(?)}(x) = 0$ for every $x \in W^1(S_{x_0})$. Again for simplicity, we assume $V_0^{(?)}(x) \neq 0$ for every $x \in S_{\bar{x}}$.
- Finally, notice that if W^1 is well defined and continuous on S_{x_0} then W^1 is also a well-defined and continuous map from S_{x_0} to \mathbb{R}^N . We show this fact in Lemma A.6, contained in Appendix A.3. Notice also that W^1 is the identity when restricted to $S_{\bar{x}}$, hence W^1 is always well defined and continuous on $S_{\bar{x}}$. What we are requiring with the last point of Hypothesis 7.5 is that the map should be continuous not only on each one of the manifolds S_{x_0} ; S_{x_0} and $S_{\bar{x}}$, but also that it should be continuous on the union of these three sets. The reason why we need continuity also on the closure of S_{x_0} is, again, technical, see proof of Lemma B.6

Before we consider the behaviour of $X_t^{(x)}$ in the case when $e^{tV_0^{(?)}}(x)$ is convergent, we must first consider the trivial case, i.e. the behaviour of the process when we start it from the “equilibrium manifold” $S_{\bar{x}}$, where $V_0^{(?)}(x) = 0$. We do this in Proposition 7.7 below, which is the analogous of Lemma 6.4.

Proposition 7.7. Let Hypothesis 7.5 [A.1], [A.3] and [A.5] hold. Let S be an integral submanifold of Λ such that $V_0^{(?)}(x) = 0$ on S . Then there exists a unique invariant measure $\bar{\mu}^S$ of P_t supported on \bar{S} such that

$$\lim_{t \uparrow \infty} P_t f(x) = \bar{\mu}^S(f); \text{ for all } x \in S; f \in C_b(\mathbb{R}^N): \quad (7.9)$$

Moreover the convergence is uniform on compact subsets of S ; that is, for every compact set $K \subset S$ and every $f \in C_b(\mathbb{R}^N)$ we have

$$\lim_{t \uparrow \infty} \sup_{x \in K} P_t f(x) - \bar{\mu}^S(f) = 0 :$$

Proof of Proposition 7.7. The proof is deferred to Appendix B.3. □

Note 7.8. The assumption that $V_0^{(?)}(x) = 0$ on S implies that, if $x \in S$, then the map $t \mapsto P_t f(x)$ is differentiable, for any $f \in C_b(\mathbb{R}^N)$. Indeed, as explained in the Introduction, in general we have that $P_t f$ is differentiable in the direction ∂_{V_0} and in the directions contained in $\hat{\Lambda}$ (see Appendix A.2) and satisfies

$$(\partial_{V_0} P_t f) = \sum_{i=1}^d V_i^2 P_t f:$$

However if $V_0^{(?)}(x) = 0$ for all $x \in S$ then $V_0(x) \in \hat{\Lambda}(x)$ for all $x \in S$ and hence $P_t f$ is also differentiable in the direction ∂_{V_0} on S . Therefore we have that P_t is also differentiable in time, i.e. as a map $t \mapsto P_t f$, and satisfies

$$\partial_t P_t f = V_0 P_t f + \sum_{i=1}^d V_i^2 P_t f:$$

□

By Hypothesis [A.7] $V_0^{(?)}(x) = 0$ on $S_{\bar{x}}$ so we can apply Proposition 7.7 to the manifold $S_{\bar{x}}$ and throughout the rest of the section we shall denote by $\mu_{S_{\bar{x}}}$ the invariant measure supported on $S_{\bar{x}}$ such that (7.9) holds for all $x \in S_{\bar{x}}$. Such a measure exists and is unique by Proposition 7.7. Similarly to what we did in Section 6, equation (6.11), we shall extend this to a measure $\mu_{S_{\bar{x}}}$ defined on \mathbb{R}^N by setting

$$\mu_{S_{\bar{x}}}(A) = \mu_{S_{\bar{x}}}(A \cap S_{\bar{x}}); \text{ for any Borel measurable set } A \subset \mathbb{R}^N:$$

For any $\bar{x} \in \mathbb{R}^N$, let $I_0(\bar{x}) = \{x \in \mathbb{R}^N : W^1(S_{\bar{x}})(x) = S_{\bar{x}}\}$. The set I_0 is contained within the basin of attraction for the measure $\mu_{S_{\bar{x}}}$. Indeed, Theorem 6.5 below shows that for all $x \in I_0(\bar{x})$ we have that $P_t f(x)$ converges to $\mu_{S_{\bar{x}}}(f)$, for all $f \in C_b(\mathbb{R}^N)$.

Theorem 7.9. Let Hypothesis 7.5 hold. Let $\bar{x} \in \mathbb{R}^N$ be such that $V_0^{(?)}(x) = 0$. Then there exists an invariant measure $\mu_{S_{\bar{x}}}$ supported on $S_{\bar{x}}$ such that for each $x_0 \in I_0(\bar{x})$, and $f \in C_b(\mathbb{R}^N)$ we have that $P_t f(x_0)$ converges to $\mu_{S_{\bar{x}}}(f)$.

Proof of Theorem 7.9. Throughout the proof we fix an arbitrary point $x_0 \in I_0$. The proof is split into 3 steps.

Step 1 : We first construct a tight evolution system of measures, μ_t , for $t \geq 0$, for the semigroup $Q_{s,t}$ which are supported on S_{x_0} . This can be done by acting analogously to what we have done in Step 1 of the proof of Theorem 6.5; in particular we may define $\mu_t := Q_{0,t} \mu_{x_0}$.²⁷ Note that $\mu_t(S_{x_0}) = 1$; indeed by Note 7.6 (second bullet point) we have that $Z_t \in S_{x_0}$ almost surely when $Z_0 = x_0$; hence for every $t \geq 0$

$$\mu_t(S_{x_0}) = Q_{0,t} \mu_{x_0}(S_{x_0}) = Q_{0,t} 1_{S_{x_0}}(x_0) = P(Z_t \in S_{x_0} | Z_0 = x_0) = 1:$$

Moreover, analogously to Step 2 in the proof of Theorem 6.5, since the family μ_t is tight, there exists a diverging sequence t_n such that μ_{t_n} converges weakly to some probability measure μ_0 as t_n tends to ∞ .

Step 2 : By construction, the measure μ_0 is a measure on S_{x_0} ; we then consider the probability measure $\mu_0(W^1)^{-1}$.²⁸ The latter measure is supported on $S_{\bar{x}}$. One needs to show that $\mu_0(W^1)^{-1} = \mu_{S_{\bar{x}}}$. Recall that $\mu_{S_{\bar{x}}}$ is the restriction of the measure $\mu_{S_{\bar{x}}}$ to $S_{\bar{x}}$. The proof of this fact is deferred to Lemma B.7. Note that this is one of the places where we use that $x_0 \in I_0(\bar{x})$. This implies that μ_t converges weakly to $\mu_{S_{\bar{x}}}$ as t tends to ∞ . Furthermore, by Hypothesis [A.3] we can take a sequence f_t such that $f_t \rightarrow 1$ and $\mu_t^{x_0}$ converges weakly to some probability measure μ_{x_0} .

²⁷ Note that using the same argument we could define $\mu_t = Q_{0,t} \mu_{x_0}$.

²⁸ Here $(W^1)^{-1}(A)$ denotes preimage of A .

Step 3 : We show that μ_{x_0} is supported on \bar{S}_x and, when we restrict it to \bar{S}_x , we have $\mu_{x_0}|_{\bar{S}_x} = \mu_0 (W^1)^{-1}$. Lemma B.8 is devoted to proving this fact. Therefore, by Step 2 and the definition of μ_{x_0} we have that

$$\mu_{x_0} = \mu_{S_x}.$$

This implies that $\mu_t^{x_0}$ converges weakly to μ_{S_x} as t tends to ∞ for any $x \in S_{x_0}$, that is, for every $f \in C_b(\mathbb{R}^N)$, $\int P_t f(x_0) dx$ converges to $\int \mu_{S_x}(f) dx$ as t tends to ∞ . \square

We now give a one dimensional example which satisfies all the assumptions we have made in this section. In particular, this example fits our framework in a non-trivial way as it exhibits many invariant measures.

Example 7.10. Consider the SDE

$$dZ_t^z = \sin(Z_t^z)dt + \frac{1}{2}(1 - \cos(Z_t^z)) dB_t; \quad Z_0 = z; \quad Z_t \in \mathbb{R};$$

where $(B_t)_{t \geq 0}$ is a one-dimensional Wiener process. In this case $V_0 = \sin(z) \otimes (1 - \cos(z)) \otimes$ and we have

$$[V_1; V_0] = [(1 - \cos(z)) \otimes; \sin(z) \otimes] = \cos(z)(1 - \cos(z)) \otimes \sin(z)^2 \otimes = V_1;$$

Therefore the vector fields $V_0; V_1$ satisfy the UFG condition; the above also shows that the obtuse angle condition (3.4) is satisfied, with $\theta_0 = 1$. Moreover, it is easy to show that the function $(V_1 P_t f)(x)$ decays exponentially fast in time, i.e. θ_0 is big enough that (3.4) implies an estimate of the type (3.5) for the fields V_1 . Because the coefficients of the equation are bounded the estimate is uniform on the whole real line, see [14, Proposition 3.1, Proposition 3.4 and Theorem 4.2] alternatively by a direct calculation, see [14, Example 4.4]. Since V_0 and V_1 both vanish whenever $z \in \mathbb{Z}$ we have that the point measures μ_n are invariant measures for any $n \in \mathbb{Z}$. However there also exist invariant measures supported on $(2n; 2(n+1))$ for any $n \in \mathbb{Z}$. Indeed let

$$\mu_n(z) := \frac{\exp\left(\frac{1}{1 - \cos(z)}\right)}{C(1 - \cos(z))} 1_{(2n; 2(n+1))}(z)$$

where C is the normalization constant and $1_{(2n; 2(n+1))}(z)$ is the characteristic function of the interval $[2n; 2(n+1))$. By direct calculation one can verify that, for every $n \in \mathbb{Z}$, $\mu_n(z)$ satisfies the stationary Fokker-Planck equation $L \mu_n = 0$, where

$$L \mu_n(z) = \frac{1}{2}(\sin(z) \mu_n(z))' + \frac{1}{2}[(1 - \cos(z)) \otimes ((1 - \cos(z)) \mu_n(z))]' :$$

Notice that if $X_0 \in [2n; 2(n+1))$ (for some fixed $n \in \mathbb{Z}$) then $X_t \in [2n; 2(n+1))$ for every $t \geq 0$. However, even if we restrict to one of the intervals $[2n; 2(n+1))$, the process still admits three invariant measures on each one of such intervals. \square

Example 7.11 (Example 6.10 continued). Recall that in this example $V_0 = \sin \otimes kz \otimes$ and $V_1 = \otimes$. While V_0 is smooth, $V_0^{(?)}$ is not continuous. Indeed, for $\epsilon > 0$ $V_0^{(?)}(z; \epsilon) = \sin(\epsilon) \otimes$, however for $\epsilon = 0$ $V_0^{(?)}(z; 0) = V_0(z; 0) = kz \otimes$. \square

We conclude this section by stating and proving Theorem 7.12 below. In order to state it, let us define the following equivalence relation on \mathbb{R}^N :

$$x \sim y, \quad x \in S_y :$$

As customary, we denote by $[x]$ the equivalence class of x under the equivalence relation \sim . Note that by Lemma 3.15, if $x \sim y$ then also $e^{tV_0^{(?)}} x \sim e^{tV_0^{(?)}} y$, therefore the flow map

$$[x] \mapsto [e^{tV_0^{(?)}} x] =: e^{tV_0^{(?)}} [x] \tag{7.10}$$

is well defined. Let now q be the map $q: \mathbb{R}^N \rightarrow \mathbb{R}^N = \mathbb{R}^N / \sim$, defined as $q(x) = [x]$. If we endow the quotient set $\mathbb{R}^N = \mathbb{R}^N / \sim$ with the σ -algebra $\mathcal{F} = \{E \subset \mathbb{R}^N = \mathbb{R}^N / \sim : q^{-1}(E) \text{ is a Borel set of } \mathbb{R}^N\}$, then q is a measurable map. If μ is a probability measure on \mathbb{R}^N , we define the pullback measure $\tilde{\mu}$ on $\mathbb{R}^N = \mathbb{R}^N / \sim$ as $\tilde{\mu}(E) = \mu(q^{-1}(E))$ for all $E \subset \mathbb{R}^N = \mathbb{R}^N / \sim$.

Theorem 7.12. Consider the SDE (1.1) and the associated semigroup P_t and assume that the vector fields V_0, \dots, V_d satisfy the UFG condition. If μ is an invariant measure for P_t , then $\tilde{\mu}$ is an invariant measure for the flow map (7.10).

Proof of Theorem 7.12. Denote by $B_b(\mathbb{R}^N = \mathbb{R}^N / \sim; \mathbb{R})$ to be the set of all bounded and measurable functions $f: \mathbb{R}^N = \mathbb{R}^N / \sim \rightarrow \mathbb{R}$. If $f \in B_b(\mathbb{R}^N = \mathbb{R}^N / \sim; \mathbb{R})$, then $f \circ q \in B_b(\mathbb{R}^N; \mathbb{R}^N)$, i.e. $f \circ q$ is a bounded and measurable function mapping from \mathbb{R}^N to \mathbb{R}^N . By the definition of invariant measure, we have

$$\begin{aligned} \int_{\mathbb{R}^N = \mathbb{R}^N / \sim} f([x]) \tilde{\mu}(d[x]) &= \int_{\mathbb{R}^N} f(q(x)) \mu(dx) \\ &= \int_{\mathbb{R}^N} (P_t(f \circ q))(x) \mu(dx) \\ &= \int_{\mathbb{R}^N} E_x[f(q(X_t^x))] \mu(dx); \end{aligned}$$

Let us now look more closely at the expected value on the right hand side of the above: for any bounded and measurable function h we can write

$$\begin{aligned} \int_{\mathbb{R}^N} E_x[h(X_t^x)] \mu(dx) &= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} h(y) P_x(X_t^x \in dy) \mu(dx) \\ &= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \frac{h(y) P_x(X_t^x \in dy)}{S_{e^{tV_0}(\cdot)}(x)} \mu(dx) \\ &= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} h(y) P_x(X_t^x \in dy) \mu(dx) \\ &\quad + \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} h(y) P_x(X_t^x \in dy) \mu(dx); \end{aligned}$$

where the second equality follows from Proposition 5.3. Now the second term in the above vanishes by Proposition 5.7 (easy to prove for positive h , if h is not positive just split into positive and negative part). Indeed if $X_t^x \in \mathbb{R}^N \setminus \mathbb{R}^N$ then $X_t^x \in \mathbb{R}^N$ (see Lemma 4.15 for a proof of this fact). Putting everything together we can write

$$\begin{aligned} \int_{\mathbb{R}^N = \mathbb{R}^N / \sim} f([x]) \tilde{\mu}(d[x]) &= \int_{\mathbb{R}^N} E_x[f(e^{tV_0}(\cdot))([x])] \mu(dx) \\ &= \int_{\mathbb{R}^N} f(e^{tV_0}(\cdot))(x) \mu(dx) \\ &= \int_{\mathbb{R}^N = \mathbb{R}^N / \sim} f(e^{tV_0}(\cdot))([x]) \tilde{\mu}(d[x]); \end{aligned}$$

where the penultimate equality follows from the fact that the object on the second line is completely deterministic and the last equality holds by the definition of the measure $\tilde{\mu}$. This concludes the proof. □

8 Existence of a density

Analogously to what we did for the study of the long-time behaviour, we split this section into two subsections. That is, in Section 8.1 we consider the setting of Section 6

and study SDEs of the form (6.1)–(6.3). In Section 8.2 we consider the general UFG-case. This section makes use of several notions from Malliavin calculus, we will recall only some basic facts and refer the reader to [46] for more detailed background material.

Note 8.1. The techniques of proof that we use here are not different from those in [46, 60, 7]; however our assumptions are significantly more general, so we can not just quote results and some proofs need to be re-sketched. Moreover, the results of this section can be seen as a generalization of the results of [7, Section 5.2] to the UFG case. We indeed recall that in [7, Section 5.2] the author assumes the validity of the Hörmander condition, which we have recalled in Section 3, as well as boundedness of the coefficients of the SDE. Here we remove both such assumptions. We emphasize that, by a geometric point of view, imposing the validity of the Hörmander condition amounts to assuming that the distribution μ_0 is equal to \mathbb{R}^N at every point. □

Let $D^{k;p} L^p(\cdot)$ denote the Malliavin Sobolev space, that is the domain of the k th order Malliavin derivative in the space $L^p(\cdot)$. We also define the space

$$D = \bigcup_{p > 1; k \geq 2N} D^{k;p}$$

We shall denote by D^0 the dual space of D , that is the space of all continuous linear maps from D to \mathbb{R} . Let us recall the following lemma, which is quoted from [46, Theorem 2.2.1].

Lemma 8.2. Fix $T > 0$, let $(X_t)_{t \in [0;T]}$ denote the solution of the SDE (1.1) and assume that V_0, V_1, \dots, V_d are smooth vector fields which are globally Lipschitz. Then X_t^i belongs to $D^{1;p}$ for any $t \in [0; T]$; $p \geq 1$ and $i = 1, \dots, N$. Moreover, for all $0 \leq r < t \leq T$; $p \geq 1$

$$\sup_{0 \leq r < t \leq T} E \sup_{r \leq s \leq T} |D_r^j X_s^i|^p < \infty$$

and the Malliavin derivative $D_r^j X_t^i$ satisfies the following SDE,

$$\begin{aligned} D_r^j X_t^i &= V_j^i(X_r) + \sum_{k=1}^d \int_r^t \partial_k V_0^i(X_s) D_r^j(X_s^k) ds \\ &+ \sum_{k=1}^d \int_r^t \partial_k V^i(X_s) D_r^j(X_s^k) dW_s^k; \end{aligned} \tag{8.1}$$

for every $r < t$.

Here we use the notation D^k to denote the Malliavin derivative operator with respect to the Brownian motion B^k .²⁹ Define the Malliavin matrix $M_t = (M_t^{ij})_{i,j=1}^N$ to be

$$M_t^{ij} = \sum_{k=1}^d \int_0^t \partial_k^i(X_s^j) \partial_k^j(X_s^i) ds$$

Again by [46, Section 2.3] we can rewrite the Malliavin matrix in terms of the Jacobian matrix $J_t := \frac{\partial X_t}{\partial x}$, details can be found in [46, Section 2.3]. There it is also shown that J_t is an invertible matrix and that the following holds

$$M_t = J_t \sum_{k=1}^d \int_0^t V_k(X_s) V_k(X_s)^T (J_s^{-1})^T ds \quad J_t^{-1} = J_t^{-1} C_t J_t^{-1}$$

²⁹Note that D^k denotes the 1st order Malliavin derivative with respect to the k th Brownian motion and is not to be confused with the k th-order Malliavin derivative.

where the matrix C_t is the reduced Malliavin covariance matrix defined as

$$C_t = \sum_{k=1}^d \int_0^t V_k(X_s) V_k(X_s)^T (J_s^{-1})^T ds;$$

8.1 Existence of a density on a suitable hyperplane

In this section we consider the SDE (6.1)–(6.3). We shall also assume Hypothesis 6.1 [H.1], which states that the set of vector fields $\{V_l(z_0; 0) : 2 \leq l \leq m\}$ span the n -dimensional hyperplane $H_0 := \{x = (z; \cdot) : \cdot = 0\}$ for all $(z_0; 0) \in \mathbb{R}^n \times \mathbb{R}$. In this setting it is clear that the law of $X_t = (Z_t; \cdot)$ does not admit a density with respect to Lebesgue measure on \mathbb{R}^{n+1} ; indeed for each fixed t , \cdot is a deterministic point which implies that $P_x(X_t \in \mathbb{R}^n \setminus \{\cdot\}) = 1$ while $\mathbb{R}^n \setminus \{\cdot\}$ is a null set with respect to Lebesgue measure on \mathbb{R}^{n+1} . We prove that for every fixed $t \geq 0$ the law of the random variable Z_t admits a density with respect to Lebesgue measure on \mathbb{R}^n . In terms of the process X_t this implies that the law of X_t admits (for every fixed $t \geq 0$) a density with respect to the Lebesgue measure on the hyperplane $H_t := \{x = (z; \cdot) : \cdot = t\}$. Moreover, since from Section 6 $X_t \in H_t$ almost surely, we have that H_t is the maximal manifold such that X_t admits a density with respect to the volume element on such a manifold. ³⁰

To prove that the law of Z_t admits a density we shall follow the same strategy of [46, Section 2.3]. Note that by Hypothesis 3.16 and Lemma 8.2 for each $t \geq 0$ and $i \in \{1, \dots, n\}$ we have that Z_t^i and \cdot belong to $D^{1,p}$ for all $p \geq 1$. First we note that the solution $X_t = (Z_t; \cdot)$ admits a Malliavin derivative.

Lemma 8.3. Let M_t denote the Malliavin matrix corresponding to the solution $X_t = (Z_t; \cdot)$ of the SDE (6.1)–(6.3). Then M_t has the form

$$M_t = \begin{pmatrix} M_t & 0 \\ 0 & 0 \end{pmatrix} \tag{8.2}$$

where the matrix M_t is the Malliavin matrix corresponding to Z_t .

Proof of Lemma 8.3. The proof is deferred to Appendix B.4. □

In [46] it is shown that if the Malliavin matrix is invertible then the law of X_t admits a density on \mathbb{R}^{n+1} . We can see from (8.2) that the matrix is not invertible; however we show that the Malliavin matrix M_t corresponding to Z_t is invertible almost surely and hence the law of Z_t admits a density on \mathbb{R}^n , for every fixed $t > 0$.

Proposition 8.4. The reduced Malliavin covariance matrix C_t corresponding to the solution $X_t = (Z_t; \cdot)$ of the SDE (6.1)–(6.3) is of the form

$$C_t = \begin{pmatrix} C_t & 0 \\ 0 & 0 \end{pmatrix};$$

where C_t is a random $n \times n$ symmetric matrix. Moreover, if we assume Hypothesis 6.1 [H.1] holds then C_t is invertible P -almost surely.

Proof of Proposition 8.4. The proof is deferred to Appendix B.4. □

Theorem 8.5. Assume Hypothesis 6.1 [H.1] and let $f(z_t, g_t, 0)$ be the solution of (6.1). Then the law of Z_t is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}^n .

³⁰Throughout our discussion we need to fix a canonical reference measure on the manifold. Here, and subsequently, when we refer to the volume element on a submanifold $M \subset \mathbb{R}^N$ we mean the measure on M which is determined from the Riemannian density associated to the induced Riemannian metric on M .

Proof of Theorem 8.5. Note the Malliavin matrix corresponding to Z_t is M_t which is invertible, indeed $M_t = J_t C_t J_t^T$ and C_t is invertible by Proposition 8.4 therefore M_t is invertible since the product of invertible matrices is invertible. By [46, Theorem 2.1.2] we have that the law of Z_t is absolutely continuous with respect to Lebesgue measure on \mathbb{R}^n , for each $t > 0$. □

8.2 Existence of a density on integral submanifolds

We now return to studying the general UFG-case. As in the previous section we cannot expect that the law of X_t will in general admit a density with respect to Lebesgue measure on \mathbb{R}^N and we will instead show that the law of X_t admits a density with respect to the volume element on a suitable manifold. Indeed, we shall show that the law of X_t admits a density with respect to the volume element on $S_{e^{tV_0^{(?)}}(x_0)}$. Note that by

Proposition 5.3 we have $X_t^{(x)} \in \overline{S_{e^{tV_0^{(?)}}(x)}}$ almost surely. In this section we shall assume Hypothesis 7.5 [A.1] and that X_t cannot hit the boundary of the integral manifold $S_{e^{tV_0^{(?)}}(x)}$, that is $X_t^{(x)} \in S_{e^{tV_0^{(?)}}(x)}$, almost surely. In the first and second comment in Note 7.6 it is shown that under Hypothesis 7.5 [A.4] implies that X_t cannot hit the boundary of the maximal integral submanifold.

Recall from Section 7 the process $fZ_t g_t$ defined by (7.2). Since $e^{tV_0^{(?)}}$ is a diffeomorphism the law of X_t admits a density with respect to the volume element on $S_{e^{tV_0^{(?)}}(x_0)}$ if and only if the law of Z_t admits a density with respect to the volume element on S_{x_0} . Let $V_{[1];t}$ be defined as in (7.6), then recall that the process $fZ_t g_t$ satisfies the SDE (7.4). Now we wish to apply [55, Theorem 3.4] to show that the law of $fZ_t g_t$ admits a density with respect to the volume measure on S_{x_0} . However, as noted in [9], there is a mistake in the proof of [55, Theorem 3.4], in particular the form of the Hörmander condition given by [55, Assumption (H)] is not sufficient for the conclusions of [55, Theorem 3.4] to hold. More precisely, they rely upon [20, Theorem 1.1.3] to show that [55, Assumption (H)] implies a suitable integration by parts formula, which is shown to be incorrect by [9]. However under our conditions there is an integration by parts formula as shown in [44, Section 3]. Therefore we may use the strategy given in [55] and the results of [44] to prove that the law of Z_t admits a density with respect to the volume measure on S_{x_0} .

A vital tool for this argument is the integration by parts formula proved in [44, Theorem 3.10]; namely, for $Z \in D$ and $\alpha_1, \dots, \alpha_M \in A_m$ we have

$$E_x [V_{[1];t} \dots V_{[M];t} f(X_t)] = t^{\frac{k_1 + \dots + k_M}{2}} E_x [\alpha_1, \dots, \alpha_M (t; x) f(X_t)];$$

for any $f \in C_V(\mathbb{R}^N)$ and for some random variable $\alpha_1, \dots, \alpha_M (t; x)$. By taking $f = g \circ e^{tV_0^{(?)}}$ we have for any $g \in C_V^1(\mathbb{R}^N)$

$$E_x [V_{[1];t} \dots V_{[M];t} g(Z_t)] = t^{\frac{k_1 + \dots + k_M}{2}} E_x [\alpha_1, \dots, \alpha_M (t; x) g(Z_t)]; \tag{8.3}$$

Let us denote by $E(S_{x_0})$ the space of all distributions (in this sentence distribution is meant in an analytic sense) on S_{x_0} with compact support. Recall that for any smooth function f we can view this as a member of $E(S_{x_0})$, denoted F_f , by setting

$$hF_f; i = \int_S f(x) \nu_{S_{x_0}}(dx); \text{ for any } f \in C_c^1(S_{x_0})$$

where $\nu_{S_{x_0}}$ denotes the volume measure on S_{x_0} .

Lemma 8.6. Assume that Z_t satisfies (8.3). Then there exists a map $\alpha_t : E(S_{x_0}) \rightarrow D^0$ with the following properties

1. If $f \in C_c^1(S_{x_0})$ then $\tau_t(f) = f(Z_t)$. Note that $f(Z_t)$ is identified as an element in D^0 by setting $h_t(f; G) = E[f(Z_t)G]$ for any $G \in D$.
2. The map τ_t is continuous as a map from $E(S_{x_0})$ to D^0 .

For a proof see [58, Proposition 2.1].

Now we shall state some properties of the map τ_t , as proven in [60, Proposition 2].

Proposition 8.7. Fix $t > 0$ and let Z_t be such that the map τ_t is well defined for every $f \in E(S_{x_0})$. Then let I be some open set

1. If $I \ni s \mapsto F_s$ is continuous (continuously differentiable), then $I \ni s \mapsto \tau_t(F_s)$ is continuous (resp. continuously differentiable). In particular, for every $G \in D$ the map $I \ni s \mapsto h_t(F_s; G)$ is continuous and respectively continuously differentiable and

$$\tau_t \left(\frac{dF_s}{ds}; G \right) = \frac{d}{ds} h_t(F_s; G) :$$

2. If $I \ni s \mapsto F_s$ is continuous then for every $G \in D$

$$\tau_t \left(\int_I F_s ds; G \right) = \int_I h_t(F_s; G) ds$$

where $\int_I T_s ds$ is a tempered distribution (here distribution is meant in an analytic sense) and is defined by $h_t \left(\int_I T_s ds; \varphi \right) = \int_I h_t(T_s; \varphi) ds$.

We can show that the law of Z_t admits a density.

Proposition 8.8. Assume Hypothesis 7.5 [A.1], and assume that $X_t^{(x_0)} \in S_{e^{V_0}(\cdot)}(x_0)$ almost surely. Then for each $t > 0$ the law of $Z_t^{(x_0)}$ admits a density with respect to the volume element on S_{x_0} .

Proof. Note that the map $x \mapsto X_t(x)$ is smooth, moreover its (weak) derivative with respect to x^i is given by $D_i X_t(x)$, where $D_i X_t(x)$ is defined by $h D_i X_t(x); \varphi = \partial_i (X_t(x) \varphi)$ for all φ . Therefore $\tau_t(X_t(x))$ is smooth and in particular $p(x) := h_t(X_t(x); 1)$ is smooth. It remains to show that $p(x)$ is the density of the law of X_t . Take $f \in C_c^1(S)$ then

$$\begin{aligned} \int_{S_{x_0}} (x) p(x) S_{x_0}(dx) &= \int_{S_{x_0}} (x) h_t(X_t(x); 1) S_{x_0}(dx) \\ &= h_t \left(\int_{S_{x_0}} (x) X_t(x) S_{x_0}(dx); 1 \right) \end{aligned}$$

Now for $f \in C_c^1(S_{x_0})$ we have

$$\int_{S_{x_0}} (x) X_t(x) S_{x_0}(dx); f = \int_{S_{x_0}} (x) h_t(X_t(x); f) S_{x_0}(dx) = \int_{S_{x_0}} (x) f(X_t(x)) S_{x_0}(dx) :$$

Therefore $\int_{S_{x_0}} (x) X_t(x) S_{x_0}(dx) = F$, and in particular $\int_{S_{x_0}} (x) X_t(x) S_{x_0}(dx) = (Z_t)$. Now we have

$$\int_{S_{x_0}} (x) p(x) S_{x_0}(dx) = h(Z_t; 1) = E[(Z_t)] : \quad \square$$

Theorem 8.9. Assume the vector fields $V_0; V_1; \dots; V_d$ are uniformly Lipschitz, satisfy the UFG condition and assume that $X_t^{(x_0)} \in S_{e^{V_0}(\cdot)}(x_0)$ almost surely.³¹ Then for each $t > 0$ the law of $X_t^{(x_0)}$ admits a density with respect to the volume element on $S_{e^{V_0}(\cdot)}(x_0)$.

³¹As we have already mentioned, the latter fact follows for example from assuming Hypothesis 7.5 [A.4].

A Some technical results

We gather in this appendix some auxiliary results. In particular, Appendix A.1 contains background material about the topology of the orbits of finitely generated smooth distributions. Appendix A.2 reports some known smoothing results on UFG semigroups, which are often used in the proofs of Appendix B. Appendix A.3 contains precise statements and proofs of further technical facts which would have been cumbersome (and detracting from the main line of thought) if presented in the main body of the work.

A.1 Topology of orbits

Here we give a brief justification of the reason why we make the standing assumption [SA.2]. In short, assuming that the manifold topology of the manifolds S is the Euclidean topology is equivalent to assuming that such manifolds are embedded manifolds. In full generality, as explained in [28, page 78], elements of a global partition induced by distributions which enjoy the integral manifold property are immersed manifolds. We briefly explain the difference between an embedded and an immersed manifold. A detailed treatment of the matter can be found in [28, Appendix A.2 and Appendix A.4]. Let $F : M_1 \rightarrow \mathbb{R}^N$ be a continuous mapping of topological spaces and let $M_2 = F(M_1)$. M_2 can be endowed with two topologies: i) the topology of M_2 as a subset of the Euclidean space \mathbb{R}^N , so that the open sets in this topology are the sets O of the form $O = O^0 \setminus M_2$ for some O^0 which is open in the Euclidean topology of \mathbb{R}^N ; ii) the topology induced by M_1 , where the open sets are the sets U of the form $U = F(U^0)$, for some U^0 which is open in the topology of M_1 . In general, the latter topology is stronger than the former. With this premise, one can give the following definition.

Definition A.1. Let $F : M_1 \rightarrow \mathbb{R}^N$ be a smooth mapping of manifolds. F is an immersion if it is injective and $\text{rank}(J_p F) = \dim(M_1)$ for every $p \in M_1$. F is an embedding if it is an immersion and the topology induced on $M_2 = F(M_1)$ by the one on M_1 coincides with the Euclidean topology of M_2 as a subset of \mathbb{R}^N .

The reason why we consider only the case in which the manifolds of the partition are embeddings comes mostly from the need to use the Stroock and Varadhan support theorem: the closure appearing in the statement of such a theorem is intended in Euclidean sense. If the manifold topology was not the Euclidean topology we would have to consider two closures, the closure in the Euclidean topology and the closure in the manifold topology. This would make the exposition much more cloudy. Moreover we point out that in all our examples the manifolds at hand are embedded manifolds. It is possible that, under the assumption of this paper that the vector $V_0^{(?)}$ is smooth and Lipschitz and that the integral curves of $V_0^{(?)}$ are convergent, one may prove that the orbits S are indeed embedded manifolds. But this is beyond the scope of this paper.

A.2 Known facts about UFG semigroups

In this appendix we gather some known facts that we use frequently.

[F.1] A semigroup P_t of bounded operators is Markov if

$$P_t 1 = 1 \quad \text{and} \quad P_t f \geq 0 \quad \text{when} \quad f \geq 0;$$

where, in the above, 1 denotes the function identically equal to one. Denoting by $\|k\|_{k_1}$ the supremum norm, the above implies that if $\|k\|_{k_1} < 1$ then $\|k P_t f\|_{k_1} \leq \|k\|_{k_1}$, i.e. the semigroup is a contraction in the supremum norm. Similarly the two parameter semigroups $f_{Q_{s,t}} g_{0_{s,t}}$ and $f_{Q_{s,t}} g_{0_{s,t}}$, considered in Section 6 and Section 7, are both contractive in the supremum norm.

[F.2] Note that if the vector fields V_0, V_1, \dots, V_d satisfy the parabolic Hörmander condition then for any $f \in C_b(\mathbb{R}^N)$, the semigroup $P_t f(x)$ is smooth in all directions in \mathbb{R}^N and moreover it is smooth in t . This is not generally the case if we assume the UFG condition. However we have that for any $f \in C_b(\mathbb{R}^N)$ and $t > 0$ the function $x \mapsto P_t f(x)$ is differentiable in the directions $V_{[j]}$ for any $j \in A$. Moreover for any compact set K , $t > 0$ there exists $C(K) > 0; \epsilon > 0$ such that

$$\sup_{x \in K} |V_{[j]} P_t f(x) - C(K) e^{\epsilon t} t^{-k} \mathbb{1}_{j \in A}| < \epsilon$$

If the vector fields $V_{[j]}$ are bounded then the above estimate holds uniformly on \mathbb{R}^N , for details see [44, Chapter 3]. In contrast to the case in which the parabolic Hörmander condition is enforced, when the UFG condition holds $P_t f$ need not be differentiable in the direction V_0 ; however it is differentiable in the direction $\mathbb{1}_{V_0}$. For more details see [14, Appendix A].

[F.3] For $f \in C_V^1(\mathbb{R}^N)$ (the set $C_V^1(\mathbb{R}^N)$ has been defined in Section 2) we have that $(x; t) \mapsto P_t f$ is smooth in both x and t , i.e. it is differentiable arbitrarily many times in every direction, see [10]. When $f \in C_b(\mathbb{R}^N)$ we may take a sequence $f_n \in C_V^1(\mathbb{R}^N)$ such that $P_t f_n \in C_V^1(\mathbb{R}^N)$ and for each compact set $K \subset \mathbb{R}^N$ we have that $P_t f_n$ and $V_{[j_1]} \dots V_{[j_k]} P_t f_n$ converge uniformly over K as n tends to ∞ to $P_t f$ and $V_{[j_1]} \dots V_{[j_k]} P_t f$ respectively for each $k \in \mathbb{N}; j_1, \dots, j_k \in A$. We shall denote by $D_V^{2;1}(\mathbb{R}^N)$ the space of all functions that can be approximated with the procedure just described. From what we have just said, the semigroup $P_t f$ belongs to $D_V^{2;1}(\mathbb{R}^N)$ for any $f \in C_b$. See [14, Appendix A] for more details.

A.3 Miscellaneous technical facts

Lemma A.2. Let X and Y be as in Example 3.8. Then the vector fields $f; X; Y; g$ do not satisfy the UFG condition, in the sense that whether we take $X = V_0$ and $Y = V_1$ or viceversa, the UFG condition is not satisfied.

Proof of Lemma A.2. In the definition of UFG condition take $Y = V_0$ and $X = V_1$ (the other case is simple to show) and assume that the UFG condition holds for some $m \in \mathbb{N}$. Denote by ad_X the map which takes a vector field Z to $[X; Z]$, then note that

$$(\text{ad}_X)^k Y = \mathbb{1}^{(k)}(x) \mathbb{1}_{V_0} \tag{A.1}$$

Here $\mathbb{1}^{(k)}$ denotes the k^{th} derivative of $\mathbb{1}$. Now $(\text{ad}_X)^k Y$ commutes with the vector field Y and hence the only non-trivial vector fields in \mathbb{R}^m are $X; Y$ and $(\text{ad}_X)^k Y$ for any $k \in \mathbb{N}$. By the UFG condition there exist smooth functions $\mathbb{1}'_X; \mathbb{1}'_{Y;k}$ such that

$$(\text{ad}_X)^{m+1} Y = \mathbb{1}'_X X + \sum_{k=0}^m \mathbb{1}'_{Y;k} (\text{ad}_X)^k Y$$

We may write this as follows using (A.1)

$$\mathbb{1}^{(m+1)} \mathbb{1}_{V_0} = \mathbb{1}'_X \mathbb{1}^{(m+1)} \mathbb{1}_{V_0} + \sum_{k=0}^m \mathbb{1}'_{Y;k} \mathbb{1}^{(k)} \mathbb{1}_{V_0}$$

By considering the direction $\mathbb{1}_{V_0}$ we have that $\mathbb{1}'_X \mathbb{1}_{V_0} = 0$, therefore we have

$$\mathbb{1}^{(m+1)} \mathbb{1}_{V_0} = \sum_{k=0}^m \mathbb{1}'_{Y;k} \mathbb{1}^{(k)} \mathbb{1}_{V_0}$$

Also note that since $\phi(x) = 0$ for all $x < 0$ we have that $\phi^{(k)}(x) = 0$ for all $x < 0$ and $k \in \mathbb{N}$; as ϕ is smooth this gives that $\phi^{(k)}(0) = 0$ for all $k \in \mathbb{N}$. In particular, ϕ solves the following initial value problem

$$\begin{aligned} \phi^{(m+1)}(x) &= \sum_{k=0}^n f_{\gamma;k}(x; y) \phi^{(k)}(x); & \text{for all } x \geq 0 \\ \phi^{(k)}(0) &= 0; & \text{for all } k \in \{0, 1, 2, \dots, m\}. \end{aligned}$$

However since ϕ is smooth and the functions $f_{\gamma;k} g_{k,0}$ are smooth, there is a (at least locally) unique solution to this initial value problem; the function which is constantly zero clearly satisfies the initial value problem. Therefore we have that $\phi = 0$ (in a neighbourhood of zero), which gives a contradiction and hence the UFG condition is not satisfied. \square

Lemma A.3. Assume that the vector fields V_0, \dots, V_d satisfy the UFG condition. Let S be a maximal integral submanifold of \mathcal{A}_0 and let $x, y \in S$. Assume that x, y lie in the same coordinate neighbourhood U_{x_0} of a coordinate transformation ψ_{x_0} constructed in Section 4.2. Then x and y lie in the same maximal integral submanifold of \mathcal{A} if and only if $\psi_{x_0}^{n+1}(x) = \psi_{x_0}^{n+1}(y)$.

Proof of Lemma A.3. Assume that x, y both lie in the same maximal integral submanifold S of \mathcal{A} . Then there is a time $T > 0$ and a path $p : [0, T] \rightarrow S$ satisfying the following ODE

$$\dot{p}(t) = \sum_{j \in A_m} V_j(p(t)) \quad (t); \quad p(0) = x; \quad p(T) = y$$

for some piecewise linear input functions $u : [0, T] \rightarrow \mathbb{R}$.

Now let $\tilde{p}(t) = \psi_{x_0}(p(t))$ and let \tilde{V} denote the representation of V in the coordinates defined by ψ_{x_0} , then we have

$$\dot{\tilde{p}}(t) = \sum_{j \in A_m} \tilde{V}_j(\tilde{p}(t)) \quad (t):$$

Now by the properties in Proposition 4.9 we have that $\tilde{V}_j^{n+1} = 0$ for all $j \in A_m$, and hence

$$\tilde{p}^{n+1}(y) = \tilde{p}^{n+1}(T) = \tilde{p}^{n+1}(0) = \tilde{p}^{n+1}(x):$$

Now assume that $\tilde{p}^{n+1}(x) = \tilde{p}^{n+1}(y)$.

Let $\tilde{\gamma}$ be any smooth curve that is contained in $(\psi_{x_0}^{-1}(U_{x_0})) \setminus (\mathbb{R}^n \setminus \psi_{x_0}^{-1}(x)g)$, and let $\tilde{\gamma}(0) = \tilde{v}$. Define $\tilde{z} = \psi_{x_0}^{-1}(\tilde{\gamma})$ and $v = \tilde{z}(0)$. Now we have

$$\tilde{z}(0) = J_{z, x_0}^{-1}(\tilde{\gamma}(0))\tilde{v} = J_{z, x_0}^{-1}(\tilde{\gamma}(0))\tilde{v}:$$

Since $\tilde{\gamma}$ is contained within $\mathbb{R}^n \setminus \psi_{x_0}^{-1}(x)g$ we have that $\tilde{v} \in \mathbb{R}^n \setminus 0g$ and hence $v \in \hat{\mathcal{A}}(x)$. Therefore the tangent space to $\psi_{x_0}^{-1}(\text{Im}(\psi_{x_0}) \setminus (\mathbb{R}^n \setminus \psi_{x_0}^{-1}(x)g))$ at each point x^0 in this set is $\hat{\mathcal{A}}(x^0)$. Therefore $J_{z, x_0}^{-1}(\text{Im}(\psi_{x_0}) \setminus (\mathbb{R}^n \setminus \psi_{x_0}^{-1}(x)g)) \subset S_x$, where S_x is the maximal integral submanifold of \mathcal{A} which passes through x . In particular, we have that $y \in S_x$ as required. \square

Lemma A.4. Assume the vector fields V_0, \dots, V_d satisfy the UFG condition. Let $x, y \in \mathbb{R}^N$ be connected by an integral curve of one of the vector fields $V_j; j \in A_m$; that is, $y = e^{TV_j}(x)$ for some $T > 0$ and $j \in A_m$. Then, for all $h \in D_V^{2;1}(\mathbb{R}^N)$,³² we have

$$h(y) - h(x) = \int_0^T (V_j h)(\gamma(s)) ds:$$

³²We recall that the set $D_V^{2;1}(\mathbb{R}^N)$ has been introduced in Appendix A.2 [F.3].

Proof of Lemma A.4. By definition of directional derivative, and because h is differentiable in the directions $V_{[j]}, 2 A_m$, one has

$$\frac{d}{ds}h(s) = V_{[j]}h(s):$$

Integrating from 0 to T , and using that $x(0) = x$ and $x(T) = y$, the statement follows. \square

Lemma A.5. With notation of Section 6, suppose Hypothesis 6.1 [H.1] holds. For any $g \in C_b(\mathbb{R}^n)$ define the functions $f(z, \cdot) = g(z)$ and $v_s(z; t) := P_t f(z; s, t)$. Then v_s is smooth as a map from $\mathbb{R}^n \times (0; 1)$ to \mathbb{R} ; moreover it satisfies

$$\partial_t v_s(z; t) = U_0 v_s(z; t) + \sum_{i=1}^{X^d} U_i^2 v_s(z; t): \tag{A.2}$$

Proof of Lemma A.5. Note that $z \in \mathbb{R}^n \neq \emptyset$ $v_s(z; t)$ is smooth (in any direction in \mathbb{R}^n) for each fixed $t > 0$ since $P_t f$ is differentiable in all the directions spanned by $V_{[j]}$ for all j (which span \mathbb{R}^n). To see that $t \neq 0$ $v_s(z; t)$ is differentiable we first consider the case when f belongs to $C_V^1(\mathbb{R}^{n+1})$ then $t \neq 0$ $P_t f$ is differentiable (see Appendix A.2 [F.3]) and hence $t \neq 0$ $v_s(z; t)$ is differentiable. Moreover using (B.15) we may differentiate v_s to find

$$\begin{aligned} \partial_t v_s(z; t) &= V_0 P_t f(z; s, t) + \sum_{i=1}^{X^d} V_i^2 P_t f(z; s, t) - W_0(s, t) \partial_t P_t f(z; s, t) \\ &= V_0 P_t f(z; s, t) + \sum_{i=1}^{X^d} V_i^2 P_t f(z; s, t) - V_0^{(?) } P_t f(z; s, t) \\ &= (V_0 - V_0^{(?) }) P_t f(z; s, t) + \sum_{i=1}^{X^d} V_i^2 P_t f(z; s, t): \end{aligned}$$

Now using the equality $V_0^{(\wedge)} = V_0 - V_0^{(?)}$ (see (1.9)), we have

$$\partial_t v_s(z; t) = V_0^{(\wedge)} v_s(z; t) + \sum_{i=1}^{X^d} V_i^2 v_s(z; t):$$

Note that, as differential operators, $V_0^{(\wedge)} = U_0$ and $V_i = U_i$ therefore we have that v_s satisfies (A.2).

To extend the proof to the case when f belongs to $C_b(\mathbb{R}^{n+1})$ we apply the argument of [14, Appendix A], so we only sketch this part of the proof. By Appendix A.2 [F.3] if $f \in C_b(\mathbb{R}^{n+1})$ may take a sequence $f_n \in C_V^1(\mathbb{R}^{n+1})$ such that f_n converges to f and $V_{[1]} \dots V_{[k]} P_t f_n$ converges uniformly on compacts of $\mathbb{R}^{n+1} \times (0; 1)$ to $V_{[1]} \dots V_{[k]} P_t f$ for any $k \geq 1$, and $1; \dots; k \in A_m$. By the above argument we have

$$\partial_t (P_t f_n(z; s, t)) = U_0 P_t f_n(z; s, t) + \sum_{i=1}^{X^d} U_i^2 P_t f_n(z; s, t):$$

For any $h > 0$ we have

$$\begin{aligned} \frac{P_{t+h} f_n(z; s, (t+h)) - P_t f_n(z; s, t)}{h} &= \frac{1}{h} \int_t^{t+h} U_0 P_r f_n(z; s, r) \\ &\quad + \sum_{i=1}^{X^d} U_i^2 P_r f_n(z; s, r) dr; \end{aligned}$$

therefore, letting n tend to ∞ , we obtain

$$\frac{P_{t+h}f(z; s_{t+h}) - P_t f(z; s_t)}{h} = \frac{1}{h} \int_t^{t+h} U_0 P_r f(z; s_r) + \sum_{i=1}^d U_i^2 P_r f(z; s_r) dr.$$

Letting now h tend to 0 we have that $(P_t f)(z; s_t)$ is differentiable with respect to t and moreover

$$\partial_t(P_t f(z; s_t)) = U_0 P_t f(z; s_t) + \sum_{i=1}^d U_i^2 P_t f(z; s_t).$$

That is, v_s is differentiable in both z and t as a map from $\mathbb{R}^n \times (0; 1)$ to \mathbb{R} and satisfies (A.2). □

Lemma A.6. With the notation of Section 7, if the map W^1 is well defined on S_{x_0} (in the sense that $S_{x_0} \subset \text{Dom}(W^1)$) and it is continuous when restricted to S_{x_0} , then W^1 is also well defined and continuous on S_{x_0} .

Proof of Lemma A.6. First note that given any point $x \in S_{x_0}$ we can find some $s \in \mathbb{R}$ and $z \in S_{x_0}$ such that $x = e^{sV_0^{(?)}}(z)$, in which case we have

$$W^1(x) = \lim_{t \downarrow 1} e^{tV_0^{(?)}}(x) = \lim_{t \downarrow 1} e^{(t+s)V_0^{(?)}}(z) = W^1(z). \tag{A.3}$$

Now $W^1(z)$ is well defined by assumption and hence $W^1(x)$ is well-defined.

To show that W^1 is continuous on S_{x_0} take $x_k \in S_{x_0}$ and $x \in S_{x_0}$ such that $x_k \rightarrow x$ as k tends to ∞ , then we must show that $W^1(x_k)$ converges to $W^1(x)$ as k tends to ∞ . Let $x_k = e^{s_k V_0^{(?)}}(z_k)$ and $x = e^{s V_0^{(?)}}(z)$ for some $s_k, s \in \mathbb{R}$ and $z_k, z \in S_{x_0}$. Without loss of generality we may assume that $s = 0$, otherwise consider the sequence $y_k := e^{-s V_0^{(?)}}(x_k)$.

Recall from Section 4.2 that we may take a local neighbourhood U_x of x and a coordinate transformation ϕ . Then for k sufficiently large we have that $x_k \in U_x$, and hence $(\phi(x_k))$ converges to $(\phi(x))$. By the uniqueness of integral curves we have that

$$(\phi(e^{s V_0^{(?)}}(y))) = e^{s V_0^{(?)}}(\phi(y)).$$

Recall that $V_0^{(?)}$ denotes the representation of $V_0^{(?)}$ in the coordinates defined by ϕ . Therefore

$$(\phi(x_k)) = (\phi(e^{s_k V_0^{(?)}}(z_k))) = e^{s_k V_0^{(?)}}(\phi(z_k)).$$

Since $V_0^{(?)}$ only acts on the last coordinate we have that the first $n-1$ components of $(\phi(e^{s_k V_0^{(?)}}(z_k)))$ are equal to the first $n-1$ components of $(\phi(z_k))$. In particular, the first $n-1$ components of $(\phi(z_k))$ converge to the first $n-1$ components of $(\phi(z))$. Now z_k and z lie on the same integral submanifold of \hat{X} and hence by Lemma A.3 the last component of $(\phi(z_k))$ is equal to the last coordinate of $(\phi(z))$. Therefore $(\phi(z_k))$ converges to $(\phi(z))$, and since ϕ is a diffeomorphism we have that z_k converges to z .

Now since W^1 is continuous on S_{x_0} and using (A.3) we have

$$W^1(x_k) = W^1(z_k) \rightarrow W^1(z) = W^1(x).$$

Therefore W^1 is continuous on S_{x_0} . □

B Proofs

This appendix contains all the proofs that we omitted in the main text.

B.1 Proofs of Section 4 and Section 5

Proof of Lemma 4.1. First we show that $\text{span}(R_m)$ is contained in $\hat{\Lambda}$. By definition $\hat{\Lambda}$ contains $V_1; \dots; V_d$ and is invariant under $V_0; \dots; V_d$, hence by Note 3.11 we have that $V_{[j]} \in \hat{\Lambda}$ for all $j \in A_m$. By linearity we have that $\text{span}(R_m) \subset \hat{\Lambda}$. We show that $\hat{\Lambda}$ is contained in $\text{span}(R_m)$. It is sufficient to show that $\text{span}(R_m)$ contains $V_1; \dots; V_d$ and is invariant under $V_0; V_1; \dots; V_d$. Since $V_1; \dots; V_d \in R_m$ it suffices to show that every vector field in $\text{span}(R_m)$ is invariant under $V_0; V_1; \dots; V_d$. Every vector field V in $\text{span}(R_m)$ can be locally expressed in the form

$$V = \sum_{j \in A_m} f_j V_{[j]} \tag{B.1}$$

for some smooth functions f_j . Therefore, again by Note 3.11, it is sufficient to show that $[V; V_j] \in R_m$ for V given by (B.1) and $j \in \{0; 1; \dots; d\}$. Note that

$$[V; V_j] = \sum_{k \in A_m} f_k [V_{[k]}; V_j] = \sum_{k \in A_m} f_k [V_{[k]}; V_j] + V_j(f_k) V_{[k]}$$

Now $[V_{[k]}; V_j]$ and $V_{[k]}$ are in $\text{span}(R_m)$ and hence $\text{span}(R_m)$ is invariant under $V_0; V_1; \dots; V_d$. Therefore $\hat{\Lambda} = \text{span}(R_m)$; similarly one can show that $\hat{\Lambda}_0 = \text{span}(R_{m;0})$. \square

Proof of Proposition 4.9. Let us start by proving i). Construct \tilde{V} as described before the statement of Proposition 4.9. After the change of coordinates the vector V is expressed as

$$\tilde{V}(z) = [(J_x)^{-1} V(x)]_{x=\tilde{x}(z)} \tag{B.2}$$

As we have already observed, the last $N-n$ rows of the Jacobian matrix J are orthogonal to vectors in $\tilde{\Lambda}$, see (4.6). Since $V \in \tilde{\Lambda}$, the statement follows.

To prove ii), we first observe that by i), the vector fields $f @_{g_{j=1}^n}$ belong to $\tilde{\Lambda}$. Moreover, by Note 3.11, we have that $[W; @_{j}] \in \tilde{\Lambda}$, for all $j = 1; \dots; n$. The field $[W; @_{j}]$ can be calculated explicitly:

$$[W; @_{j}] = \sum_{i=1}^n W^i @_{j}; @_{i} = \sum_{i=1}^n \frac{\partial W^i}{\partial z} @_{i}$$

Because $[W; @_{j}] \in \tilde{\Lambda}$, one must have

$$\frac{\partial W^i}{\partial z} = 0 \text{ for all } j = 1; \dots; n; i = n+1; \dots; N$$

This concludes the proof. \square

Proof of Lemma 4.15. Since $S \subset \tilde{S}$ we have that $\overline{S} \subset \overline{\tilde{S}}$, therefore it is sufficient to show that if $x \in \tilde{S}$ then $x \in S$. Assume for a contradiction there exists some $x \in \tilde{S} \setminus S$. Since $x \in S$ there exists a neighbourhood $U \subset \tilde{S}$ which contains x and on which the coordinate transformation constructed at the beginning of Section 4.2 is well defined. Now $x \in \tilde{S} \setminus S$ implies there exists a sequence $f(x_k) \in S$ such that x_k converges to x . For k sufficiently large x_k belongs to U and, since the coordinate transformation is smooth, we have

$$x^{n+1} = \lim_{k \rightarrow \infty} x_k^{n+1} \tag{B.3}$$

having used the notation 4.2. However x_k all belong to the same maximal integral submanifold of $\hat{\Lambda}$ and hence x_k^{n+1} is constant for k large enough, by Lemma A.3. However this implies, for k large enough, that $x^{n+1} = x_k^{n+1}$; so by (B.3) and Lemma A.3 we have that x and x_k lie in the same maximal integral submanifold of $\hat{\Lambda}$. However this gives a contradiction, since $x_k \in S$ and $x \notin S$. \square

Proof of Lemma 4.16. We will prove that the set of points $K := \{x \in S_{x_0} : V_0^{(?)}(x_0) = 0\}$ is both open and closed in (the topology of) S_{x_0} , hence it has to be the whole manifold S_{x_0} – see [SA.2] and Appendix A.1 for clarifications on the manifold topology. Such a set is clearly closed (in \mathbb{R}^N and hence in the manifold topology) as it is the intersection between S_{x_0} and the preimage of 0 through a continuous function. To prove that it is also open, we will show that for any $x \in K$ there exists an open neighbourhood of x , O_x , which is contained in K . Let $x \in \mathbb{R}^N$ such that $V_0^{(?)}(x) = 0$ and let $n = n(x)$ be the rank of $\hat{\Lambda}_0$ at x ; then there exist n vectors in $\hat{\Lambda}_0(x)$ which span $\hat{\Lambda}_0(x)$ at x . Notice that, by construction, such vectors must belong to $\hat{\Lambda}(x)$, as by Lemma 4.1 $\hat{\Lambda}(x) = \hat{\Lambda}_0(x)$ if $V_0^{(?)}(x) = 0$. By the smoothness of the vector fields and because x is a regular point for both distributions, there exists a neighbourhood O_x of x such that the same n vectors span $\hat{\Lambda}_0(y)$ for every $y \in O_x$. Because $V_0^{(?)}(y)$ is orthogonal to all the vectors in $\hat{\Lambda}_0(y) (= \hat{\Lambda}(y))$, it must be the case that $V_0^{(?)}(y) = 0$ on O_x (otherwise the rank of $\hat{\Lambda}_0$ would increase, which is impossible as the rank stays constant on the orbits). Therefore $O_x \subset K$ and the proof is concluded. \square

Proof of Proposition 5.1. We emphasize that this proof heavily relies on the fact that the integral manifolds of $\hat{\Lambda}_0$ coincide with the orbits of $\hat{\Lambda}_0$, see Proposition 4.3.

Proof of i). Let S be one of the integral manifolds of $\hat{\Lambda}_0$ and suppose $x \in \mathcal{S}$. To begin with, we show that $S_x \subset S$. To this end, let y be any point in S_x . We want to show that $y \in S$. By Proposition 4.3 the integral manifold S_x is given by the orbit through x of the vector fields in $\hat{\Lambda}_0$, and hence y can be written as the end point of a curve which starts from x and is a piecewise integral curve for vector fields in $\hat{\Lambda}_0$. By considering each piece of the integral curve separately, if needed, we may assume that $y = e^{TV}(x)$ for some $T > 0$ and $V \in \hat{\Lambda}_0$. Since $x \in \overline{S}$, there is a sequence $\{x_k\}_k$ converging to x and such that $\{x_k\}_k \subset S$. Set $y_k := e^{TV}(x_k)$ and note that $\{y_k\}_k$ belongs to S since S is an orbit of $\hat{\Lambda}_0$. We have that y_k converges to y since the map $z \mapsto e^{TV}(z)$ is continuous. Therefore $y \in \overline{S}$ which implies that $S_x \subset \overline{S}$ as y is an arbitrary point in S_x . However S and S_x are both maximal integral submanifolds so they are either disjoint or they coincide; since $x \in S_x$ and $x \in S$ they must be disjoint, hence $S_x \subset \mathcal{S}$.

Proof of ii). Note that by the Stroock and Varadhan Theorem, Theorem 4.4, we have $P_x(X_t \in \overline{S_x}) = 1$ for any $x \in \mathbb{R}^N$. From the reasoning in the proof of point i), we know that if $x \in \mathcal{S}$ then $S_x \subset \mathcal{S}$, so that $\overline{S_x} \subset \mathcal{S}$. Therefore for any $x \in \mathcal{S}$ we have

$$P_x(X_t \in \mathcal{S}) = P_x(X_t \in \overline{S_x}) = 1 : \square$$

Proof of Proposition 5.3. Here we consider the case $V_0^{(?)}(x_0) \neq 0$ (which, by Lemma 4.16, implies $V_0^{(?)}(x) \neq 0$ for every $x \in S_{x_0}$). Consider the control problem (4.2) associated with the SDE (1.1). If we can show that any solution $p(t)$ of (4.2) has the property that $p(t) \in S_{e^{tV_0^{(?)}}(x_0)}$ for all t , the result then follows by the Stroock and Varadhan Support Theorem.³³ Let us now define the set

$$C := \{t \in \mathbb{R} : p(t) \in S_{e^{tV_0^{(?)}}(x_0)}\} :$$

Note that C is non-empty since $0 \in C$; if we can show that C is open and closed as a subset of \mathbb{R} then we must have that $C = \mathbb{R}$ which implies the desired result. Let us start by showing that C is open in \mathbb{R} . To this end, let t_0 be an arbitrary point $t_0 \in C$; without loss of generality we may assume that $t_0 = 0$ (otherwise we consider the path

³³Note that by Theorem 4.4 the path $\{X_t^{(x_0)}(l)g_{t \in [0;T]}\}$ is a limit, in $C([0;T]; k, k_1)$, of solutions to the control problem (4.2). Because uniform convergence implies pointwise convergence, for each fixed $t \geq 0$ the point X_t is a limit of $\{p(t) : p \text{ is a solution to (4.2)}\}$.

$q(t) := p(t + t_0)$). We will show that there exists an open neighbourhood of 0 which is contained in C . To show this fact we will make use of the (local) change of coordinates defined in Section 4.2. Let $\psi_{x_0} : U_{x_0} \rightarrow \mathbb{R}^{n+1}$ be the coordinate transformation defined on a local neighbourhood of x_0 . Take $\epsilon > 0$ sufficiently small that $p(t) \in U_{x_0}$ for all $t \in [0, \epsilon]$. It is sufficient to show that $p(t) \in S_{e^{tV_0^{(?)}}(x_0)}$ for all $t \in [0, \epsilon]$. Let $\tilde{p}(t) = \psi_{x_0}(p(t))$; consistently with the notation set in Section 2, we shall denote the first n components of $\tilde{p}(t)$ by $z(t)$, the $(n + 1)^{\text{th}}$ component by $\tau(t)$ and the last $N - (n + 1)$ components by $a(t)$. That is, $\tilde{p}(t) = (z(t); \tau(t); a(t)) = \psi_{x_0}(p(t))$; hence, in particular,

$$\tau(t) = \tau_{x_0}^{n+1}(p(t)) \tag{B.4}$$

By Lemma A.3 and (B.4), to prove that $p(t) \in S_{e^{tV_0^{(?)}}(x_0)}$ it is sufficient to show that the following holds

$$\tau(t) = \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)) \tag{B.5}$$

Differentiating the equation $\tilde{p}(t) = \psi_{x_0}(p(t))$ with respect to t and using (4.2) and (B.2), we see that \tilde{p} satisfies the equation

$$\frac{d\tilde{p}(t)}{dt} = V_0(\tilde{p}(t)) + \sum_{i=1}^d \frac{p_i}{2} X_i^d V_i(\tilde{p}(t))$$

Since $V_i \wedge V_j$ for $i = 1, \dots, d$ and $V_0 \wedge V_i$ is invariant under V_0 we have, using Proposition 4.9 (ii) and the notation (2.2)–(2.3),

$$\begin{aligned} \frac{dz(t)}{dt} &= U_0(z(t); \tau(t); a(t)) + \sum_{i=1}^d \frac{p_i}{2} X_i^d U_i(z(t); \tau(t); a(t)) \\ \frac{d\tau(t)}{dt} &= W_0(\tau(t); a(t)); \\ \frac{da(t)}{dt} &= 0; \end{aligned} \tag{B.6}$$

(The above is completely analogous to what we have done to obtain (4.7)–(4.9)). From equation (B.6) we then have

$$\tau(t) = e^{tW_0}(\tau(0)) \stackrel{(B.4)}{=} \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0))$$

In order to prove (B.5) it remains to show that $\tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)) = \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0))$. By uniqueness of the integral curves, to prove this equality we must show

$$\frac{d}{dt} \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)) = W_0(\tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)))$$

This follows from

$$\begin{aligned} \frac{d}{dt} \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)) &= \tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0)) V_0^{(?)}(e^{tV_0^{(?)}}(x_0)) \\ &= W_0(\tau_{x_0}^{n+1}(e^{tV_0^{(?)}}(x_0))); \end{aligned}$$

where the last equality is a consequence of (B.2) and of the fact that $V_0^{(?)} = (0; \dots; 0; W_0; 0; \dots; 0)$ (see comments after (4.7)–(4.9)). This proves (B.5), so C is open. Now we show that C is closed by showing that $\mathbb{R}^n \cap C$ is open.

Assume there exists t_0 such that $p(t_0) \notin S_{e^{t_0 V_0^{(?)}}(x_0)}$. Now we may take ϵ sufficiently small that $p(t) \in U_{p(t_0)}$ whenever $|t - t_0| < \epsilon$. It is sufficient to show that $p(t) \notin C$

whenever $|t - t_0| < \epsilon$. For a contradiction assume that there exists some $\bar{t} \in (t_0 - \epsilon; t_0 + \epsilon)$ with $\bar{t} \in C$ i.e. $p(\bar{t}) \in S_{e^{t_0 - \bar{t}} V_0^{(?)}}(x_0)$; then by the same argument as above, we have that

$p(t_0) \in S_{e^{(t_0 - \bar{t}) V_0^{(?)}}(p(\bar{t}))}$. Now by Lemma 3.15 applied to the points $p(\bar{t})$ and $e^{t_0 - \bar{t}} V_0^{(?)}(x_0)$ and to the vector field $V_0^{(?)}$ we have that

$$S_{e^{(t_0 - \bar{t}) V_0^{(?)}}(p(\bar{t}))} = S_{e^{(t_0 - \bar{t}) V_0^{(?)}}(e^{t_0 - \bar{t}} V_0^{(?)}(x_0))} = S_{e^{t_0 - \bar{t}} V_0^{(?)}}(x_0);$$

Therefore $p(t_0) \in S_{e^{t_0 - \bar{t}} V_0^{(?)}}$ and we have a contradiction since $t_0 \notin C$ therefore there is no $\bar{t} \in (t_0 - \epsilon; t_0 + \epsilon) \setminus C$, hence $(t_0 - \epsilon; t_0 + \epsilon) \cap C = \emptyset$. That is, C is closed in \mathbb{R} and we have $C = \mathbb{R}$ as required. \square

Proof of Proposition 5.7. For every $x \in \mathbb{R}^N$ and $t > 0$, let $g_t(x) := P_x(X_t^{(x)} \in S_x)$ and notice that $E_t = \{x \in \mathbb{R}^N : P_x(X_t^{(x)} \in S_x) > 0\} = \{x \in \mathbb{R}^N : g_t(x) > 0\}$. Suppose the SDE (1.1) admit an invariant measure, μ . Because $E = \bigcup_{t>0} E_t$, if we prove that $(E_t) = 0$ for every $t > 0$, then it follows that $(E) = 0$ (as $\int E_t g_t \mu$ is an increasing sequence of sets). So we concentrate on proving the first statement. To this end, define S^ℓ to be the union of all the maximal integral submanifolds of $V_0^{(?)}$ of dimension ℓ and notice that $\bigcup_{\ell=0}^N S^\ell = \mathbb{R}^N$; moreover, for every (arbitrary but fixed) $t > 0$, set $E_t^\ell := \{x \in S^\ell : P_x(X_t^{(x)} \in S^\ell) > 0\}$. We now proceed in two steps.

Step 1 : show that

$$E_t = \bigcup_{\ell=0}^N E_t^\ell;$$

Note that $E_t^\ell \subset E_t$; indeed, if $x \in E_t^\ell$ then $S_x \subset S^\ell$ and $P_x(X_t^{(x)} \in S_x) = P_x(X_t^{(x)} \in S^\ell) > 0$. Therefore $x \in E_t$. It remains to show that if $x \in E_t$ then there exists some ℓ such that $x \in E_t^\ell$. Fix $x \in E_t$ and let ℓ denote the dimension of S_x . By the Stroock and Varadhan Support Theorem (Theorem 4.4) we have that $P_x(X_t^{(x)} \in \overline{S_x}) = 1$ for every $t > 0$, hence

$$P_x(X_t^{(x)} \in \overline{S_x}) = P_x(X_t^{(x)} \in S_x) = g_t(x);$$

By Proposition 5.1 we have that $\overline{S_x}$ is contained in the set $\bigcup_{k \leq \ell} S^k$. In particular, we have that $\overline{S_x}$ is disjoint from S^ℓ and hence

$$g_t(x) = P_x(X_t^{(x)} \in \overline{S_x}) = P_x(X_t^{(x)} \in S^\ell);$$

Since $x \in E_t$ we have that $g_t(x) > 0$ and therefore $P_x(X_t^{(x)} \in S^\ell) > 0$, which, by definition, gives that $x \in E_t^\ell$.

Step 2 : show that $(E_t^\ell) = 0$ for all $\ell \in \{0, \dots, N\}$. To this end, set $g_t^\ell(x) := P_x(X_t^{(x)} \in S^\ell)$; then the set of $x \in S^\ell$ such that $g_t^\ell(x) > 0$ is the set E_t^ℓ . Therefore it is sufficient to show that $\int_{S^\ell} g_t^\ell(x) \mu(dx) = 0$ for all $\ell \in \{0, \dots, N\}$. Assume this is not the case; that is, assume there exists some ℓ such that

$$\int_{S^\ell} P_x(X_t^{(x)} \in S^\ell) \mu(dx) > 0; \tag{B.7}$$

We will let ℓ be the maximum index such that (B.7) holds. Since μ is an invariant measure we have that

$$(S^\ell) = \int_{\mathbb{R}^N} P_x(X_t^{(x)} \in S^\ell) \mu(dx) = \sum_{k=0}^{\ell-1} \int_{S^k} P_x(X_t^{(x)} \in S^\ell) \mu(dx); \tag{B.8}$$

Fix $k \in \{0, \dots, \ell-1\}$ and first consider the case when $k > \ell$. Since ℓ was chosen to be maximal such that (B.7) holds we must have

$$\int_{S^k} P_x(X_t^{(x)} \in S^\ell) \mu(dx) = 0 \quad \text{if } k > \ell;$$

This is equivalent to saying that the ν -measure of the set $E_t^k = \{x \in S^k : P_x(X_t^x \notin S^k)\} > 0$ is zero. Since $k \in \mathbb{N}$, we have $\{x \in S^k : P_x(X_t^x \in S^k)\} > 0$ for E_t^k , so the ν -measure of the set $\{x \in S^k : P_x(X_t^x \in S^k)\} > 0$ is zero as well. Therefore

$$\int_{S^k} P_x(X_t \in S^k) \nu(dx) = 0 \quad \text{for all } k > \ell;$$

so that

$$\sum_{k=\ell+1}^{\infty} \int_{S^k} P_x(X_t^x \in S^k) \nu(dx) = 0; \tag{B.9}$$

Now consider the case $k < \ell$. In this case we have

$$\sum_{k=0}^{\ell-1} \int_{S^k} P_x(X_t^x \in S^k) \nu(dx) = 0; \tag{B.10}$$

as by Proposition 5.1 the dimension of the manifold in which X_t evolves can only either decrease or stay the same along the paths of the SDE. Putting together (B.8), (B.9) and (B.10), one has

$$\nu(S^{\ell}) = \int_{S^{\ell}} P_x(X_t^x \in S^{\ell}) \nu(dx);$$

Writing $P_x(X_t^x \in S^{\ell})$ as $1 - P_x(X_t^x \notin S^{\ell})$ we obtain

$$\nu(S^{\ell}) = \int_{S^{\ell}} P_x(X_t \in S^{\ell}) \nu(dx) = \nu(S^{\ell}) - \int_{S^{\ell}} P_x(X_t \notin S^{\ell}) \nu(dx);$$

which gives

$$\int_{S^{\ell}} P_x(X_t \notin S^{\ell}) \nu(dx) = 0;$$

This contradicts (B.7) and hence we must have that the statement holds. \square

Proof of Lemma 5.9. Fix $x, y \in S$ and $f \in C_b(\mathbb{R}^N)$ and assume first that x, y are such that there exists a path $\gamma : [0, T] \rightarrow \mathbb{R}^N$ with $\gamma(0) = x$, $\gamma(T) = y$ and $\dot{\gamma}(t) = V_{[\cdot]}(\gamma(t))$, for some $T \in A_m$. Clearly the final time T will depend on x and y , $T = T_{x,y}$. By Lemma A.4 and by Appendix A.2 [F.3] we have

$$P_t f(y) - P_t f(x) = \int_0^{T_{x,y}} (V_{[\cdot]} P_t f)(\gamma(s)) ds;$$

Take a compact set K such that $K \subset ([0, T_{x,y}])$, then by (3.5) we have

$$|P_t f(y) - P_t f(x)| \leq \sup_{x \in K} |V_{[\cdot]} P_t f(x)| T_{x,y} \leq c_{t_0, K}^{\frac{1}{2}} e^{-(t-t_0)=2T_{x,y}} \|f\|_K;$$

Letting $t \rightarrow \infty$ we obtain the result. For any $x, y \in S$ we can take a piecewise integral curve connecting x and y , hence applying the above argument to each piece of the curve we obtain (5.4). \square

Proof of Proposition 5.10. Assume there exists an invariant measure ν with $\nu(S) = 1$; we must show that ν is the unique invariant measure such that $\nu(S) = 1$. Integrating (5.4) with respect to ν we obtain

$$\lim_{t \rightarrow \infty} P_t f(x) - \int_S P_t f(y) \nu(dy) = 0; \quad \text{for every } x \in S; \tag{B.11}$$

(Here exchanging the integral and limit is justified by the dominated convergence theorem and using Appendix A.3 [F.1]). The invariance of μ and (B.11) imply (5.5). Because (5.5) holds for every $x \in S$, by the uniqueness of the limit we have that μ must be the only invariant measure supported on S .

It remains to show that μ is ergodic. Suppose there exists $\epsilon > 0$ and a Borel set $E \subset \mathbb{R}^n$ such that $P_t 1_E = 1_E - \epsilon$ almost everywhere. Then by the semigroup property we have for every $n \in \mathbb{N}$ $P_{nt} 1_E = 1_E - n\epsilon$ almost everywhere. Now squaring and integrating (5.5) with respect to μ we have that $P_t f$ converges to $\int_S f d\mu$ in L^2 for each $f \in C_b(\mathbb{R}^n)$. Then since $C_b(\mathbb{R}^n)$ is dense in L^2 (see [54, Theorem 3.14]), for all $f \in L^2$ we have

$$\lim_{n \rightarrow \infty} \int_S P_{nt} f^2 d\mu - \left(\int_S f d\mu \right)^2 = 0:$$

By taking a subsequence if necessary we get convergence μ -almost everywhere. In particular, taking $f = 1_E$ we get $\mu(E) = 1_E(x)$ μ -almost everywhere. Hence $\mu(E) = 0$ or 1 , and μ is ergodic. \square

B.2 Proofs of Section 6

We include here the proofs of Lemma B.1 and Lemma B.4, on which the proof of Theorem 6.5 hinges.

Lemma B.1. Let Hypothesis 6.1 [H.1] and [H.3] hold and assume the semigroup $\{Q_{s,t}\}_{0 \leq s \leq t}$ admits an evolution system of measures $\{f_t g_t\}_{t \geq 0}$; then, for each $g \in C_b(\mathbb{R}^n)$, $z \in \mathbb{R}^n$, and $s \geq 0$ we have

$$\lim_{t \rightarrow \infty} \int_{\mathbb{R}^n} Q_{s,t} g(z) g(y) d\mu_t = 0: \tag{B.12}$$

Proof of Lemma B.1. Fix $g \in C_b(\mathbb{R}^n)$ and let $f \in C_b(\mathbb{R}^{n+1})$ be a function that does not depend on the last variable and such that $f(z; \cdot) = g(z)$ for every $z \in \mathbb{R}^n$; note that by (6.6) we have

$$Q_{s,t} g(z) = P_{t-s} f(z; \cdot): \tag{B.13}$$

Now for every fixed $z, y \in \mathbb{R}^n$ we can write

$$|Q_{s,t} g(z) - Q_{s,t} g(y)| = |P_{t-s} f(z; \cdot) - P_{t-s} f(y; \cdot)|$$

By Note 6.2 we have that the hyperplane $S := \{x = (z; \cdot) \in \mathbb{R}^{n+1} : \cdot = s\}$ is the orbit of the vector fields $V_{[j]}$, $j \in A_m$. Since $(z; s)$ and $(y; s)$ belong to S we may take a piecewise integral curve connecting them. Without loss of generality we may take an integral curve $\gamma: [0; T] \rightarrow \mathbb{R}^{n+1}$ connecting $(z; s)$ and $(y; s)$, with $\dot{\gamma}_t = V_{[j]}(\gamma_t)$. Clearly the time T will depend on z and y , i.e. $T = T_{z,y}$. Let K be a compact set such that $\gamma([0; T_{z,y}]) \subset K$; by Lemma A.4 applied to the function $h = P_{t-s} f$, which is in $D_V^{2;1}(\mathbb{R}^n)$ by Appendix A.2 [F.3], we have

$$|P_{t-s} f(z; s) - P_{t-s} f(y; s)| \leq \int_0^{T_{z,y}} V_{[j]}(P_{t-s} f(\gamma_u)) du:$$

Because we let $t \rightarrow \infty$, we can restrict to the case $t > s$. So $\epsilon > 0$ such that $t - s > s_0$; by (3.5) we then have

$$|P_{t-s} f(z; s) - P_{t-s} f(y; s)| \leq c_{s_0, r} e^{-(t-s-s_0)k} k_1 T_{z,y}:$$

Letting t tend to ∞ and using (B.13) we obtain

$$\lim_{t \rightarrow \infty} |Q_{s,t} g(z) - Q_{s,t} g(y)| = 0: \tag{B.14}$$

The proof can now be concluded as follows: because $\{P_t g_t\}$ is an evolution system of measures (see 6.12), we can write

$$Q_{s,t} g(z) = \int_{\mathbb{R}^n} g(y) P_{t-s}(z, dy) = \int_{\mathbb{R}^n} Q_{s,t} g(y) P_s(z, dy)$$

Using (B.14) and the dominated convergence theorem (which is applicable by Appendix A.3 [F.1]) we may take the limit as t tends to ∞ and obtain (B.12). \square

In order to prove Lemma B.4, which is the core of the last step of the proof of Theorem 6.5, we must first prove the following two results, Lemma B.2 and Lemma B.3.

Lemma B.2. Assume Hypothesis 6.1 holds. Then, for each $g \in C_b(\mathbb{R}^n)$ and $z \in \mathbb{R}^n$ we have

$$Q_{s-t,s} g(z) \rightarrow Q_t g(z)$$

as s tends to ∞ uniformly on compact subsets of $\mathbb{R}^n \times (0, 1)$. That is, for every fixed $T > 0$ and $r > 0$ we have

$$\lim_{s \rightarrow \infty} \sup_{z \in B_r} \sup_{t \in [1, T]} |Q_{s-t,s} g(z) - Q_t g(z)| = 0$$

Proof of Lemma B.2. Fix $g \in C_b(\mathbb{R}^n)$ and consider

$$v_s(z; t) = (Q_{s-t,s} g)(z); \quad z \in \mathbb{R}^n; t > 0$$

Like in the Proof of Lemma B.1, define $f \in C_b(\mathbb{R}^{n+1})$ by $f(z; s) = g(z)$ for all $z \in \mathbb{R}^n$ and $s \in \mathbb{R}$; then by (B.13) we have

$$v_s(z; t) = P_t f(z; s-t); \tag{B.15}$$

Since $(P_t f)(x)$ is a continuous function, we may take the limit as s tends to infinity to obtain

$$\lim_{s \rightarrow \infty} v_s = P_t f(z; \infty) = Q_t g(z) =: v(z; t)$$

We now wish to show that the above limit is uniform on compact subsets of $\mathbb{R}^n \times (0, 1)$; that is, we wish to show that

$$\lim_{s \rightarrow \infty} \sup_{t \in [\frac{1}{s}, T]} \sup_{z \in B_R} |v_s(z; t) - v(z; t)| = 0; \quad \text{for every fixed } R > 0; T > 0$$

To show this fact we shall use the Ascoli-Arzelà Theorem. Indeed, assuming for the moment that we can apply such a theorem, then we can find a subsequence s_k such that v_{s_k} converges uniformly on $B_R \times [1, T]$. Since v_{s_k} converges pointwise to v we have that the limit is independent of the choice of sequence hence v_s converges uniformly in $B_R \times [1, T]$ to v , as s tends to ∞ . So, if we show that the derivatives of v_s are bounded on $B_R \times [\frac{1}{s}, T]$, uniformly in s , then we may apply Arzelà-Ascoli Theorem and the proof is concluded by the above line of reasoning. By Lemma A.5 the function v_s is smooth in $(z; t) \in \mathbb{R}^n \times \mathbb{R}$ and satisfies (A.2).

For any point $(z; t) \in \mathbb{R}^n \times (0, 1)$ there exist an open neighbourhood of $(z; t)$ and smooth functions ψ_i such that, for any $i \in \{1, \dots, n\}$, the derivative ∂_i can be expressed as

$$\partial_i = \sum_{j=1}^n \psi_{ij} V_j$$

Therefore, to show that the derivatives ∂V_s are bounded on $B_R \times [0; T]$ it is sufficient to show that $V_{[\cdot]_j} V_s$ is bounded in $B_R \times [1=; T]$. This follows from the estimates recalled in Appendix A.2 [F.2]. In particular, there exists constants $C(R); \delta(R) > 0$ such that

$$\begin{aligned} \sup_{z \in B_R; t \in [1=; T]} V_{[\cdot]_j} V_s(z; t) &= \sup_{z \in B_R; t \in [1=; T]} |V_{[\cdot]_j} P_t f(z; s; t)| \\ &\leq \sup_{t \in [1=; T]} \sup_{x \in B_R} |V_{[\cdot]_j} P_t f(x)| \\ &\leq C(R) |T|^{k-k} e^{-\delta(R)T} k f k_1. \end{aligned}$$

Here we have used that P_t is convergent and hence $B_R \times [t; t + T]$ is a compact subset of \mathbb{R}^{n+1} . Similarly we may bound the second order derivatives $V_i^2 V_s$, and using (A.2) we obtain a bound for the derivative with respect to s , which is independent of s . \square

Using the tightness of the family $\{f_t g_t\}_0$, there exists a divergent sequence t_k such that t_k converges weakly to some measure μ_k as $k \rightarrow \infty$, for each $k \in \mathbb{N}$. (We emphasise that, by a diagonal argument, the sequence t_k can be chosen to be independent of k). Moreover, $\{f_{t_k} g_{t_k}\}_k$ is tight since $\{f_t g_t\}_0$ is tight (see [1, Step 2 in the proof of Theorem 6.2]).

Lemma B.3. Assume Hypothesis 6.1 holds and construct $\{f_k g_{k,2N}\}$ as above. Then,

$$\int_{\mathbb{R}^n} Q_k g(z) \mu_k(dz) = \int_{\mathbb{R}^n} g(z) \mu_0(dz);$$

for any $g \in C_b(\mathbb{R}^n)$ and every $k \in \mathbb{N}$.

Proof of Lemma B.3. We will consider the integral $\int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz)$ and show the following: for every $k \in \mathbb{N}$

$$\int_{\mathbb{R}^n} g(z) \mu_0(dz) = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz) = \int_{\mathbb{R}^n} Q_k g(z) \mu_k(dz); \tag{B.16}$$

Let us start with showing the first equality in (B.16). Because $\{f_t g_t\}_0$ is an evolution system of measures (and taking k sufficiently large that $t_k > k$), we have

$$\int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz) = \int_{\mathbb{R}^n} g(z) \mu_{t_k}(dz);$$

The above, combined with the fact that t_k converges weakly to μ_0 , gives the first identity in (B.16). To prove the second equality in (B.16), observe the following:

$$\begin{aligned} &\int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz) - \int_{\mathbb{R}^n} Q_k g(z) \mu_k(dz) \\ &= \int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz) - \int_{\mathbb{R}^n} Q_k g(z) \mu_{t_k}(dz) \\ &+ \int_{\mathbb{R}^n} Q_k g(z) \mu_{t_k}(dz) - \int_{\mathbb{R}^n} Q_k g(z) \mu_k(dz) \\ &= I_{1,k} + I_{2,k}; \end{aligned}$$

having set

$$\begin{aligned} I_{1,k} &:= \int_{\mathbb{R}^n} Q_{t_k} g(z) \mu_{t_k}(dz) - \int_{\mathbb{R}^n} Q_k g(z) \mu_{t_k}(dz); \\ I_{2,k} &:= \int_{\mathbb{R}^n} Q_k g(z) \mu_{t_k}(dz) - \int_{\mathbb{R}^n} Q_k g(z) \mu_k(dz); \end{aligned}$$

Now $I_{2,\epsilon}$ converges to 0 as $\epsilon \rightarrow 1$ since Q_{t-k} converges weakly to Q_k , by definition of Q_k . To see that $I_{1,\epsilon}$ vanishes when ϵ tends to 1, let $x \in \mathbb{R}^n$ and take a ball B_r such that $Q_{t-k}(B_r) \geq 1 - \epsilon$ for all $t > k$. This is possible since the family $f_t : t \geq 0$ is tight. By Lemma B.2 we know that $Q_{t-k;t} g(z)$ converges uniformly on compacts to $Q_k g(z)$; hence, if ϵ is sufficiently large, we have

$$\sup_{z \in B_r} |Q_{t-k;t} g(z) - Q_k g(z)| < \epsilon$$

We can therefore derive the following estimate in $I_{1,\epsilon}$:

$$\begin{aligned} I_{1,\epsilon} &= \int_{\mathbb{R}^n} |Q_{t-k;t} g(z) - Q_k g(z)| \mu_{t-k}(dz) \\ &= \int_{B_r} |Q_{t-k;t} g(z) - Q_k g(z)| \mu_{t-k}(dz) \\ &\quad + \int_{\mathbb{R}^n \setminus B_r} |Q_{t-k;t} g(z) - Q_k g(z)| \mu_{t-k}(dz) \\ &\leq \epsilon + 2kgk_1 \end{aligned}$$

As ϵ is arbitrary we have that $I_{1,\epsilon}$ converges to 0 as ϵ tends to 1, and the claim follows. \square

Lemma B.4. Assume Hypothesis 6.1 holds and, as described before the statement of Lemma B.3, let μ_0 be the weak limit of the sequence μ_t . Then $\mu_0 = \mu$.

Proof of Lemma B.4. Take $g \in C_b(\mathbb{R}^n)$. By Lemma 6.4 we know that $Q_k g(z) \rightarrow (g)$ as k tends to ∞ for each $z \in \mathbb{R}^n$ and $g \in C_b(\mathbb{R}^n)$. By an argument analogous to the one used in the proof of Lemma B.2 we have that $Q_k g(z)$ converges to (g) locally uniformly for $z \in \mathbb{R}^n$.

Now $x \in \mathbb{R}^n$; since $f_k g_k$ is a tight sequence, we may take $B_r \subset \mathbb{R}^n$ such that $Q_k(B_r) \geq 1 - \epsilon$ for all $t \geq 0$. Moreover, for k sufficiently large we have

$$\sup_{x \in B_r} |Q_k g(x) - (g)| < \epsilon$$

Then

$$\begin{aligned} \int_{\mathbb{R}^n} Q_k g(z) \mu_{t-k}(dz) - (g) &= \int_{B_r} Q_k g(z) \mu_{t-k}(dz) - (g) \\ &\quad + \int_{\mathbb{R}^n \setminus B_r} Q_k g(z) \mu_{t-k}(dz) \\ &\leq \epsilon + 2kgk_1 \end{aligned}$$

Since ϵ is arbitrary we deduce

$$\int_{\mathbb{R}^n} Q_k g(z) \mu_{t-k}(dz) \rightarrow (g); \text{ as } k \rightarrow \infty$$

However by Lemma B.3 we also have

$$\int_{\mathbb{R}^n} g(z) \mu_0(dz) = \int_{\mathbb{R}^n} Q_k g(z) \mu_{t-k}(dz); \text{ for every } k \in \mathbb{N}$$

Therefore $\mu_0 = \mu$. \square

B.3 Proofs of Section 7

Proof of Proposition 7.3. First note to prove (7.6) it is sufficient to show that for any $z \in \mathbb{R}^N$,

$$[\text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}; \mathbb{Q} + V_{0;t}] = \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_0]; \tag{B.17}$$

$$[\text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}; V_{i;t}] = \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_i]; \quad \text{for any } i \in \{1, \dots, d\}; \tag{B.18}$$

By [38, Proposition 8.30], for any three vector fields $U; V; W$ we have

$$[\text{Ad}_W U; \text{Ad}_W V] = \text{Ad}_W [U; V]; \tag{B.19}$$

Therefore setting $W = tV_0^{(\cdot)}$, $U = V_{[\cdot]}$ and $V = V_i$ for any $i \in \{1, \dots, d\}$ we have that (B.18) holds. To prove (B.17) note that

$$[\text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}; \mathbb{Q}] = \mathbb{Q} \text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]} = \text{Ad}_{tV_0^{(\cdot)}} [V_0^{(\cdot)}; V_{[\cdot]}] = \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_0^{(\cdot)}]; \tag{B.20}$$

In the above we have used [56, Lemma 4.4.2] which states that $\mathbb{Q} \text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]} = \text{Ad}_{tV_0^{(\cdot)}} [V_0^{(\cdot)}; V_{[\cdot]}]$. Using (B.19) and (B.20) we have

$$\begin{aligned} [\text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}; \mathbb{Q} + \text{Ad}_{tV_0^{(\cdot)}} V_0^{(\cdot)}] &= \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_0^{(\cdot)}] + \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_0^{(\cdot)}] \\ &= \text{Ad}_{tV_0^{(\cdot)}} [V_{[\cdot]}; V_0]; \end{aligned}$$

Therefore (B.17) holds. Hence we have that (7.6) holds.

It remains to show that for any $z \in \mathbb{R}^N$ there exists smooth and bounded functions $\tilde{V}_{[\cdot];t} : \mathbb{R}^N \times \mathbb{R} \rightarrow \mathbb{R}$ such that

$$V_{[\cdot];t}(x) = \sum_{z \in \mathbb{R}^N} \tilde{V}_{[\cdot];t}(x; z) V_{[\cdot];t}(z);$$

Since the vector fields $V_0; V_1; \dots; V_d$ satisfy the UFG condition (3.3), applying the operator $\text{Ad}_{tV_0^{(\cdot)}}$ to (3.3) we have

$$\text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}(x) = \sum_{z \in \mathbb{R}^N} \text{Ad}_{tV_0^{(\cdot)}} (\tilde{V}_{[\cdot];t}(x; z) V_{[\cdot];t}(z)) = \sum_{z \in \mathbb{R}^N} \tilde{V}_{[\cdot];t}(x; z) (e^{tV_0^{(\cdot)}}(x)) \text{Ad}_{tV_0^{(\cdot)}} V_{[\cdot]}(z);$$

Therefore if we show that the functions $\tilde{V}_{[\cdot];t}(x; z) := \tilde{V}_{[\cdot];t}(x; z) (e^{tV_0^{(\cdot)}}(x))$ are smooth, bounded and belong to the sets $C_V^1(\mathbb{R}^N \times \mathbb{R})$ then we have that the vector fields $\mathbb{Q} + V_{[0];t}; V_{[1];t}; \dots; V_{[d];t}$ satisfy the UFG condition when viewed as vector fields in both the time variable t and spatial variables z . Since $\tilde{V}_{[\cdot];t}$ is smooth and bounded, and $V_0^{(\cdot)}$ is smooth we have that $\tilde{V}_{[\cdot];t} (e^{tV_0^{(\cdot)}})$ is smooth and bounded, it remains to show that for any $k \in \{1, \dots, d\}$ and $z \in \mathbb{R}^N$ we have

$$\sup_{x \in \mathbb{R}^N; t \in \mathbb{R}} (V_{[\cdot];t}(z) \dots (V_{[k];t}(z)) (\tilde{V}_{[\cdot];t}(x; z) (e^{tV_0^{(\cdot)}}(x))) < 1$$

By [38, Proposition 8.30] for every smooth function f we have for all $z \in \mathbb{R}^N, z \in S_{x_0}, s \in \mathbb{R}$ that

$$V_{[\cdot];s}(f (e^{sV_0^{(\cdot)}}))(z) = (V_{[\cdot]} f)(e^{sV_0^{(\cdot)}}(z)); \tag{B.21}$$

Therefore

$$\begin{aligned} &\sup_{x \in \mathbb{R}^N; s \in \mathbb{R}} (V_{[\cdot];s}(z) \dots (V_{[k];s}(z)) (\tilde{V}_{[\cdot];s}(x; z) (e^{sV_0^{(\cdot)}}(x))) \\ &= \sup_{x \in \mathbb{R}^N; s \in \mathbb{R}} (V_{[\cdot]} \dots (V_{[k]} \tilde{V}_{[\cdot];s}(x; z) (e^{sV_0^{(\cdot)}}(x))) \\ &= \sup_{y \in \mathbb{R}^N} (V_{[\cdot]}(y) \dots (V_{[k]}(y) \tilde{V}_{[\cdot];s}(x; z) (e^{sV_0^{(\cdot)}}(x))) < 1; \end{aligned}$$

Therefore the UFG condition is satisfied. □

Proof of Proposition 7.4. For $f \in C^1_V(\mathbb{R}^N)$ we have that $P_t f$ is smooth (in every direction, see [F.3]) so by (7.5) we obtain

$$V_{[j];s} Q_{s;t} f(z) = V_{[j];s} P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z) \stackrel{(B.21)}{=} V_{[j]} P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z): \tag{B.22}$$

By differentiating (7.5) with respect to s we have, with a calculation analogous to the one in Lemma A.5,

$$\begin{aligned} \partial_s Q_{s;t} f(z) &= L P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z) + V_0^{(?)} P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z) \\ &\stackrel{(1.9)}{=} V_0^{(?)} P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z) + \sum_{i=1}^d V_i^2 P_{t-s}(f e^{tV_0^{(?)}})(e^{sV_0^{(?)}})(z) \\ &= V_{0;s} Q_{s;t} f(z) + \sum_{i=1}^d V_{i;s} Q_{s;t} f(z) = L_s Q_{s;t} f(z): \end{aligned}$$

Now by a density argument analogous to the one in the proof of Lemma A.5 we obtain the result for $f \in C_b(\mathbb{R}^N)$. To prove the result for $g \in C_b(\bar{S}_{x_0})$ we may apply the Tietze Extension Theorem, see [21, Chapter 2 Theorem 5.4], to extend g to a function $f \in C_b(\mathbb{R}^N)$ such that $f = g$ on \bar{S}_{x_0} (this is where we need g to be continuous up to and including the closure of S_{x_0} , as functions that are continuous on open sets do not necessarily admit a continuous extension to the whole \mathbb{R}^N , i.e. Tietze Extension Theorem would not apply). Since Z_t takes values in S_{x_0} for every $t \geq 0$ we have that $Q_{s;t} g(z) = Q_{s;t} f(z)$ for any $z \in \bar{S}_{x_0}$, hence the claim follows. \square

Proof of Proposition 7.7. By Hypothesis 7.5 [A.3], the family of measures $\{p_t^x g_t\}_{t \geq 0}$ is tight and hence, by Prokhorov's theorem, there exist a measure μ^{-S} and a diverging sequence $t_k \rightarrow \infty$ such that $p_{t_k}^x$ converges weakly to μ^{-S} as $t_k \rightarrow \infty$. Note that in general the sequence t_k and the measure μ^{-S} may depend on the choice of $x \in S_{x_0}$; however, by Lemma 5.9, $p_{t_k}^x(\cdot) = (P_{t_k} 1_{f_g})(x)$ converges weakly to μ^{-S} for any choice of $x \in S$. We now show that such a convergence is also independent of the choice of divergent sequence. Let s_k be a sequence such that $s_k \rightarrow \infty$ and $x \in f \in C_b(\mathbb{R}^N)$ and $x \in S$; then

$$P_{t_k} f(x) - P_{s_k} f(x) = \int_{s_k}^{t_k} \partial_t P_t f(x) dt = \int_{s_k}^{t_k} L P_t f(x) dt:$$

By (3.5) (and (3.7)) there exists a constant $C = C(t_0; x) > 0$ such that for all $t > t_0$ we have

$$V_0^{(?)} P_t f(x) \leq C(t_0; x) k f k_1 e^{-t}; \quad V_i^2 P_t f(x) \leq C(t_0; x) k f k_1 e^{-t}:$$

Using that $V_0^{(?)} = 0$ we have that there exists a constant $C = C(t_0; x) > 0$ such that

$$| \int_{s_k}^{t_k} L P_t f(x) dt | \leq C(t_0; x) k f k_1 e^{-s_k}; \quad \text{for all } t > t_0:$$

Therefore

$$| P_{t_k} f(x) - P_{s_k} f(x) | \leq C(x; t_0) k f k_1 \int_{s_k}^{t_k} e^{-t} dt = \frac{C(x; t_0) k f k_1}{e^{-s_k} - e^{-t_k}} e^{-s_k}:$$

Letting k tend to ∞ we have that $| P_{t_k} f(x) - P_{s_k} f(x) |$ vanishes in the limit and hence $P_{s_k} f(x)$ converges to $\mu^{-S}(f)$. Therefore $P_t f(x)$ converges to $\mu^{-S}(f)$ as t tends to ∞ .

To show that μ^{-S} is an invariant measure take an arbitrary $s > 0$ and $f \in C_b(\mathbb{R}^N)$; then for every $s \geq 0$

$$\mu^{-S}(P_s(f)) = \lim_{t \downarrow 1} P_t P_s f(x) = \lim_{t \downarrow 1} P_{t+s} f(x) = \lim_{t \downarrow 1} P_t f(x) = \mu^{-S}(f);$$

Hence μ^{-S} is an invariant measure. To show that the convergence is uniform on compact subsets of S we apply Arzela-Ascoli. Indeed x a compact set $K \subset S$ then it is sufficient to show that $P_t f(x)$ has bounded derivatives uniformly in t on K . However $x \notin \partial S$ $P_t f(x)$ is differentiable in the directions $V_{[j]}$ for all $j \in A$ which span the tangent space of S and, by the Obtuse Angle Condition, Assumption [A.5], we have for all $t > t_0$ that (3.5) holds. Hence we have that $P_t f(x)$ converges to $\mu^{-S}(f)$ uniformly on compact subsets of S . Note that since (7.9) holds for all $f \in C_b(\mathbb{R}^N)$, there is at most one measure satisfying (7.9). \square

We now move on to prove Lemma B.7 and Lemma B.8, which are the backbone of the proof of Theorem 7.9). Throughout this section, for any $f \in C_b(\bar{S}_x)$, we let

$$f^\wedge = f \circ W^{-1}; \tag{B.23}$$

In order to prove Lemma B.7 we first state and prove the following two results.

Lemma B.5. Let Hypothesis 7.5 [A.1], [A.2], [A.7] and [A.6] hold. For any fixed $f \in C_b(\bar{S}_x)$, define f^\wedge as in (B.23). Then for any compact $K \subset S_{x_0}$ and $T > 0$ we have

$$Q_{t-s;t} f^\wedge(z) \rightarrow P_{s,t} f^\wedge \left(\lim_{t \downarrow 1} e^{V_0^{(?)}}(z) \right) \tag{B.24}$$

uniformly for $s \in [1-T; T]$ and $z \in K$, whenever $\lim_{t \downarrow 1} e^{V_0^{(?)}}(z)$ exists for all $z \in K$.

Proof of Lemma B.5. The proof is analogous to the proof of Lemma B.2, so we only sketch it and point out the main differences. Note that $P_t f^\wedge$ is continuous, using (7.5) we have that (B.24) holds pointwise. To obtain convergence uniform on compact subsets of $S_{x_0} \setminus (0; 1)$, we use Arzela-Ascoli and the following estimate.

Fix a compact set $K \subset S_{x_0}$ and $T > 0$ then using (B.22) and the short time estimates from [44, Corollary 3.13] (which have been recalled in [F.2]), there exists some constant $C(K; T)$ such that the following holds:

$$\begin{aligned} & \sup_{z \in K; s \in [1-T; T]} |V_{[j];t-s}(Q_{t-s;t} f^\wedge)(z)| \\ & \stackrel{(B.22)}{=} \sup_{z \in K; s \in [1-T; T]} |V_{[j];t-s}(P_s(f^\wedge e^{tV_0^{(?)}}))(e^{(t-s)V_0^{(?)}}(z))| \\ & = \sup_{s \in [1-T; T]; x \in e^{(t-s)V_0^{(?)}}(K)} |V_{[j];t-s}(P_s(f^\wedge e^{tV_0^{(?)}}))(x)| \\ & = \sup_{s \in [1-T; T]; x \in K^0} |V_{[j];t-s}(P_s(f^\wedge e^{tV_0^{(?)}}))(x)| \\ & \leq C(K^0; T) k f^\wedge k_1; \end{aligned}$$

here K^0 is defined as $K^0 = \bigcup_{t \in [1-T; T]} e^{tV_0^{(?)}}(K)$. The set K^0 is compact under our assumptions, as for each x the diffeomorphism $x \mapsto e^{tV_0^{(?)}}(x)$ is a continuous function, the curve $e^{tV_0^{(?)}}(x)$ is convergent and the map W^{-1} is assumed continuous. \square

Define μ_k to be the probability measure such that μ_{t-k} converges weakly to μ_k , this measure is constructed analogously to the comment above Lemma B.3.

Lemma B.6. Let Hypothesis 7.5 [A.1] and [A.7] hold, and assume the semigroup $\{Q_{s;t}g_s\}_t$ admits a tight evolution system of measures $\{f_{t;k}\}_t$ supported on \bar{S}_{x_0} . Then

$$\int_{\bar{S}_x} P_k f^\wedge(x) (\cdot_k (W^1)^{-1})(dx) = \int_{\bar{S}_x} f(x) (\cdot_0 (W^1)^{-1})(dx);$$

for any $f \in C_b(\bar{S}_x)$ and f^\wedge defined as in (B.23).

Proof of Lemma B.6. This proof is completely analogous to the proof of Lemma B.3, so we only point out the main differences. It suffices to prove the following two expressions

$$\int_{\bar{S}_{x_0}} f^\wedge(z) \cdot_0(dz) = \lim_{\ell \uparrow 1} \int_{\bar{S}_{x_0}} Q_{t_\ell; k; t_\ell} f^\wedge(z) \cdot_{t_\ell; k}(dz) = \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_k(dz); \quad (B.25)$$

compare to (B.16) for comparison. Let us start with the first equality in (B.25). Since $\{f_{t;k}\}_t$ is an evolution system of measures we have

$$\int_{\bar{S}_{x_0}} Q_{t_\ell; k; t_\ell} f^\wedge(z) \cdot_{t_\ell; k}(dz) = \int_{\bar{S}_{x_0}} f^\wedge(z) \cdot_{t_\ell}(dz);$$

Since \cdot_{t_ℓ} converges weakly to \cdot_0 and W^1 is a continuous map from \bar{S}_{x_0} to \mathbb{R}^N , by the continuous mapping theorem we have that $\cdot_{t_\ell} (W^1)^{-1}$ converges weakly to $\cdot_0 (W^1)^{-1}$ and hence we obtain (B.25). To prove the second equality in (B.25) like in the proof of Lemma B.3 we write

$$\begin{aligned} & \int_{\bar{S}_{x_0}} Q_{t_\ell; k; t_\ell} f^\wedge(z) \cdot_{t_\ell; k}(dz) - \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_k(dz) \\ &= \int_{\bar{S}_{x_0}} Q_{t_\ell; k; t_\ell} f^\wedge(x) P_k f^\wedge(W^1(x)) \cdot_{t_\ell; k}(dz) \\ &+ \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_{t_\ell; k}(dz) - \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_k(dz) \\ &= I_{1;\ell} + I_{2;\ell}; \end{aligned}$$

having set

$$\begin{aligned} I_{1;\ell} &:= \int_{\bar{S}_{x_0}} Q_{t_\ell; k; t_\ell} f^\wedge(z) P_k f^\wedge(W^1(z)) \cdot_{t_\ell; k}(dz) \\ I_{2;\ell} &:= \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_{t_\ell; k}(dz) - \int_{\bar{S}_{x_0}} P_k f^\wedge(W^1(z)) \cdot_k(dz); \end{aligned}$$

Observe that on the image of W^1 we have $V_0^{(?) } = 0$, by Hypothesis 7.5 [A.7], and hence $P_k f^\wedge(W^1(z)) = P_k f(W^1(z))$ therefore we can rewrite $I_{2;\ell}$ as

$$I_{2;\ell} = \int_{\bar{S}_x} P_k f(z) \cdot_{t_\ell; k} (W^1)^{-1}(dz) - \int_{\bar{S}_x} P_k f^\wedge(W^1(z)) \cdot_k (W^1)^{-1}(dz);$$

Now $I_{2;\ell}$ converges to 0 as $\ell \uparrow 1$ since $\cdot_{t_\ell; k} (W^1)^{-1}$ converges weakly to $\cdot_k (W^1)^{-1}$. The term $I_{1;\ell}$ can be studied analogously to what we have done in the proof of Lemma B.3, up to modifications in the same spirit of those made so far, so we omit the details. \square

Lemma B.7. Suppose Hypothesis 7.5 holds. Let \cdot_{-S_x} be defined as in the comment above Theorem 7.9. Then $\cdot_0 (W^1)^{-1} = \cdot_{-S_x}$.

Proof of Lemma B.7. This proof is completely analogous to the proof of Lemma B.4 so we omit the details. \square

Lemma B.8. Assume Hypothesis 7.5 holds, let x_0 be an arbitrary point in $I_0(x)$ and let $f_t; g_0$ be constructed as in the proof of Theorem 7.9. Let $f_t; g$ be a divergent sequence such that $p_t^{x_0}$ converges weakly to some probability measure \bar{x}_0 . Then $\bar{x}_0 \int_{\bar{S}_x} = \int_0 (W^1)^{-1}$.

Proof of Lemma B.8. First we note that \bar{x}_0 and $\int_0 (W^1)^{-1}$ are both supported on \bar{S}_x . Indeed for the measure $\int_0 (W^1)^{-1}$ this follows from Lemma B.7. It is sufficient to show that given a function $f \in C_b(\mathbb{R}^N)$ such that $f(x) = f(y)$ whenever $x \in \bar{S}_y$ then $\bar{x}_0(f) = \int_0 (W^1)^{-1}(x_0)$. Let f be such a function then by Proposition 5.3 we have

$$P_t f(x_0) = E_{x_0}[f(X_t)] = E_{x_0}[f(e^{t \cdot V_0^{(?)}}(x_0))] = f(e^{t \cdot V_0^{(?)}}(x_0)):$$

Now letting t tend to ∞ we have $\bar{x}_0(f) = \int_0 (W^1)^{-1}(x_0)$ and hence \bar{x}_0 must be supported on \bar{S}_x . We now show that \bar{x}_0 and $\int_0 (W^1)^{-1}$ coincide on \bar{S}_x . Take a function $f \in C_b(\mathbb{R}^N)$ and let $f^\wedge = f \circ W^1$ then it is sufficient to show that $\bar{x}_0(f^\wedge) = \int_0(f^\wedge)$. This follows from

$$\begin{aligned} \bar{x}_0(f^\wedge) &= \lim_{t \rightarrow \infty} P_t f^\wedge(x_0) = \lim_{t \rightarrow \infty} Q_{0;t} (f^\wedge \circ e^{t \cdot V_0^{(?)}})(x_0) \\ &= \lim_{t \rightarrow \infty} Q_{0;t} f^\wedge(x_0) = \int_0(f^\wedge): \end{aligned} \quad \square$$

B.4 Proofs of Section 8

Proof of Lemma 8.3. Recall that $V_i = (U_i; 0)$ for $i = 1; \dots; d$ and $V_0 = (U_0; W_0)$ where W_0 is independent of the variable Z . By (8.1) we have

$$D_r^j X_t^{n+1} = D_r^j \int_0^t \otimes_{n+1} W_0(s) D_r^j(s) ds:$$

The only solution to this differential equation is $D_r^j \int_0^t = 0$. Therefore we have $M_t^{i;n+1} = 0$ and $M_t^{n+1;j} = 0$ for any $i; j = 1; \dots; n+1$. Hence M_t has the form (8.2). The $(i; j)$ th entry of the matrix M_t is

$$M_t^{ij} = M_t^{ij} = \int_{k=1}^d \int_0^t D_s^k(X_t^i) D_s^k(X_t^j) ds = \int_{k=1}^d \int_0^t D_s^k(X_t^i) D_s^k(Z_t^j) ds:$$

Therefore M_t is the Malliavin matrix corresponding to Z_t . □

Proof of Proposition 8.4. Note that $J_t^{n+1;j} = \frac{\partial}{\partial x} \int_0^t = 0$ for any $i \in \{1; \dots; n\}$ therefore J_t has the form

$$J_t = \begin{pmatrix} J_t & a \\ 0 & b \end{pmatrix}$$

for some random real numbers $a; b$ and a random $n \times n$ invertible matrix J_t . This implies that

$$J_t^{-1} = \begin{pmatrix} J_t^{-1} & J_t^{-1} a b^{-1} \\ 0 & b^{-1} \end{pmatrix}$$

Now by Lemma 8.3 we have that

$$\begin{aligned} C_t &= J_t^{-1} M_t (J_t^{-1})^T \\ &= \begin{pmatrix} J_t^{-1} & J_t^{-1} a b^{-1} \\ 0 & b^{-1} \end{pmatrix} \begin{pmatrix} M_t & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} (J_t^{-1})^T & 0 \\ J_t^{-1} a b^{-1} & b^{-1} \end{pmatrix} \\ &= \begin{pmatrix} J_t^{-1} M_t (J_t^{-1})^T & 0 \\ 0 & 0 \end{pmatrix} : \end{aligned}$$

Let $C_t = J_t^{-1} M_t (J_t^{-1})^T$ then it remains to show that C_t is invertible.

It is sufficient to show that $\ker(C_t) = \{0\}$ almost surely, that is if there exists $v \in \mathbb{R}^n$ such that $v^T C_t v = 0$ implies $v = 0$ almost surely. Note that

$$0 = v^T C_t v = \int_0^t \sum_{k=1}^d |v^T J_t^{-1} V_k(X_s; Y_s)|^2 ds;$$

Therefore if $v \in \ker(C_t)$ then v is orthogonal to the space $K_s := \text{span}\{J_r^{-1} V_k(X_r) : 0 \leq r \leq s; k = 1, \dots, d\}$. Hence it is sufficient to show that $K_s = \mathbb{R}^n$.

Note that the family of vector spaces $\{K_s : s \geq 0\}$ is increasing and set $K_{0+} := \bigcup_{s>0} K_s$. By the Blumenthal zero-one law, see Theorem 7.17 in [29], K_{0+} is a deterministic space with probability one. Define the stopping time

$$\tau = \inf\{s > 0 : \dim K_s > \dim K_{0+}\};$$

Note that $\tau > 0$ with probability one. Let v be orthogonal to K_{0+} and non-zero, then we have $v \in K_s$ if $s < \tau$, that is,

$$v^T J_t^{-1} V_k(X_s; Y_s) = 0; \quad k = 1, \dots, d; s < \tau;$$

This follows since $K_{0+} = K_s$ for all $s > 0$ and for $s < \tau$ we have that $\dim(K_{0+}) = \dim(K_s)$.

Recall the set R_m was defined in (3.1), we shall denote by $R_k(x)$ to be the vector space spanned by the vectors of R_k evaluated at the point x .

By following the proof of [46, Theorem 2.3.2] we obtain that v is orthogonal to $R_k(x_0)$ and hence obtain that $R_k(x_0) \subset K_{0+}$ for all $k \in \mathbb{N}$. By setting $k = m$ we have that $\mathbb{R}^n \subset K_{0+} = K_s \subset \mathbb{R}^n$. Therefore $\ker(C_t) = \{0\}$ and we have that C_t is invertible. \square

References

- [1] L. Angiuli, L. Lorenzi, A. Lunardi. Hypercontractivity and asymptotic behaviour in nonautonomous Kolmogorov equations .
- [2] L. Bañas, and Z. Brzeźniak, M. Neklyudov, M. Ondreját, and A. Prohl. Ergodicity for a stochastic geodesic equation in the tangent bundle of the 2D sphere . Czechoslovak Mathematical Journal 65.3 (2015): 617–657. MR-3407597
- [3] D. Bakry and M. Emery. Diffusions hypercontractives . In: Sém. de Probab. XIX. Lecture Notes in Math., vol. 1123. Springer, Berlin, 1985. MR-0889476
- [4] D. Bakry, I. Gentil and M. Ledoux. Analysis and geometry of Markov Diffusion operators . Springer, 2014. MR-3155209
- [5] A. Bellaïche and J.J. Risler. Sub-Riemannian Geometry . Progress in Mathematics, Birkhäuser Basel, 2012. MR-1421821
- [6] L.R. Bellet. Ergodic properties of Markov processes . In: Open quantum systems II. Springer, Berlin, Heidelberg, 2006, pp. 1–39. MR-2248986
- [7] J.-M. Bismut. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions . Wahrscheinlichkeitstheorie verw Gebiete, 1981. MR-0621660
- [8] A. Bonfilioli, E. Lanconelli, and F. Uguzzoni, Stratified Lie groups and potential theory for their sub-Laplacians . Springer Monographs in Mathematics, Springer, Berlin, 2007. MR-2363343
- [9] P. Cattiaux and L. Mesnager. Hypoelliptic non-homogeneous diffusions . Probab. Theory Relat. Fields 123 (2002): 453–483. MR-1921010
- [10] D. Crisan, F. Delarue. Sharp derivative bounds for solutions of degenerate semi-linear partial differential equations . J. Funct. Anal. 263.10 (2012): 3024–3101. MR-2973334
- [11] D. Crisan and S. Ghazali. On the convergence rates of a general class of weak approximations of SDEs . In: Stochastic Differential Equations: Theory and Applications, 2007, pp. 221–248. MR-2393578

- [12] D. Crisan, K. Manolarakis, C. Nee. Cubature methods and applications . Paris-Princeton Lectures on Mathematical Finance, 2013. MR-3183925
- [13] D. Crisan and E. McMurray. Cubature on Wiener Space for McKean-Vlasov SDEs with Smooth Scalar Interaction . arXiv:1703.04177v1. MR-3910002
- [14] D. Crisan, M. Ottobre. Pointwise gradient bounds for degenerate semigroups (of UFG type) . Proc. R. Soc. A 472.2195 (2016): 20160442. MR-3592268
- [15] D. Crisan, C. Litterer, T. Lyons. Kusuoka–Stroock gradient bounds for the solution of the filtering equation . Journal of Functional Analysis 268.7 (2015): 1928–1971. MR-3315583
- [16] G. da Prato and M. Roeckner. A note on evolution systems of measures for time-dependent stochastic differential equations . In: Seminar on Stochastic Analysis, Random Fields and Applications V. Birkhäuser Basel, 2007. MR-2401953
- [17] F. Dragoni, V. Kontis, B. Zegarliński. Ergodicity of Markov semigroups with Hörmander type generators in finite dimensions . J. Pot. Anal. 37 (2011): 199–227. MR-2969300
- [18] A.B. Duncan, T. Lelièvre, and G.A. Pavliotis. Variance reduction using nonreversible langevin samplers . Journal of Statistical Physics 163.3 (2016): 457–491. MR-3483241
- [19] J.P. Eckmann and M. Hairer. Spectral properties of hypoelliptic operators . Comm. Math. Phys. 235 (2003): 233–253. MR-1969727
- [20] P. Florchinger, Malliavin calculus with time dependent coefficients and application to nonlinear filtering . Probability Theory and Related Fields 86.2 (1990): 203–223. MR-1065279
- [21] T.W. Gamelin, and Robert Everist Greene. Introduction to topology . Courier Corporation, 1983. MR-0696379
- [22] Matthias Geissert, Alessandra Lunardi. Asymptotic behavior and hypercontractivity in non-autonomous Ornstein–Uhlenbeck equations . Journal of the London Mathematical Society 79.1 (1 February 2009): 85–106. MR-2472135
- [23] M. Hairer. How hot can a heat bath get? Communications in Mathematical Physics 292.1 (2009): 131–177. MR-2540073
- [24] M. Hairer. On Malliavin's proof of Hörmander's theorem . Bulletin des Sciences Mathématiques 135.6-7 (2011): 650–666. MR-2838095
- [25] F. Hérau. Short and long time behavior of the Fokker-Planck equation in a confining potential and applications . J. Funct. Anal. 244.1 (2007): 95–118. MR-2294477
- [26] R. Hermann. On the accessibility problem in control theory . In: Internat. Sympos. Nonlinear Differential Equations and Nonlinear Mechanics, Academic Press, New York, 1963, pp. 325–332. MR-0149402
- [27] L. Hörmander. Hypoelliptic second order differential equations . Acta Math. 119 (1967): 147–171. MR-0222474
- [28] A. Isidori. Nonlinear control systems . Springer Science & Business Media, 2013. MR-0895138
- [29] I. Karatzas, and S. Shreve. Brownian motion and stochastic calculus . Vol. 113. Springer Science & Business Media, 2012. MR-3184878
- [30] H. Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms . In: École d'Été de Probabilités de Saint-Flour XII-1982. Springer, Berlin, Heidelberg, 1984, pp. 143–303. MR-0876080
- [31] J. Kuntz, M. Ottobre and A.M. Stuart. Diffusion limit for the Random Walk Metropolis algorithm out of stationarity . arXiv preprint arXiv:1405.4896 (2014). MR-4010946
- [32] M. Kunze, L. Lorenzi, and A. Lunardi. Nonautonomous Kolmogorov parabolic equations with unbounded coefficients . Transactions of the American Mathematical Society 362.1 (2010): 169–198. MR-2550148
- [33] S. Kusuoka and D.W. Stroock. Applications of the Malliavin Calculus – I . In: Stochastic Analysis (Katata/Kyoto, 1982), 1982, pp. 271–306. MR-0780762
- [34] S. Kusuoka and D.W. Stroock. Applications of the Malliavin Calculus – II . Journal of the Faculty of Science, Univ. of Tokyo 1 (1985): 1–76. MR-0783181
- [35] S. Kusuoka, D.W. Stroock. Applications of the Malliavin Calculus – III . Journal of the Faculty of Science, Univ. of Tokyo 2 (1987): 391–442. MR-0914028

- [36] S. Kusuoka. Malliavin calculus revisited . *J. Math. Sci. Univ. Tokyo* 10 (2003): 261–277. MR-1987133
- [37] S. Kusuoka. Approximation of expectations of diffusion processes based on Lie algebra and Malliavin calculus . *UTMS*, 34, 2003. MR-2079333
- [38] John M. Lee. Smooth manifolds . In: *Introduction to Smooth Manifolds*. Springer, New York, NY, 2003, pp. 1–29. MR-2954043
- [39] C. Lobry. Controlabilite des systemes non lineaires . *SIAM J. Control* 8 (1970): 573–605. MR-0271979
- [40] L. Lorenzi, M. Bertoldi. Analytical methods for Markov Semigroups . Chapman and Hall, 2006. MR-2313847
- [41] T. Lyons and N. Victoir. Cubature on Wiener space . *Proc. Royal Soc. London* 468 (2004): 169–198. MR-2052260
- [42] Y.-A. Ma, T. Chen and E. Fox. A complete recipe for stochastic-gradient MCMC . In: *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 2899–2907.
- [43] Patrick Malsom and Frank Pinski. Pinned Brownian Bridges in the Continuous-Time Limit . arXiv preprint arXiv:1704.01991 (2017).
- [44] C. Nee. Sharp gradient bounds for the diffusion semigroup . PhD Thesis, Imperial College London, 2011.
- [45] S. Ninomyia and N. Victoir. Weak approximation scheme of stochastic differential equations and applications to derivatives pricing . *Applied Mathematical Finance*, 15.2 (2008): 107–121. MR-2409419
- [46] D. Nualart. The Malliavin calculus and related topics . Vol. 1995. Berlin: Springer, 2006. MR-1344217
- [47] M. Ottobre. Asymptotic Analysis for Markovian models in non-equilibrium Statistical Mechanics, Ph.D Thesis, Imperial College London, 2012.
- [48] M. Ottobre. Markov chain Monte Carlo and irreversibility . *Reports on Math. Phys.* (2016) MR-3517569
- [49] V. Kontis, M. Ottobre, B. Zegarliński. Markov semigroups with hypocoercive-type generator in finite dimensions: ergodicity and smoothing . *Journal of Functional Analysis* (2016). MR-3475455
- [50] M. Ottobre and G. Pavliotis. Asymptotic analysis for the generalized Langevin equation . *Nonlinearity* 24.5 (2011): 1629. MR-2793823
- [51] L. Rey-Bellet. Ergodic properties of Markov processes . In: *Open Quantum Systems II*. Springer, Berlin, Heidelberg, 2006, pp. 1–39. MR-2248986
- [52] Luc Rey-Bellet and Konstantinos Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach . *Nonlinearity* 28(7) (2015): 2081–2103. MR-3366637
- [53] Luc Rey-Bellet and Konstantinos Spiliopoulos. Variance reduction for irreversible Langevin samplers and diffusion on graphs. *Electronic Communications in Probability* 20 (2015). MR-3314650
- [54] W. Rudin. Real and complex analysis . Tata McGraw-Hill Education, 1987. MR-0924157
- [55] J. Schiltz. Time depending Malliavin calculus on manifolds and application to nonlinear filtering . *Probability and Mathematical Statistics* 18.2 (1998): 319–334. MR-1671588
- [56] E. Sontag. Mathematical control theory: deterministic finite dimensional systems . Vol. 6. Springer Science & Business Media, 2013. MR-1640001
- [57] H.J. Sussmann. Orbits of families of vector fields and integrability of distributions . *Transactions of the American Mathematical Society* 180 (1973). MR-0321133
- [58] S. Taniguchi. Malliavin's stochastic calculus of variations for manifold-valued Wiener functionals and its applications . *Probability Theory and Related Fields* 65.2 (1983): 269–290. MR-0722132
- [59] C. Villani. Hypocoercivity . *Mem. Amer. Math. Soc.* 202.950 (2009). MR-2562709
- [60] S. Watanabe. Malliavin's calculus in terms of generalized Wiener functionals . *Theory and Application of Random Fields* (1983): 284–290. MR-0799952

Acknowledgments. P. Dobson was supported by the Maxwell Institute Graduate School in Analysis and its Applications (MIGSAA), a Centre for Doctoral Training funded by the UK Engineering and Physical Sciences Research Council (grant EP/L016508/01), the Scottish Funding Council, Heriot-Watt University and the University of Edinburgh. This work was partially supported by the grant 346300 for IMPAN from the Simons Foundation and the matching 2015–2019 Polish MNiSW fund.