INVARIANT DENSITIES FOR DYNAMICAL SYSTEMS WITH RANDOM SWITCHING

A Thesis
Presented to
The Academic Faculty

by

Tobias Hurth

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
August 2014

INVARIANT DENSITIES FOR DYNAMICAL SYSTEMS WITH RANDOM SWITCHING

Approved by:

Professor Yuri Bakhtin, Advisor
School of Mathematics
Georgia Institute of Technology

Professor Leonid Bunimovich
School of Mathematics
Georgia Institute of Technology

Professor Rafael de la Llave
School of Mathematics
Georgia Institute of Technology

Professor Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Professor Jonathan C. Mattingly
Department of Mathematics
Duke University

Professor Ionel Popescu
School of Mathematics
Georgia Institute of Technology

Date Approved: 2 May 2014

ACKNOWLEDGMENTS

I would like to thank my advisor Yuri Bakhtin for suggesting the topic of this thesis, for teaching me the mathematics needed to complete the thesis, and for collaborating with me and supporting me along the way. The results presented herein are the outcome of this collaboration. I would also like to thank our co-author and friend Jonathan Mattingly for enlightening discussions and helpful advice. His input can be seen in many parts of this thesis, most notably in Chapter 3, Chapter 4 and Chapter 6. In addition, I would like to thank Stéphane Le Borgne, Florent Malrieu and Pierre-André Zitt for sharing and discussing their work with us. Finally, thanks are due to my friend Siegfried Beckus, who volunteered to proofread an earlier draft of the thesis.

My research was partially funded through my advisor's grant CAREER DMS-0742424, and I am grateful to the National Science Foundation for this support.

TABLE OF CONTENTS

SUMMARY

We study invariant measures and invariant densities for dynamical systems with random switching (switching systems, in short). An early example of a switching system related to the telegrapher's equation was analyzed by Goldstein in [18], and later by Kac in [25]. The first systematic study of switching systems was undertaken by Davis in [12]. Davis coined the term piecewise deterministic Markov processes for them.

In this thesis, we study a class of switching systems with the following specifics: Given a finite collection of smooth vector fields on a finite-dimensional smooth manifold, we fix an initial vector field and a starting point on the manifold. We follow the solution trajectory to the corresponding initial-value problem for a random, exponentially distributed time until we switch to a new vector field chosen at random from the given collection. Again, we follow the trajectory induced by the new vector field for an exponential time until another switch occurs. This procedure is iterated. The resulting two-component process whose first component records the position on the manifold, and whose second component records the driving vector field at any given time, is a Markov process.

We identify sufficient conditions for its invariant measure to be unique and absolutely continuous with respect to the product of Lebesgue measure on the manifold and counting measure on an index set associated to the collection of vector fields. These conditions consist of a Hörmander-type hypoellipticity condition as well as a recurrence condition.

In the one-dimensional case, where the manifold is the real line or some subset

thereof, we examine regularity properties of the invariant densities of absolutely continuous invariant measures. In particular, we show that invariant densities are smooth away from critical points of the vector fields. At critical points, we derive the asymptotically dominant term for invariant densities under the additional assumption that the vector fields are analytic.

CHAPTER I

INTRODUCTION

This chapter gives an introduction to dynamical systems with random switching. In Section 1.1, we sketch the role these systems play in modeling various phenomena in the sciences and engineering, and explain why they are also of intrinsically mathematical interest. In Section 1.2, we give several examples of dynamical systems with random switching. Throughout the thesis, we will revisit these examples to better illustrate some of our results (and also those of others). The main terminology and notation is introduced in Section 1.3. In particular, we will describe the construction of a dynamical system with random switching in detail. Section 1.4 is devoted to the questions addressed in this thesis. We sketch our most important statements and survey some interesting results in the existing literature.

1.1 Dynamical systems with random switching

This thesis is about dynamical systems with random switching. We will often refer to these random dynamical systems by the shorter term "switching systems". The class of switching systems we study can be described in terms of a finite family of vector fields D. The vector fields are defined on a finite-dimensional smooth manifold M. We assume that at any given time, the evolution of the system is driven by one of these vector fields, and at random times the driving vector field changes to another vector field that is randomly selected from D. Systems of this nature arise naturally in applications. In physics, switching systems can be used to model the overdamped motion of a particle in a viscous fluid, subject to alternating forces (see [17]). They also have applications to biochemistry as models for molecular motors and gene regulation ([17]), to neuronal activity and to modeling Internet traffic ([6]). Switching systems with one-dimensional manifold M appear in Markovian fluid models (see [2] and [22]). Additional motivation for studying switching systems, as well as an extensive bibliography on the subject, can be found in the monograph [37]. Switching systems are also interesting from a purely mathematical point of view. They can exhibit several somewhat counterintuitive features, such as divergence of the switching system to infinity even though all involved vector fields and their averages converge to 0, see [27]. My advisor's motivation for studying switching systems was the idea that they could serve as an introduction to hypoellipticity.

Dynamical systems with random switching were introduced by Davis in [12] under the name piecewise deterministic Markov processes (PDMPs), but examples of switching systems appear in the literature much earlier, e.g. in relation to the telegrapher's equation (see [18] and [25]). A collection of results on PDMPs can be found in [13]. Recently, there has been increased activity with regard to the ergodic theory of processes with random switching, see [6] and [10].

1.2 Examples

In this section, we present several examples of switching systems.

Example 1 Let $M = \mathbb{R}$ and let D be the collection of vector fields $u_1(\eta) := -\eta$, $u_2(\eta) := 1$ and $u_3(\eta) := -1$. At any given time, the process X is either attracted to the critical point 0 or moves to the left or to the right at constant speed.

Example 2 Let $M = \mathbb{R}$ and let D be the collection of vector fields $u_1(\eta) := -\eta$ and $u_2(\eta) := 1 - \eta$. The process X is alternately attracted by 0 and 1, and is eventually confined to the interval $(0, 1)$.

Example 3 Let M be the n-dimensional torus $\mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$, and let $D = \{e_1, \ldots, e_n\}$ be the standard basis in $\mathbb{R}^n$. At any given time, the process X moves at constant speed in the direction of one of the coordinate axes.

Example 4 Let $M = \mathbb{R}^3$, and let D consist of two Lorenz vector fields u and v with different parameter values. A Lorenz vector field is a vector field of the form

$$u(x, y, z) := \begin{pmatrix} \sigma \cdot (y - x) \\ rx - y - xz \\ xy - bz \end{pmatrix},$$

where σ, r and b are physical parameters. Assume that u has Rayleigh number $r_u = 28$ and that v has a Rayleigh number r_v that is different from, but close to, 28. We assume for both vector fields that $\sigma = 10$ and that $b = \frac{8}{3}$, which is the classical parameter choice for the Lorenz system.

Example 5 Let $M = \mathbb{R}^2$, and let D consist of the vector fields $u_1 := e_1$, $u_2 := e_2$ and $u_3(\eta) := -\eta$. The process X either moves parallel to the x- or y-axis at constant speed, or is attracted to the origin at an exponential rate. Since the vector fields $-e_1$ and $-e_2$ are not included in D, the process X while eventually be confined to the open first quadrant.

Example 6 The following example is taken from [27]. Let $M = \mathbb{R}^2$, and let D consist of the two linear vector fields u_1 and u_2 given by the matrices

$$U_1 := \begin{pmatrix} -a & c \\ 0 & -a \end{pmatrix}$$

and

$$U_2 := \begin{pmatrix} -a & 0 \\ -c & -a \end{pmatrix},$$

respectively. Here, we assume that a and c are positive parameters. Notice that both matrices are defective, in the sense that their only eigenvalue $-a$ has geometric multiplicity 1 and algebraic multiplicity 2.

1.3 Definitions and notation

We consider a finite collection D of vector fields on an n-dimensional smooth manifold M. We do not assume that M is compact. We denote these vector fields by $u_i, i \in S := \{1, \ldots, k\}$. Each vector field u_i in D induces an ordinary differential equation of the form

$$\dot{x}(t) = u_i(x(t)). \tag{1}$$

We assume that (1) is uniquely solvable if equipped with an initial condition

$$x(0) = \xi \in M.$$

This is for instance the case if u_i is Lipschitz continuous. For most of our results, we need a higher degree of regularity than Lipschitz continuity, at least continuous differentiability. We also assume that each vector field u_i is forward complete, which means that the solution trajectories to (1) are well-defined for all times $t \geq 0$.

We define a stochastic process $X = (X_t)_{t \geq 0}$ on M as follows: Given an initial point $\xi \in M$ and an initial vector field $u_i \in D$, the process X follows the solution trajectory to the corresponding initial-value problem for an exponentially distributed random time τ with parameter $\lambda_i > 0$, i.e. the distribution of τ has density $\rho_\tau(t) = \lambda_i e^{-\lambda_i t}, t \geq 0$. Then, a new driving vector field u_j is selected at random from $D \setminus \{u_i\}$, and X follows the solution trajectory to the initial-value problem

$$\dot{x}(t) = u_j(x(t))$$
$$x(0) = X_\tau$$

for an exponentially distributed random time with parameter $\lambda_j > 0$. We call these random times switching times. Iterating the construction above, we obtain a continuous trajectory $(X_t)_{t \geq 0}$ on M that is defined for all positive times and driven by one of the vector fields from D between any two switches. If the vector fields in D are smooth, the trajectory is piecewise smooth.

We make the following assumptions on the switching mechanism:

(a) All switching times are exponentially distributed and independent conditioned on the sequence of driving vector fields.

(b) The parameter $\lambda_j > 0$ of the exponential time between any two switches depends only on the current state $j \in S$. In particular, it does not depend on the value of X at the given time.

(c) For any two indices i and j in S, there is a positive probability of switching from i to j.

We call the parameters $(\lambda_i)_{i\in S}$ switching rates. For $j \neq i$, let $\lambda_{i,j}$ be the rate of switching from u_i to u_j. Then,

$$\lambda_i = \sum_{j\neq i} \lambda_{i,j}.$$

In many papers on dynamical systems with random switching, the switching rates are allowed to depend on the location of the process X, and it is only required that the transition mechanism on S be irreducible (see for instance [17], [6] and [10]). It is interesting to note that even if the switching rates of a process (X, A) do not depend on X, the switching rates for the time-reversed version of (X, A) are in general X-dependent (see [17]). If we consider for instance the switching system in Example 2 with constant switching rates $\lambda_1 = \lambda_2$, we observe that if we let time run backwards, the rate of switching from $-u_1$ to $-u_2$ explodes near the critical point 1 of $-u_2$ and the rate of switching from $-u_2$ to $-u_1$ explodes near 0. This is not hard to see: Since the original process is confined to the interval $(0, 1)$, so is its time-reversed version. However, the trajectories of $-u_1$ and $-u_2$ become unbounded as t goes to infinity, so the fast switching becomes necessary to keep the trajectories inside $(0, 1)$.

We work with exponentially distributed switching times because the exponential distribution is memoryless, i.e. $P(T > s + t|T > s) = P(T > t)$ if T is an exponentially distributed random variable on a probability space $(\Omega, \mathcal{F}, P)$. This ensures the Markov property for the stochastic process that, at any time t, keeps track of the position of X on M and of the driving vector field in D. If the switching times are not exponentially distributed, we can still construct a Markov process if we also record the time elapsed since the last switch. Our results can be extended to this more general setting, but we do not carry out these straightforward extensions to avoid heavy notation.

The process X is not a Markov process: The shape of the trajectory leading up to a fixed point on M allows us to infer the current driving vector field. We can build a Markov process by adjoining a second stochastic process $A = (A_t)_{t\geq 0}$ that captures the driving vector field at any given time. More precisely, we define $A_t \in S$ as the index of the driving vector field at time t. We will also refer to this index as the regime or the state at time t. The process A is a continuous-time Markov process on the finite state space S. Under our assumptions on the switching rates (in particular the fact that the rates do not depend on X), it has a unique stationary distribution. The trajectories of A are right-continuous and piecewise constant. The two-component process (X, A) is then a Markov process with state space $M \times S$. We call X the continuous and A the discrete component of (X, A). We denote elements of the associated Markov family, i.e. the distribution on paths emitted at $(\xi, i) \in M \times S$ and generated by the iterative random procedure outlined above, by $\mathsf{P}_{\xi,i}$. The corresponding transition probability measures are denoted by $\mathsf{P}^t_{\xi,i}$, $t \geq 0$, and the Markov semigroup associated to the process (X, A) is denoted by $(\mathsf{P}^t)_{t\geq 0}$. The transition probability measures are defined on the product σ-algebra $\mathcal{B}(M) \otimes \mathcal{P}(S)$, where $\mathcal{B}(M)$ is the Borel σ-algebra on M and $\mathcal{P}(S)$ is the power set of S. We write $\mathsf{E}_{\xi,i}$ for expectation with respect to $\mathsf{P}_{\xi,i}$.

If the initial distribution of the Markov process (X, A) is μ, then the distribution of the process at time t is given by the measure $\mu \mathsf{P}^t$ on $M \times S$, defined by

$$\mu \mathsf{P}^t(E \times \{j\}) := \sum_{i \in S} \int_M \mathsf{P}^t_{\xi,i}(E \times \{j\}) \ \mu(d\xi \times \{i\}). \tag{2}$$

From here on, we will denote the projection $\mu(\cdot \times \{i\})$ by μ_i.

A probability measure μ on $M \times S$ is called invariant for $(\mathsf{P}^t)_{t \geq 0}$ if $\mu = \mu \mathsf{P}^t$ for all $t \geq 0$. For real-valued and $(\mathsf{P}^t_{\xi,i})_{(\xi,i) \in M \times S}$-integrable functions f on $M \times S$, we can also define the left action of $\mathsf{P}^t, t \geq 0$, on f by

$$\mathsf{P}^t f(\xi, i) := \sum_{j \in S} \int_M f(\eta, j) \ \mathsf{P}^t_{\xi,i}(d\eta \times \{j\}), \quad (\xi, i) \in M \times S. \tag{3}$$

A probability measure μ is then invariant for $(\mathsf{P}^t)_{t \geq 0}$ if and only if

$$\sum_{i \in S} \int_M \mathsf{P}^t f(\xi, i) \ \mu_i(d\xi) = \sum_{i \in S} \int_M f(\xi, i) \ \mu_i(\xi) \tag{4}$$

for all $t \geq 0$ and for all bounded $\mathcal{B}(M) \otimes \mathcal{P}(S)$-measurable functions f on $M \times S$. The infinitesimal generator $\mathscr{L}$ of the Markov semigroup $(\mathsf{P}^t)_{t \geq 0}$ is the linear operator defined by

$$\mathscr{L} f(\cdot) := \lim_{t \downarrow 0} \frac{1}{t} (\mathsf{P}^t f(\cdot) - f(\cdot))$$

for those functions f for which the limit exists (see for instance [16]). If f is a function on $\mathbb{R}^n \times S$ such that $f(\cdot, i)$ is a smooth function on $\mathbb{R}^n$ for every $i \in S$, we have

$$\mathscr{L} f(\xi, i) = \langle u_i(\xi), \nabla_\xi f(\xi, i) \rangle + \sum_{j \neq i} \lambda_{i,j} \cdot (f(\xi, j) - f(\xi, i)), \tag{5}$$

see [6, Formula 2]. Here, $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{R}^n$ and ∇_ξ is the gradient with respect to ξ. We will not work directly with the infinitesimal generator, but we point out that equation (5) is the starting point for deriving the Fokker–Planck equations for continuously differentiable invariant densities of the semigroup. This derivation is carried out in [17, Proposition 3.1]. We will take up the discussion of the Fokker–Planck equations in earnest in Section 6.4, and already hint at them in Section 1.4.

For $i \in S$, we denote the flow function of the vector field u_i by Φ_i. Due to forward completeness of u_i, the flow function is uniquely defined for any $t \geq 0$ and for any $\xi \in M$ by

$$\frac{d}{dt}\Phi_i^t(\xi) = u_i(\Phi_i^t(\xi)),$$
$$\Phi_i^0(\xi) = \xi.$$

We write $\mathbb{R}_+$ to denote the positive real line $(0, \infty)$. For any index vector $\mathbf{i} = (i_1, \ldots, i_m) \in S^m$ and for any corresponding vector of switching times $\mathbf{t} = (t_1, \ldots, t_m) \in \mathbb{R}_+^m$, we define

$$\Phi_{\mathbf{i}}^{\mathbf{t}}(\xi) := \Phi_{i_m}^{t_m}(\Phi_{i_{m-1}}^{t_{m-1}}(\ldots\Phi_{i_1}^{t_1}(\xi))\ldots)$$

as the cumulative flow along the trajectories of $u_{i_1}, \ldots, u_{i_m}$ with starting point $\xi \in M$. Through much of the thesis, we will restrict ourselves to positive switching times, but we will need to admit flows backwards in time in Chapters 4 and 6. In these instances, we will extend the definition of the cumulative flow to sequences of switching times in $\mathbb{R}^m$, provided that each of the flow functions is defined for the respective negative time.

1.4 Questions addressed in this thesis and prior work

The questions addressed in this thesis concern the ergodic theory of dynamical systems with random switching. Our main objects of study will be invariant measures of the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$ associated to the process (X, A). Many long-term asymptotic properties of dynamical systems and random dynamical systems can be described in terms of invariant measures. The existence of invariant measures can often be derived by constructing a Lyapunov function and by subsequently establishing recurrence properties or tightness for the process (see for instance [37, Sections 3.3–3.4]). On a compact state space, existence of an invariant measure is often shown using the Krylov–Bogoliubov method, see Section 3.2.

Uniqueness and absolute continuity of invariant measures are often related to each other and more subtle. In Chapter 2, we establish sufficient conditions for uniqueness and absolute continuity of the invariant measure of the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$. These consist in easily verifiable properties of the vector fields in D. For dynamical systems with random switching, a major obstacle to uniqueness and absolute continuity of the invariant measure is the fact that the only source of randomness is the sequence of driving vector fields, i.e. the process (X, A) evolves deterministically most of the time. Our conditions for uniqueness and absolute continuity are formulated in terms of Lie algebras associated to the driving vector fields. They are analogues of the classical Hörmander condition guaranteeing absolute continuity of transition densities of hypoelliptic diffusions, and it is thus natural to refer to them as hypoellipticity conditions. In the diffusion context, absolute continuity of transition densities is usually derived from the variational analysis of diffusion paths known as Malliavin calculus, see for instance [4, Chapter VIII], [5] and [30].

In 2.1, we will formulate a weak and a strong hypoellipticity condition. For the weak hypoellipticity condition, we assume that the tangent space at some point $\xi \in M$ is generated by the smallest Lie algebra of smooth vector fields on M that contains all vector fields in D. We shall denote this Lie algebra by $\mathcal{I}(D)$. For the strong hypoellipticity condition, we assume that the tangent space at some $\xi \in M$ is generated by the derived algebra associated to $\mathcal{I}(D)$. Since the derived algebra is a subalgebra of $\mathcal{I}(D)$, the strong hypoellipticity condition is indeed stronger than the weak one. If the weak hypoellipticity condition holds at a point $\xi \in M$ that can be approached from any initial point using the given vector fields as admissible controls, then there exists at most one invariant measure, and this measure is absolutely continuous. This is Theorem 2 in Section 2.2. It was derived independently by Benaïm, Le Borgne, Malrieu and Zitt in [6].

The central part in establishing absolute continuity and uniqueness is the analysis of transition probabilities of switching systems. Under the strong hypoellipticity condition, we prove that all transition probabilities for the system have nontrivial absolutely continuous components. The weak hypoellipticity condition allows to prove the existence of absolutely continuous components not for the transition probabilities themselves, but for their time averages. The extraction of these absolutely continuous components is based on classical control theory results that can be found in Chapter 3 of [24]. These control theory results rely on earlier work by Chow [9], Sussmann and Jurdjevic [32], and Krener [26]. Our conditions and the structure of our proofs match those of [24], where the nondegeneracy of certain maps is exploited to establish accessibility of an open set of points, either at a fixed time t (under the strong hypoellipticity condition) or for $t \geq 0$ (under the weak hypoellipticity condition). We use the same nondegeneracy to prove absolute continuity, and one can interpret our result as filling the control theory with probabilistic content.

Existence of and convergence to an invariant measure are questions of general interest and complement our work on uniqueness and absolute continuity. While we did not study these questions, we give an overview of some interesting and important results on ergodicity for switching systems in Chapter 3. Most of the results we present were developed by Benaïm, Le Borgne, Malrieu and Zitt in [6] and [7]; and by Cloez and Hairer in [10]. We do not claim that our survey is comprehensive.

If an invariant measure of the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$ is absolutely continuous, it has a probability density function according to the Radon–Nikodym Theorem (see for instance [15]). We call the density of an invariant measure an invariant density. In Chapter 4, we study the regularity theory for invariant densities of switching systems with one-dimensional continuous component. In particular, we assume that the manifold M is the real line. We show that smoothness of the vector fields in D translates into smoothness of invariant densities away from critical points of the vector fields (Theorem 11).

In the literature, regularity properties of invariant densities are often assumed in order to derive other features of the densities. For instance, it is shown in [17, Proposition 3.1] that if invariant densities are $\mathscr{C}^1$ on a set Ω, they satisfy the Fokker–Planck equations associated to the switching system in the interior of Ω. From this differential characterization, the authors deduce time-reversibility of stationary piecewise deterministic Markov processes and derive explicit formulas for the invariant densities of certain switching systems that they call exactly solvable. A result similar to [17, Proposition 3.1] can be found in [22, Theorem 1]. Our Theorem 11 in Section 4.1 gives sufficient conditions for continuity and differentiability of invariant densities that are stated in terms of the vector fields, and are easily verifiable. In particular, we show that if none of the vector fields vanish at a point $\xi \in \mathbb{R}$ and if all vector fields are $\mathscr{C}^{n+1}$ in a neighborhood of ξ, then the invariant densities are $\mathscr{C}^n$ at ξ.

In Chapter 5, we give a detailed description of the support of invariant measures for switching systems whose continuous component X lives on $\mathbb{R}$. While this description is interesting in its own right, it also serves as a tool to analyze how invariant densities behave at critical points of the vector fields. This analysis is carried out in Chapter 6. In the case of two vector fields on a bounded interval that point in opposite directions (such as Example 2), [17, Proposition 3.12] gives an explicit formula for the invariant densities. From this formula, one obtains the exact asymptotic behavior of the densities close to critical points. However, computing invariant densities explicitly is in general very difficult ([17, Section 3.3]). Finding necessary and sufficient conditions for boundedness of invariant densities is already challenging. In the one-dimensional case, invariant densities are bounded away from critical points (Lemma 13), but we expect to find switching systems with two-dimensional continuous component whose invariant densities become unbounded along curves that do not contain any critical points. For an appropriate choice of switching rates, this phenomenon should occur in Example 5, where we expect the invariant densities to become unbounded along the coordinate axes. If the continuous component is one-dimensional, [2, Theorem 1] provides sufficient conditions for boundedness of an invariant density close to a critical point of its associated vector field. For vector fields that behave linearly close to a critical point, we give necessary conditions and sufficient conditions for boundedness in terms of the vector fields and the switching rates (Corollary 4). These conditions recover part of the results in [2]. For analytic vector fields, we also derive the asymptotically dominant term of an invariant density as its argument approaches a critical point of the corresponding vector field (Theorem 13). Even if the vector fields in D are not analytic, we can derive some asymptotics at critical points, but the results are not as sharp as in the analytic case, see Theorem 12. The basic tools in our investigation of invariant densities (both in Chapter 4 and in Chapter 6) are two integral equations satisfied by invariant densities. These equations are closely related to the Kolmogorov forward equations (see Appendix B), but do not require differentiablity of the densities. When deriving the asymptotically dominant terms in the case of analytic vector fields, we use the theory of regular singular points for systems of linear ordinary differential equations. We follow [33, Section 3.11].

We now highlight some important questions that are not covered in this thesis. It is natural to ask how a switching process behaves if the switching rates diverge to $+\infty$ (i.e. in the limiting regime of very fast switching), and how it behaves if the switching rates converge to 0 (i.e. in the limiting regime of very slow switching). For the simple switching system in Example 2, these questions were studied in [23]. If we assume that switches from u_1 to u_2 occur with the same frequency λ as switches from u_2 to u_1, we find that the invariant densities of the switching process have the form

$$\rho_1(\xi) = c(\lambda) \cdot \xi^{\lambda-1} \cdot (1-\xi)^{\lambda} \tag{6}$$

and

$$\rho_2(\xi) = c(\lambda) \cdot (1-\xi)^{\lambda-1} \cdot \xi^{\lambda}, \tag{7}$$

where $c(\lambda)$ is a normalizing constant. Formulas (6) and (7) can be derived by solving the Fokker–Planck equations associated to the switching system. This is only possible because one can reduce the Fokker–Planck equations for Example 2 to a single first-order linear ODE — for switching systems with more than two vector fields, one obtains a linear system with nonconstant coefficients that may have singularities at critical points, see Section 6.4. In [17, Proposition 3.12], similar formulas were derived independently for switching between two real-valued vector fields pointing in opposite directions. Exploiting the explicit representations in (6) and (7), we established the following limit theorems in the spirit of laws of large numbers for fast and slow switching (see [23]).

Theorem 1 *For $\lambda > 0$, let $\mu^{(\lambda)}$ denote the invariant measure for the switching system in Example 2 with switching rates $\lambda_1 = \lambda_2 = \lambda$.*

(i) As λ goes to 0, $\mu_1^{(\lambda)}$ converges weakly to the Dirac measure δ_0 and $\mu_2^{(\lambda)}$ converges weakly to the Dirac measure δ_1.

(ii) As λ goes to $+\infty$, $\mu_1^{(\lambda)}$ and $\mu_2^{(\lambda)}$ converge weakly to the Dirac measure $\delta_{\frac{1}{2}}$.

For fast switching, Proposition 3.6, part (i), in [17] is comparable to part (ii) of Theorem 1, but is stated for a broad class of switching systems, not just for one particular system. To give an idea of Proposition 3.6, we define the family of processes $(\chi_t)_{t\geq 0}$ on the set of functions from S to $\{0,1\}$ by

$$\chi_t(i) := \mathbb{1}_{\{i\}}(A_t), \quad i \in S, t \geq 0.$$

Loosely writing, the proposition then states that as λ goes to $+\infty$, the process (X, χ), conditioned on starting at $(\xi, \mathbb{1}_{\{i\}})$, approaches a limiting process (X^*, χ^*), where X^* is the solution to the entirely deterministic initial-value problem

$$\begin{aligned} \dot{X}_t^* &= \sum_{i \in S} \nu(\{i\}) \cdot u_i(X_t^*), \\ X_0^* &= \xi \end{aligned} \tag{8}$$

and where

$$\chi_t^*(j) := \nu(\{j\}), \quad j \in S, t \geq 0.$$

The measure ν on S is the stationary distribution of the Markov process A. The vector field in (8) is called the mean vector field in [17] and can be thought of as the natural average of the vector fields in D under the dynamics of A. In Example 2 with equal switching rates, the stationary distribution of A assigns probability $\frac{1}{2}$ to both states 1 and 2. The mean vector field is then

$$\bar{u}(\xi) = \frac{1}{2}u_1(\xi) + \frac{1}{2}u_2(\xi) = \frac{1}{2} - \xi.$$

Clearly, $\bar{u}$ has a critical point at $\xi = \frac{1}{2}$ that is globally attracting. Compare this result to weak convergence to $\delta_{\frac{1}{2}}$ in part (ii) of Theorem 1.

In [23], we also established a large-deviation result for the switching system in Example 2. For the entropy function

$$I(\xi) := -\ln(4\xi \cdot (1-\xi)),$$

we showed that

$$\liminf_{\lambda\to\infty} \frac{1}{\lambda} \cdot \ln(\mu_1^{(\lambda)}(G)) \geq -\inf_{\xi\in G} I(\xi)$$

for any nonempty open set $G \subset (0,1)$ and that

$$\limsup_{\lambda\to\infty} \frac{1}{\lambda} \cdot \ln(\mu_1^{(\lambda)}(F)) \leq -\inf_{\xi\in F} I(\xi)$$

for any closed set $F \subset (0,1)$. A variety of large-deviation results for much broader classes of switching systems can be found in [17].

CHAPTER II

UNIQUENESS AND ABSOLUTE CONTINUITY

In this chapter, we identify conditions on the vector fields in D that guarantee uniqueness and absolute continuity of the invariant measure associated to the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$. The chapter is based on [1] and is organized as follows: In Section 2.1, we introduce the main notions from differential geometry and geometric control theory needed to formulate sufficient conditions for uniqueness and absolute continuity of the invariant measure of $(\mathsf{P}^t)_{t\geq 0}$. In Section 2.2, we state the main result on uniqueness and absolute continuity of the invariant measure (Theorem 2), as well as two auxiliary results on regularity of transition probabilities each based on one of the Hörmander-type assumptions. We prove these regularity results in Section 2.3. Section 2.4 contains the proof of Theorem 2. In Section 2.5, we apply Theorem 2 to Examples 3 and 4. Throughout Chapter 2, we assume that the vector fields in D are $\mathscr{C}^\infty$. We also assume that M is an n-dimensional $\mathscr{C}^\infty$-manifold, where n can be any positive integer.

2.1 Hypoellipticity

Let $\mathcal{V}(M)$ denote the set of real smooth vector fields on the manifold M, and let $C^\infty(M)$ denote the set of real-valued smooth functions on M. As explained above, we assume that D is contained in $\mathcal{V}(M)$. Any element of $\mathcal{V}(M)$ corresponds uniquely to a derivation on $C^\infty(M)$, that is to a linear operator δ on $C^\infty(M)$ satisfying the Leibniz rule

$$\delta(f \cdot g) = \delta(f) \cdot g + f \cdot \delta(g).$$

The Lie bracket of two vector fields u and v in $\mathcal{V}(M)$ is defined as the vector field

$$[u,v](f) := u(v(f)) - v(u(f))$$

for test functions f in $C^\infty(M)$. Alternatively, using the symbol Φ_u for the flow function associated to vector field u, we can define the Lie bracket $[u,v]$ as the vector field

$$[u,v](\xi) := \frac{1}{2}\frac{d^2}{dt^2}\Phi_v^{-t}(\Phi_u^{-t}(\Phi_v^t(\Phi_u^t(\xi))))|_{t=0}. \tag{9}$$

Formula (9) can be interpreted as follows: Given two vector fields u and v, we obtain the value of the Lie bracket $[u,v]$ at a point $\xi \in M$ by making appropriately scaled infinitesimal switches between u and v. The set $\mathcal{V}(M)$ equipped with the bilinear operation $[.,.]$ becomes a Lie algebra over the reals. That means $\mathcal{V}(M)$ is a real vector space endowed with the bilinear and alternating operation $[.,.]$ that satisfies the Jacobi identity

$$[u,[v,w]] + [w,[u,v]] + [v,[w,u]] = 0, \quad u,v,w \in \mathcal{V}(M).$$

A subset of $\mathcal{V}(M)$ is called involutive if it is closed under taking Lie brackets of its elements. An involutive subspace of $\mathcal{V}(M)$ is called a subalgebra of $\mathcal{V}(M)$. We denote the smallest subalgebra of $\mathcal{V}(M)$ that contains D by $\mathcal{I}(D)$. The derived algebra $\mathcal{I}'(D)$ is the smallest algebra containing Lie brackets of vector fields in $\mathcal{I}(D)$. We have $\mathcal{I}'(D) \subset \mathcal{I}(D)$, but $\mathcal{I}'(D)$ might not contain any elements of D and may therefore be strictly contained in $\mathcal{I}(D)$. In addition, we define $\mathcal{I}_0(D)$ as the set of vector fields of the form

$$v + \sum_{i=1}^{k} \nu_i u_i,$$

where $v \in \mathcal{I}'(D)$, $u_1, \ldots, u_k \in D$ and $\sum_{i=1}^k \nu_i = 0$. Finally, we set

$$\mathcal{I}(D)(\xi) := \{u(\xi) : \; u \in \mathcal{I}(D)\}$$

and

$$\mathcal{I}_0(D)(\xi) := \{u(\xi) : \; u \in \mathcal{I}_0(D)\}$$

for any $\xi \in M$. The sets $\mathcal{I}(D)(\xi)$ and $\mathcal{I}_0(D)(\xi)$ are finite-dimensional vector spaces.

Our results on uniqueness and absolute continuity of the invariant measure are based on the following assumptions that can naturally be called hypoellipticity conditions in analogy with Hörmander's theory. We say that a point $\xi \in M$ satisfies the strong hypoellipticity condition if $\dim \mathcal{I}_0(D)(\xi) = n$. We say that a point $\xi \in M$ satisfies the weak hypoellipticity condition if $\dim \mathcal{I}(D)(\xi) = n$. The set of points satisfying the strong hypoellipticity condition is open, and so is the set of points satisfying the weak hypoellipticity condition.

For our absolute continuity results we will need a reference measure on M that will play the role of Lebesgue measure. As a smooth manifold, M can be endowed with a Riemannian metric. The metric tensor can be used to define measures on coordinate patches of M. One can then use a partition of unity (see for instance [34, Section 7]) to construct a Borel measure on M whose pushforward to $\mathbb{R}^n$ under any chart map is equivalent to Lebesgue measure. We call the uniquely defined measure on M obtained through this construction Lebesgue measure, denote it by λ^M, and use it as the main reference measure, often omitting "with respect to Lebesgue measure" when writing about absolute continuity. The product of the Lebesgue measure on M and counting measure on S will be called the Lebesgue measure on $M \times S$. We denote the Lebesgue measure on $\mathbb{R}^m$ by λ^m.

It remains to introduce the notions of reachability and accessibility. Recall our definition of the flow function Φ_i associated to the vector field u_i in Section 1.3. Also recall how we defined the cumulative flow along the trajectories of vector fields $u_{i_1}, \ldots, u_{i_m}$ with a given starting point on M. In this chapter, we will only work with positive switching times, both for single and cumulative flows. We call a point $\eta \in M$ D-reachable from a point $\xi \in M$ if there exist an index vector $\mathbf{i}$ and a corresponding vector of positive switching times $\mathbf{t}$ such that

$$\eta = \Phi^{\mathbf{t}}_{\mathbf{i}}(\xi).$$

If the components of $\mathbf{t}$ sum up to t, we say that η is D-reachable from ξ at time t. For $\xi \in M$ and $t > 0$, let $L_t(\xi)$ denote the set of D-reachable points from ξ at time t, and let $L(\xi) := \bigcup_{t>0} L_t(\xi)$ denote the set of D-reachable points from ξ. We call the points in the closure $\overline{L(\xi)}$ D-accessible from ξ. Let $L := \bigcap_{\xi \in M} \overline{L(\xi)}$ denote the set of points that are D-accessible from all other points in M. In [1], we used the term D-appraochable instead of D-accessible. We make this change because the term D-accessible is already established in the literature (see for instance [6, Remark 3.8]). Notice that if one of the vector fields in D has a minimal global attractor, then this attractor is a subset of L.

2.2 Sufficient conditions for uniqueness and absolute continuity of the invariant measure

We are ready to state sufficient conditions for uniqueness and absolute continuity of the invariant measure.

Theorem 2 *Suppose the weak hypoellipticity condition is satisfied at a point $\xi \in L$. If $(\mathsf{P}^t)_{t\geq 0}$ has an invariant measure, then the invariant measure is unique and absolutely continuous with respect to the product of Lebesgue measure on M and counting measure on S.*

The main task in the proof of Theorem 2 is to establish regularity for transition probabilities under the weak hypoellipticity condition. Under the weak hypoellipticity condition, it may happen that none of the transition probability measures $(\mathsf{P}^t_{\xi,i})_{t\geq 0}$ has a nonzero absolutely continuous component. We refer the reader to the discussion of Example 3 in Section 2.5 to illustrate this point. Nevertheless, the weak hypoellipticity condition guarantees that time averages of transition probabilities have nontrivial absolutely continuous components. Specifically, we will establish this for the resolvent probability kernel $\mathsf{Q}_{\xi,i}$ defined by

$$\mathsf{Q}_{\xi,i}(E \times \{j\}) := \int_{\mathbb{R}_+} e^{-t} \cdot \mathsf{P}^t_{\xi,i}(E \times \{j\})\ dt. \tag{10}$$

Theorem 3 *If the weak hypoellipticity condition is satisfied at some point $\xi \in M$, then for any $i \in S$, the measure $\mathsf{Q}_{\xi,i}$ defined by* (10) *has a nonzero absolutely continuous component with respect to Lebesgue measure on $M \times S$.*

Resolvent kernels are useful in the study of invariant distributions due to the following straightforward result.

Lemma 1 *If a measure μ is invariant with respect to the semigroup $(\mathsf{P}^t)_{t\geq 0}$, it is also invariant with respect to Q, i.e. $\mu = \mu\mathsf{Q}$, where the convolution $\mu\mathsf{Q}$ is defined in analogy to* (2) *by*

$$\mu\mathsf{Q}(E \times \{j\}) := \sum_{i\in S} \int_M \mathsf{Q}_{\xi,i}(E \times \{j\})\ \mu_i(d\xi).$$

Under the strong hypoellipticity condition, we can establish a much stronger regularity property of the transition probabilities.

Theorem 4 *If the strong hypoellipticity condition is satisfied at a point $\xi \in M$, then for any $i \in S$ and any $t > 0$, the transition kernel $\mathsf{P}^t_{\xi,i}$ has a nonzero absolutely continuous component with respect to Lebesgue measure on $M \times S$.*

2.3 Proof of Theorems 4 and 3

Our proofs of Theorems 4 and 3 use classical results from geometric control theory that can be found in [24]. The statements we present below are derived from Theorems 3.1, 3.2, and 3.3 in [24]. Analogous results for the special case of analytic vector fields on a real analytic manifold are first stated in [32, Theorems 3.1 and 3.2]. In their paper, Sussmann and Jurdjevic were able to build on prior work by Chow (see [9]) who considered symmetric families of analytic vector fields. Krener generalized these results to C^∞-vector fields in [26].

Recall that a regular point of a function $f : \mathbb{R}^m \to M$ is a point $\mathbf{t} \in \mathbb{R}^m$ such that the differential $Df(\mathbf{t})$ has full rank. If $Df(\mathbf{t})$ has deficient rank, $\mathbf{t}$ is called a critical point of f.

Theorem 5 *Assume that the strong hypoellipticity condition holds at some $\xi \in M$. Then:*

1. *For any $i, j \in S$, there are an integer $m > n$ and a vector $\mathbf{i} \in S^{m+1}$ with $i_1 = i$ and $i_{m+1} = j$ such that for any $t > 0$ the mapping $f_{\mathbf{i}} : \mathbb{R}_+^m \to M$ defined by*

$$f_{\mathbf{i}}(t_1, \ldots, t_m) := \Phi_{\mathbf{i}}\left(t_1, \ldots, t_m, t - \sum_{l=1}^{m} t_l, \xi\right) \tag{11}$$

   has a nonempty open set of regular points in the simplex

$$\Delta_{t,m} := \left\{(t_1, \ldots, t_m) \in \mathbb{R}_+^m : \sum_{l=1}^{m} t_l < t\right\}.$$

2. *The interior of $L(\xi)$ is nonempty and dense in $L(\xi)$.*

Theorem 6 *Assume that the weak hypoellipticity condition holds at some $\xi \in M$. Then:*

1. *For any $i, j \in S$, there are an integer $m > n$ and a vector $\mathbf{i} \in S^{m+1}$ with $i_1 = i$ and $i_{m+1} = j$ such that for any $t > 0$ the mapping $F_{\mathbf{i}} : \mathbb{R}_+^{m+1} \to M$ defined by*

$$F_{\mathbf{i}}(t_1, \ldots, t_{m+1}) := \Phi_{\mathbf{i}}(t_1, \ldots, t_{m+1}, \xi)$$

   has a nonempty open set of regular points in $\Delta_{t,m+1}$.

2. *The interior of $L(\xi)$ is nonempty and dense in $L(\xi)$.*

Appendix A contains a discussion of Theorems 5 and 6, including proof sketches. More detailed proofs and considerably more background information can be found in [24, Chapter 3]. The strong hypoellipticity condition is stronger than the weak hypoellipticity condition, so it is not surprising that the conclusion of Theorem 5 implies the conclusion of Theorem 6.

Theorem 5 shows that under the strong hypoellipticity condition, we can find a sequence of driving vector fields such that using that sequence and varying only the switching times we can generate an open set of terminal positions for any fixed terminal time $t > 0$. Moreover, the map assigning the terminal position at time t to the switching time sequence is regular, i.e. its Jacobian has full rank. We will use this theorem to conclude that, under this map, the pushforward of an absolutely continuous measure is also absolutely continuous. Under the weak hypoellipticity condition, such regularity for a fixed time t is not guaranteed. However, Theorem 6 shows that if it is allowed to vary also the terminal time t, we can still generate an open set of terminal positions and the Jacobian of the corresponding map still has full rank. This means that although the pushforward measures themselves do not necessarily have the desired regularity, their averages over terminal times t do, and we will use this argument to study the regularity of the resolvent measure of (X, A).

In addition to Theorems 5 and 6, we need the following result on the pushforward of an absolutely continuous measure under a regular transformation.

Lemma 2 *Let n and m be positive integers with $n \leq m$. Suppose that B and Δ are nonempty open sets in $\mathbb{R}^m$, $B \subset \Delta$, and that M is an n-dimensional smooth manifold. If $f : \Delta \to M$ is differentiable on B and all points in B are regular for f, then for any absolutely continuous probability measure μ on Δ satisfying $\mu(B) > 0$, the pushforward μf^{-1} is not singular with respect to λ^M.*

We will prove Lemma 2 only for the case $M = \mathbb{R}^n$. Our proof can be easily modified to include the general case by using coordinate patches on M.

We will use the following statement (see, e.g., Proposition 4.4 in [14]):

Lemma 3 *Let $f: B \to \mathbb{R}^m$ be a Borel function that is differentiable almost everywhere on an open set $B \subset \mathbb{R}^m$ and satisfies $\lambda^m\{\mathbf{t} \in B : \det Df(\mathbf{t}) = 0\} = 0$. If μ is absolutely continuous with respect to λ^m, then μf^{-1} is absolutely continuous with respect to λ^m, and*

$$\frac{d(\mu f^{-1})}{d\lambda^m}(\mathbf{s}) = \sum_{\mathbf{t}\in B: f(\mathbf{t})=\mathbf{s}} |\det Df(\mathbf{t})|^{-1} \frac{d\mu}{d\lambda^m}(\mathbf{t}).$$

Proof of Lemma 2: We can find an open set $B' \subset B$ such that $\mu(B') > 0$ and there are n columns of $Df(\mathbf{t})$ such that for any $\mathbf{t} \in B'$, the columns are linearly independent. We can assume without loss of generality that these columns are the first n columns of $Df(\mathbf{t})$. For $\rho: B' \to \mathbb{R}^n \times \mathbb{R}^{m-n}$ defined by

$$\rho: \mathbf{t} = (t_1, \ldots, t_m) \mapsto (f(\mathbf{t}), t_{n+1}, \ldots, t_m),$$

and any $\mathbf{t} \in B'$, we have

$$\det D\rho(\mathbf{t}) \neq 0.$$

Therefore, by Lemma 3, the pushforward of the restriction of μ to B' under ρ is a positive absolutely continuous measure on $\mathbb{R}^n \times \mathbb{R}^{m-n}$. Integrating over $\mathbb{R}^{m-n}$, we obtain that the pushforward of the restriction of μ to B' under f is a positive absolutely continuous measure on $\mathbb{R}^n$. This completes the proof. $\square$

We can now proceed to proving Theorems 4 and 3.

Proof of Theorem 4: To establish Theorem 4, we need to show that for any $t > 0$ and $i \in S$, the measure $\mathsf{P}^t_{\xi,i}$ is not singular.

Fix an index $i \in S$. We call a finite sequence $\mathbf{i}$ of indices in S with initial index i an admissible sequence. For any admissible $\mathbf{i}$, let $C_{\mathbf{i}}$ be the event that the driving vector fields up to time t appear in the order determined by $\mathbf{i}$. Since $\mathsf{P}_{\xi,i}(C_{\mathbf{i}}) > 0$ for any admissible $\mathbf{i}$, it suffices to find an admissible sequence $\mathbf{i}$ such that $\mathsf{P}^t_{\xi,i}(\cdot|C_{\mathbf{i}})$ is not singular. We claim that this holds true for the the sequence $\mathbf{i}$ provided by Theorem 5.

According to Theorem 5, there is an admissible sequence $\mathbf{i} = (i_1, i_2, \dots, i_{m+1})$ with $i_1 = i$ such that the function $f_{\mathbf{i}}$ has a regular point in $\Delta_{t,m}$. Since the set of regular points of a differentiable function is open in its domain, the function $f_{\mathbf{i}}$ is regular in a nonempty open set $B \subset \Delta_{t,m}$.

Let $T_1, T_2, \dots, T_{m+1}$ be independent and exponentially distributed random variables such that T_j has parameter λ_{i_j} for $1 \le j \le m+1$. On $C_{\mathbf{i}}$ we have $A_t = i_{m+1}$, and the distribution of X_t under $\mathsf{P}_{\xi,i}(\cdot|C_{\mathbf{i}})$ coincides with the distribution of $f_{\mathbf{i}}(T_1, \dots, T_m)$ conditioned on the event

$$R := \left\{ \sum_{j=1}^{m} T_j < t \le \sum_{j=1}^{m+1} T_j \right\}. \tag{12}$$

The distribution of the random vector $(T_1, \dots, T_m)$ conditioned on R, is equivalent to the uniform distribution on the simplex

$$\Delta_{t,m} := \left\{ (t_1, \dots, t_m) \in \mathbb{R}_+^m : \quad \sum_{j=1}^{m} t_j < t \right\}.$$

Now, Theorem 4 directly follows from Lemma 2. □

Proof of Theorem 3: We need to show that $\mathsf{Q}_{\xi,i}$ is not a singular measure. The proof is based on Theorem 6.

For the S-valued process A, we define $I_t(A)$ as the sequence of states visited by A between 0 and t. For any $m \in \mathbb{N}$ and any sequence $\mathbf{i} \in S^m$, we can introduce an auxiliary measure $\mathsf{Q}_{\xi,i,\mathbf{i}}$ on M by

$$\mathsf{Q}_{\xi,i,\mathbf{i}}(B) := \int_{R_+} e^{-t}\, \mathsf{P}_{\xi,i}\{X_t \in B \text{ and } I_t(A) = \mathbf{i}\}\, dt, \quad B \in \mathcal{B}(M).$$

Since

$$\mathsf{Q}_{\xi,i}(B \times \{j\}) = \sum_m \sum_{\mathbf{i}=(i,i_2,\dots,i_{m-1},j)\in S^m} \mathsf{Q}_{\xi,i,\mathbf{i}}(B), \tag{13}$$

it is sufficient to find $\mathbf{i} = (i_1, \dots, i_m)$ with $i_1 = i$ such that $\mathsf{Q}_{\xi,i,\mathbf{i}}(M) > 0$ and

$$\overline{\mathsf{Q}}_{\xi,i,\mathbf{i}}(\cdot) := \frac{\mathsf{Q}_{\xi,i,\mathbf{i}}(\cdot)}{\mathsf{Q}_{\xi,i,\mathbf{i}}(M)}$$

is a nonsingular probability measure. To apply Lemma 2, we need to represent $\overline{\mathsf{Q}}_{\xi,i,\mathbf{i}}$ as the pushforward of a measure equivalent to Lebesgue measure under a smooth map with a nonempty set of regular points.

Since the weak hypoellipticity condition holds at ξ, Theorem 6 yields an integer $m > n$ and a sequence $\mathbf{i} = (i_1, i_2, \dots, i_{m+1})$ with $i_1 = i$, such that the function $F_{\mathbf{i}} : \mathbb{R}_+^{m+1} \to M$ defined by

$$F_{\mathbf{i}}(\mathbf{t}) := \Phi_{\mathbf{i}}(\mathbf{t}, \xi)$$

has a regular point. For this $\mathbf{i}$ provided by Theorem 6, $\overline{\mathsf{Q}}_{\xi,i,\mathbf{i}}$ is the distribution of $\Phi_{\mathbf{i}}\left(T_1, \dots, T_m, T - \sum_{j=1}^m T_j, \xi\right)$ conditioned on

$$R := \left\{ \sum_{j=1}^{m} T_j < T \leq \sum_{j=1}^{m+1} T_j \right\}, \tag{14}$$

where $T_1, \dots, T_{m+1}$ and T are independent random variables that are exponentially distributed with parameters $\lambda_{i_1}, \dots, \lambda_{i_{m+1}}$ and 1, respectively.

Since the joint distribution of $T_1, \dots, T_{m+1}, T$ is equivalent to Lebesgue measure and since event R has positive probability, the distribution μ of $T_1, \dots, T_m, T$ conditioned on R induces a measure on

$$\Delta := \left\{ (t_1, \dots, t_m, t) \in \mathbb{R}_+^{m+1} : \quad \sum_{j=1}^{m} t_j < t \right\}$$

that is equivalent to Lebesgue measure. The regularity of $F_{\mathbf{i}}$, guaranteed by Theorem 6, implies that the function $f_{\mathbf{i}} : \ \Delta \to M$ defined by

$$f_{\mathbf{i}}(t_1, \dots, t_m, t) := F_{\mathbf{i}}\left(t_1, \dots, t_m, t - \sum_{j=1}^{m} t_j \right)$$

has a nonempty open set of regular points in Δ, and the proof is completed by an application of Lemma 2, since $\overline{\mathsf{Q}}_{\xi,i,\mathbf{i}}$ is the pushforward of μ under $f_{\mathbf{i}}$. $\square$

2.4 Proof of Theorem 2

According to the Ergodic Decomposition Theorem, all invariant measures for a Markov semigroup can be represented in terms of ergodic ones (see for instance [19, Theorem 1.7]). We will use this to derive the absolute continuity part of Theorem 2 from absolute continuity of ergodic invariant distributions.

To define ergodic measures, we need to recall the notion of μ-invariant sets. Let μ be an invariant measure for the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$. A set $A \in \mathcal{B}(M) \otimes \mathcal{P}(S)$ is μ-invariant if for every $t \geq 0$, we have $\mathsf{P}^t_{\xi,i}(A) = 1$ for μ-almost every $(\xi, i) \in A$. An invariant measure μ is called ergodic if for every μ-invariant set A, either $\mu(A) = 1$ or $\mu(A) = 0$. The Ergodic Decomposition Theorem then states that for any invariant measure μ, there is a unique probability measure P on the set of invariant measures I such that P is supported on the ergodic measures in I and

$$\mu = \int_I \nu P(d\nu).$$

The following is a basic result on systems with Markov switchings that does not use Conditions A or B.

Theorem 7 *If μ is invariant with respect to $(\mathsf{P}^t)_{t\geq 0}$ and ergodic, it is either absolutely continuous or singular.*

Proof: Consider the Lebesgue decomposition $\mu = \mu_{ac} + \mu_s$, where μ_{ac} is absolutely continuous and μ_s is singular with respect to Lebesgue measure. Let us show that both μ_{ac} and μ_s are invariant. For any $t > 0$, using the invariance of μ, we can write

$$\mu_{ac} + \mu_s = \mu = \mu\mathsf{P}^t = \mu_{ac}\mathsf{P}^t + \mu_s\mathsf{P}^t = \sum_{j=1}^{k} \nu_j + \mu_s\mathsf{P}^t, \tag{15}$$

where

$$\nu_j(\cdot) := \int_M \mathsf{P}^t_{\xi,j}(\cdot)\mu_{ac}(d\xi \times \{j\}), \quad j \in S. \tag{16}$$

We claim that the measures ν_j, $j \in S$, are absolutely continuous. To see this, we check that for any sequence $\mathbf{i} = (i_1, \ldots, i_{m+1})$ with $i_1 = j$, the measure $\nu_{\mathbf{i}}$ defined by

$$\begin{aligned} \nu_{\mathbf{i}}(E) &:= \int_M \mathsf{P}_{\xi,j}(X_t \in E | C_{\mathbf{i}}) \mu_{ac}(d\xi \times \{j\}) \\ &= \int_M \mathsf{P}\left(\Phi_{\mathbf{i}}\left(T_1, \ldots, T_m, t - \sum_{l=1}^m T_l, \xi \right) \in E \,\middle|\, R \right) \mu_{ac}(d\xi \times \{j\}) \end{aligned} \tag{17}$$

is absolutely continuous. Here we use the notation introduced in Section 2.3. In particular, we use the definition of R given in (12). Suppose that $\lambda^M(E) = 0$. For fixed $T_1, \ldots, T_m, T_{m+1}$, the map $\Phi_{\mathbf{i}}$ is a diffeomorphism in ξ. Therefore, on event R, we have

$$\mu_{ac}\left(\left\{ \xi \times \{j\} : \ \Phi_{\mathbf{i}}\left(T_1, \ldots, T_m, t - \sum_{l=1}^m T_l, \xi \right) \in E \right\} \right) = 0,$$

and $\nu_{\mathbf{i}}(E) = 0$ follows from disintegrating the right side of (17) and changing the order of integration.

Now, using (15) and the absolute continuity of ν_j, $j \in S$, we can write

$$\mu_{ac} = \sum_{j=1}^k \nu_j + (\mu_s \mathsf{P}^t)_{ac}. \tag{18}$$

Since $\mathsf{P}^t_{\xi,j}(M \times S) = 1$ for all ξ and j, (16) implies $\sum_{j=1}^k \nu_j(M \times S) = \mu_{ac}(M \times S)$. Therefore, applying (18) to $M \times S$, we obtain that the absolutely continuous component of the measure $\mu_s \mathsf{P}^t$ is zero. In other words, $\mu_s \mathsf{P}^t$ is singular, and from (15) and the absolute continuity of ν_j, $j \in S$, we obtain $\mu_s = \mu_s \mathsf{P}^t$. In other words, μ_s is invariant for $(\mathsf{P}^t)_{t \geq 0}$. It follows from (15) that μ_{ac} is also invariant. Since μ is ergodic, it cannot be represented as a sum of two nontrivial invariant measures. This means that either $\mu = \mu_{ac}$ or $\mu = \mu_s$. □

We endow the state space S with the discrete topology and recall that a point $(\xi, i) \in M \times S$ is contained in the support of a measure if and only if the measure of every open neighborhood of (ξ, i) is positive.

Theorem 8 *Let μ be an ergodic invariant measure for $(\mathsf{P}^t)_{t\geq 0}$. Assume that the support of μ contains a point (η, i) such that the weak hypoellipticity condition holds at η. Then, μ is absolutely continuous with respect to Lebesgue measure on $M \times S$.*

In order to prove Theorem 8, we need the following lemma.

Lemma 4 *Let ν be a finite Borel measure on $M \times S$ with support K. If U is any open set in $M \times S$ whose intersection with K is nonempty, we have*

$$\nu(U \cap K) > 0.$$

Proof: Assume that $\nu(U \cap K) = 0$. The complement of the support K has measure zero. Therefore,

$$\nu(U) = \nu(U \cap K) + \nu(U \cap K^c) = 0.$$

Thus, U^c is a closed subset of $M \times S$ whose complement has measure zero. From the definition of the support, we obtain that $K \subset U^c$. But then, $U \cap K$ must be empty, a contradiction. □

Proof of Theorem 8: According to Theorem 7, we need to show that μ is not singular. If μ is singular, it is entirely supported on a set $G \subset M \times S$ of Lebesgue measure 0, so $\mu(G^c) = 0$. Since μ is invariant with respect to $(\mathsf{P}^t)_{t\geq 0}$, it is also invariant with respect to Q. Therefore, $\mu(G^c) = \mu\mathsf{Q}(G^c)$, and we see that $\mu(V) = 0$, where

$$V := \{(\xi, j) \in M \times S : \ \mathsf{Q}_{\xi,j}(G^c) > 0\}.$$

Let U be the set of points $\xi \in M$ where the weak hypoellipticity condition holds. Due to Theorem 3, $U \times S \subset V$, and we conclude that $\mu(U \times S) = 0$. Recall that U is an open subset of M, and $(U \times S) \cap \operatorname{supp} \mu \neq \emptyset$ by assumption. Lemma 4 implies that $\mu((U \times S) \cap \operatorname{supp} \mu) > 0$. This contradicts $\mu(U \times S) = 0$, completing the proof. □

If one replaces the weak hypoellipticity condition in Theorem 8 with the strong hypoellipticity condition, the resulting statement holds automatically, but one can give a proof that does not involve the resolvent Q, see [1, Theorem 8].

Next, we establish two properties of the set $E := L \cap U$, where U is the open set of points satisfying the weak hypoellipticity condition and L is the set of points that are D-accessible from all other points in M.

Lemma 5 *The set E has nonempty interior.*

Proof: By assumption, $\xi \in E$, so $U \neq \emptyset$ and $L(\xi) \cap U \neq \emptyset$ because the vector fields in D are continuous. Since $\xi \in U$, Theorem 6 implies that $L(\xi)$ has nonempty interior that is dense in $L(\xi)$. Therefore, the set

$$V := L(\xi)^\circ \cap U$$

is nonempty and open. Clearly, $V \subset U$, and it remains to prove that $L(\xi)^\circ \subset L$. In fact, we even have that $L(\xi) \subset L$. To see this, let us fix any $\zeta \in L(\xi)$, $\eta \in M$, and prove that $\zeta \in \overline{L(\eta)}$. Since $\zeta \in L(\xi)$, we have

$$\zeta = \Phi_{\mathbf{i}}(\mathbf{t}, \xi)$$

for some index sequence $\mathbf{i}$ and some time sequence $\mathbf{t}$. Let us fix a neighborhood W of ζ. Since the mapping $x \mapsto \Phi_{\mathbf{i}}(\mathbf{t}, x)$ is continuous, the inverse image of W under this map is an open neighborhood of ξ. Since ξ is D-accessible from η, this open neighborhood of ξ contains a point that is D-reachable from η. Hence, W contains a point that is D-reachable from η. $\square$

As an immediate corollary of Lemma 5, the set L has nonempty interior.

Lemma 6 *Suppose μ is an invariant measure for $(\mathsf{P}^t)_{t \geq 0}$. If G is a nonempty open subset of L and $j \in S$, then $\mu_j(G) > 0$.*

Proof: Let us assume that $\mu(G \times \{j\}) = 0$. Since μ is invariant with respect to $(\mathsf{P}^t)_{t\geq 0}$, it is also invariant with respect to Q, and we have

$$0 = \mu_j(G) = \sum_{i=1}^{k} \int_M \mathsf{Q}_{\eta,i}(G \times \{j\})\mu_i(d\eta).$$

For all $i \in S$ and μ_i-almost every $\eta \in M$, we thus obtain

$$\mathsf{Q}_{\eta,i}(G \times \{j\}) = 0. \tag{19}$$

Choose a point $\eta \in M$ for which (19) holds. By assumption, $G \subset L \subset \overline{L(\eta)}$. Since G is open, $G \cap L(\eta) \neq \emptyset$. So, there exist a sequence $\mathbf{i} = (i, i_2, \ldots, i_m, j)$ and a vector of switching times $\mathbf{t} = (t_1, \ldots, t_m, t_{m+1})$ such that $\Phi_{\mathbf{i}}(\mathbf{t}, \eta) \in G$. By continuity of $\Phi_{\mathbf{i}}$, there is a neighborhood W of $\mathbf{t}$ in $\mathbb{R}_+^{m+1}$ such that $\Phi_{\mathbf{i}}(\mathbf{s}, \eta) \in G$ for all $\mathbf{s} \in W$. Defining $s := s_1 + \ldots + s_{m+1}$ and using the representation of $\mathsf{P}^s_{\eta,i}(\cdot|C_{\mathbf{i}})$ via exponentially distributed times from the proof of Theorem 4, we conclude that $\mathsf{P}^s_{\eta,i}(G \times \{j\}) > 0$ for s sufficiently close to $t := t_1 + \ldots + t_{m+1}$. Therefore, $\mathsf{Q}_{\eta,i}(G \times \{j\}) > 0$, contradicting (19). □

Proof of Theorem 2: By the Ergodic Decomposition Theorem, it suffices to show absolute continuity and uniqueness of an ergodic invariant measure.

We first derive absolute continuity. If μ is an ergodic invariant measure that satisfies the assumptions of Theorem 2, it suffices to show that $L \subset \operatorname{supp} \mu$ in light of Theorem 8. Let $j \in S$, $\xi \in L$, and let U be a neighborhood of ξ in M. By Lemma 6, we have $\mu_j(U) > 0$, hence $\xi \in \operatorname{supp} \mu$.

Next, we show uniqueness of the ergodic invariant measure. Let us assume that $\mu^{(1)}$ and $\mu^{(2)}$ are two distinct ergodic invariant probability measures, and lead this assumption to a contradiction. The Ergodic Decomposition Theorem implies that $\mu^{(1)}$ and $\mu^{(2)}$ are mutually singular. Hence, the set $M \times S$ can be partitioned into two sets H_1 and H_2 with $\mu^{(1)}(H_2) = \mu^{(2)}(H_1) = 0$. The two sets can be represented as

$$H_a = \bigcup_{j=1}^{k} M_{a,j} \times \{j\}, \quad a = 1, 2$$

for some measurable sets $M_{a,j}$, $j \in S$, $a = 1, 2$. It is clear that $M_{1,j} \cup M_{2,j} = M$ for all $j \in S$. For all $a \in \{1, 2\}$ and $j \in S$,

$$\mu^{(a)}(M_{a,j} \times \{j\}) = \mu^{(a)}(M \times \{j\}) > 0,$$

since the left side is a stationary distribution for the Markov chain on S and by our assumptions, transitions between all states happen with positive probability.

Fix a j in S. By Lemma 5, the set E° is nonempty. By Lemma 6, we have $\mu^{(1)}(E^{\circ} \times \{j\}) > 0$ for all $j \in S$. Since $\mu^{(1)}(M_{2,j} \times \{j\}) = 0$, we deduce that $\mu^{(1)}(E_1 \times \{j\}) > 0$, where $E_1 := E^{\circ} \cap M_{1,j}$. The measure $\mu^{(1)}$ is invariant with respect to $(\mathsf{P}^t)_{t \geq 0}$, hence it is also invariant with respect to Q, and we have

$$0 = \mu^{(1)}(M_{2,j} \times \{j\}) \geq \int_{E_1} \mathsf{Q}_{\eta,j}(M_{2,j} \times \{j\}) \mu^{(1)}(d\eta \times \{j\}). \tag{20}$$

Since $\mu^{(1)}(E_1 \times \{j\}) > 0$, it suffices to show that $\mathsf{Q}_{\eta,j}(M_{2,j} \times \{j\}) > 0$ for all $\eta \in E_1$ to obtain a contradiction with (20).

Since η satisfies the weak hypoellipticity condition, Theorem 6 guarantees that there exist an integer $m > n$ and a vector $\mathbf{i} = (j, i_2, \ldots, i_m, j)$ such that the function $f: \mathbb{R}_+^{m+1} \to M$ defined by

$$f(\mathbf{t}) := \mathbf{\Phi}_{\mathbf{i}}(\mathbf{t}, \eta) \tag{21}$$

has an open set O of regular points and

$$\{\mathbf{t} = (t_1, \ldots, t_{m+1}) \in O: \; t_1 + \ldots + t_{m+1} < t\} \neq \emptyset$$

for all $t > 0$. Therefore, the map F defined by

$$F(t_1, \ldots, t_{m+1}, t) := f\left(t_1, \ldots, t_m, t - \sum_{l=1}^{m} t_l\right)$$

on

$$\Delta := \left\{(t_1, \ldots, t_{m+1}, t) \in \mathbb{R}_+^{m+2}: \; \sum_{l=1}^{m} t_l < t < \sum_{l=1}^{m+1} t_l\right\}$$

has an open set $V \subset \Delta$ of regular points such that

$$\{\mathbf{t} = (t_1, \ldots, t_{m+1}, t) \in V: \; t < s\} \neq \emptyset, \quad s > 0. \tag{22}$$

The fact that F is regular on V implies that $F(V)$ is open. Using the representation of Q via (13) and the family of exponentially distributed times $T_1, \ldots, T_{m+1}, T$, we obtain that it is sufficient to prove that

$$\mathsf{P}\{F(T_1, \ldots, T_{m+1}, T) \in M_{2,j} | R\} > 0, \tag{23}$$

where R was introduced in (14). Since E° is an open set containing η, and $F(V)$ is an open set such that $\eta \in \overline{F(V)}$ (due to (22) and continuity of F at 0), we obtain that $G := E^\circ \cap F(V)$ is also a nonempty open set.

Let us choose a vector $\mathbf{r} \in V$ such that $F(\mathbf{r}) \in E^\circ$. Since $\mathbf{r}$ is a regular point for F, we see that for an arbitrary choice of local smooth coordinates around $\mathbf{r}$, there are n independent columns of the matrix $DF(\mathbf{s})$ for $\mathbf{s}$ in a small neighborhood of $\mathbf{r}$. Without loss of generality, we can assume that these are the first n columns. Then, the map $\rho : \mathbb{R}^{m+2} \to M \times \mathbb{R}^{m+2-n}$ defined by

$$\rho(s_1, \ldots, s_{m+1}, s) := (F(s_1, \ldots, s_{m+1}, s), s_{n+1}, \ldots, s_{m+1}, s)$$

has nonzero Jacobian in that neighborhood. Therefore, we can choose an open set W_V containing $\mathbf{r}$ so that ρ is a diffeomorphism between W_V and $W_G \times W_{m+2-n}$, where $W_G \subset G$ and $W_{m+2-n} \subset \mathbb{R}_+^{m+2-n}$ are some open sets.

The set W_G is an open subset of L. It is also not empty since it contains $F(\mathbf{r})$. Lemma 6 implies that $\mu^{(2)}(W_G \times \{j\}) > 0$. Since $\mu^{(2)}(M_{2,j}^c \times \{j\}) = 0$, we conclude that $\mu^{(2)}(J \times \{j\}) > 0$, where $J := M_{2,j} \cap W_G$. Since $\mu^{(2)}$ is an ergodic measure, it is absolutely continuous, so

$$\lambda^M(J) > 0. \tag{24}$$

Since $J \subset M_{2,j}$, the desired inequality (23) will follow from

$$\mathsf{P}\{F(T_1, \ldots, T_{m+1}, T) \in J | R\} > 0. \tag{25}$$

Since the joint distribution of $T_1, \ldots, T_{m+1}, T$ is equivalent to the Lebesgue measure on Δ, Lemma 3 implies that $\rho(T_1, \ldots, T_{m+1}, T)$ has positive density almost everywhere in $W_G \times W_{m+2-n}$. Integrating over W_{m+2-n}, we see that $F(T_1, \ldots, T_{m+1}, T)$ has positive density almost everywhere in W_G. Now (25) follows from (24). □

Of course, Theorem 2 remains true if one replaces the weak hypoellipticity condition with the strong hypoellipticity condition. Under the strong hypoellipticity conition, one can prove this result without referring to the resolvent Q. Namely, one can use the regularity of transition probabilities established in Theorem 4 and invoke Theorems 5 and [1, Theorem 8] instead of Theorems 6 and 8.

2.5 Examples

Let us first consider Example 3. For any fixed time $t > 0$, the set of points D-reachable from the origin at time t is the image of

$$\left\{(s_1, \ldots, s_n) \in [0, \infty)^n : \quad \sum_{j=1}^{n} s_j = t\right\}$$

under the covering map $\mathbb{R}^n \to \mathbb{T}^n$, and has Lebesgue measure zero. Thus, $\mathsf{P}^t_{\xi,i}$ is a purely singular measure. This implies that the strong hypoellipticity condition does not hold for this system: If the strong hypoellipticity condition was satisfied at some point $\xi \in \mathbb{T}^n$, the transition probability measures $\mathsf{P}^t_{\xi,i}$ would not be singular with respect to Lebesgue measure, according to Theorem 4.

It is also instructive to show directly why the strong hypoellipticity condition does not hold. As all vector fields in D are constant, the derived algebra $\mathcal{I}'(D)$ contains only the zero vector field. Thus, for any $\xi \in \mathbb{T}^n$,

$$\mathcal{I}_0(D)(\xi) = \left\{\sum_{i=1}^{n} \nu_i u_i : \ \sum_{i=1}^{n} \nu_i = 0\right\}.$$

Due to the constraint $\sum_{i=1}^{n} \nu_i = 0$, the algebra $\mathcal{I}_0(D)(\xi)$ does not have full dimension, so the strong hypoellipticity condition is violated at every point in $\mathbb{T}^n$.

On the other hand, the weak hypoellipticity condition is clearly satisfied at any point $\xi \in \mathbb{T}^n$, as the standard basis of $\mathbb{R}^n$ applied to ξ yields a full-dimensional set of vectors in the tangent space. Also note that any point in $\mathbb{T}^n$ is D-reachable from any other point. Therefore, Theorem 2 guarantees that the associated Markov semigroup has a unique invariant measure, provided that such a measure exists. In this elementary example, it is possible to point out the invariant measure explicitly. If all switching rates are equal, the invariant measure is given by

$$\mu(E \times \{i\}) = \lambda(E), \quad E \in \mathcal{B}(\mathbb{T}^n), i \in S,$$

where λ denotes Lebesgue measure on $\mathbb{T}^n$.

Example 4 provides a situation where the number of vector fields in D is less than the dimension of the manifold M, each individual vector field in D gives rise to dynamics with a strange attractor and no absolutely continuous invariant measures, but the switching system has a unique invariant measure that is absolutely continuous.

In [35], Tucker shows that the Lorenz system with parameters $\sigma = 10$, $r = 28$ and $b = \frac{8}{3}$, corresponding to vector field u, admits a robust strange attractor Λ as well as a unique SRB-measure supported on Λ (see [38] for background information on SRB-measures). Robustness implies that the dynamical structure of the system remains intact under small parameter changes, so the dynamics induced by v share these features if r_v is sufficiently close to r_u. Moreover, the SRB-measure on Λ satisfies a dissipative ergodic theorem, which can be inferred from [3, Section 5.1], using Tucker's result. It follows that any point $\xi \in \Lambda$ is $\{u\}$-accessible (and thus D-accessible) from every point in a set $S_\xi \subset \mathbb{R}^3$ whose complement has Lebesgue measure zero.

Assisted by a computer algebra system, we checked that the strong hypoellipticity condition is satisfied for this system at any point in $\mathbb{R}^3$ that does not lie on the z-axis. Since the z-axis is invariant under the flows of both vector fields, we disregard it and set M to be $\mathbb{R}^3$ without points on the z-axis. With this provision, every point on the attractor Λ is D-accessible from any point in M:

Consider a point $\xi \in \Lambda$ and a point $\eta \in M$. By Theorem 5, there is a nonempty open set of D-reachable points from η (recall that the strong hypoellipticity condition holds at every point in M). And since this open set has positive Lebesgue measure, it contains a point belonging to S_ξ. Hence, ξ is D-accessible from η.

The only remaining condition of Theorem 2 that we need to check is existence of an invariant distribution. An elementary calculation similar to that for the case of one vector field (see, e.g., [21, Section 14.2]) shows that if r_v is sufficiently close to r_u, then the function

$$V(x, y, z) := r_u x^2 + \sigma y^2 + \sigma(z - 2r_u)^2$$

plays the role of a Lyapunov function for both vector fields u and v. Namely, there is a number $\nu > 0$ such that $\langle u, \nabla V\rangle < 0$ and $\langle v, \nabla V\rangle < 0$ if $V \geq \nu$. In particular, the compact set $\{(x, y, z) : \ V(x, y, z) \leq \nu\}$ is invariant for both vector fields, and a standard application of the Krylov–Bogoliubov method (see Section 3.2) shows that the system has an invariant distribution. As in Example 3, uniqueness and absolute continuity of an invariant measure follow now from Theorem 2.

CHAPTER III

ERGODICITY

In this chapter, we collect several conditions that guarantee exponential convergence of the distribution of $(X, A)_t$ to an invariant measure in a suitable metric as t goes to infinity. The chapter is based on work by Michel Benaïm, Stéphane Le Borgne, Florent Malrieu and Pierre-André Zitt. In Section 3.1, we state a lemma due to Benaïm, Le Borgne, Malrieu and Zitt that guarantees existence of minorizing measures for compact subsets of M under the strong hypoellipticity condition. In Section 3.2, we use this lemma to derive exponential convergence to the invariant measure in total-variation distance on a compact manifold M. This result is also due to Benaïm, Le Borgne, Malrieu and Zitt. Besides, we present the Krylov–Bogoliubov method to establish existence of an invariant measure if M is compact. In Section 3.3, we briefly discuss exponential convergence to the invariant measure in a noncompact setting. Finally, we apply some of the ergodicity results in this chapter to our examples.

3.1 Existence of a minorizing measure

We first address under which assumptions there exists a minorizing measure on $M \times S$. We call a probability measure ν on $M \times S$ minorizing with respect to the Markov semigroup $(\mathsf{P}^t)_{t \geq 0}$ of (X, A) and a compact set $K \subset M$ if there exist a constant $c > 0$ and a time $t > 0$ such that

$$\inf_{\eta \in K, i \in S} \mathsf{P}^t_{\eta,i}(E \times \{l\}) \geq c \cdot \nu(E \times \{l\})$$

for all measurable sets $E \subset M$ and for all $l \in S$. Notice that the lower bound is uniform in (η, i), at least over a compact subset of $M \times S$. See also [20, Assumption 2]. Existence of a minorizing measure is reminiscent of Doeblin's condition in the case of a discrete state space ([20, page 2]).

Under the strong hypoellipticity condition, we have the following statement.

Lemma 7 *Assume that the strong hypoellipticity condition holds at a D-accessible point ξ, and let $K \subset M$ be compact. Then, there exists a measure ν that is minorizing with respect to $(\mathsf{P}^t)_{t\geq 0}$ and K.*

Lemma 7 is due to Benaïm, Le Borgne, Malrieu and Zitt, see [6]. In [6], the authors assume that M is compact. Our version of the statement is a minimal extension that follows immediately.

Sketch of Proof: We give an idea of how to prove Lemma 7. See [6] for the details. If the strong hypoellipticity condition holds at a point $\xi \in M$, we have the following local regularity result (see [6, Theorems 4.2, 4.4]): There exist an open neighborhood U of ξ, an open set $V \subset M$, a time $T > 0$, an index $j \in S$ and constants $\bar{c}, \epsilon > 0$ such that

$$\inf_{\eta \in U, i \in S, t \in [T, T+\epsilon]} \mathsf{P}^t_{\eta,i}(E \times \{l\}) \geq \bar{c} \cdot \lambda^M(E \cap V) \cdot \delta_{l,j}$$

for all measurable sets $E \subset M$ and for all $l \in S$. Here, λ^M denotes Lebesgue measure on M and $\delta_{l,j}$ is the Kronecker delta. In addition, one has to establish global lower bounds on transition probabilities to neighborhoods of D-accessible points. If the strong hypoellipticity condition is satisfied at a D-accessible point ξ, the following statement holds (see [6, Equation (20), page 16]): For any open neighborhood U of ξ and for any compact set $K \subset M$, there exist a time $t > 0$ and a constant $\alpha > 0$ such that

$$\inf_{\eta \in K, i \in S} \mathsf{P}_{\eta,i}(X_t \in U) \geq \alpha.$$

Using these global lower bounds on transition probabilities, we can extend the local regularity result with the help of the Chapman–Kolmogorov equations to obtain existence of a minorizing measure. □

3.2 The Krylov–Bogoliubov method and convergence in total variation

Suppose that M is compact or that the process X is eventually confined to a bounded subset of M. Under this assumption, existence of an invariant measure is guaranteed by the Krylov–Bogoliubov method, see for instance [11, Theorem 3.1.1]. The argument for switching systems goes as follows:

Assume without loss of generality that M is compact. Under our general assumptions on the switching system, the semigroup $(\mathsf{P}^t)_{t\geq 0}$ is Feller, see [6, Proposition 2.1]. This means that for any bounded and continuous function $f:\ M\times S\to\mathbb{R}$, the function $\mathsf{P}^t f$, defined by (3), is also bounded and continuous for all $t\geq 0$. For more general classes of piecewise deterministic Markov processes, $(\mathsf{P}^t)_{t\geq 0}$ doesn't have to be Feller, see [13, Example 27.5]. Fix a point $(\xi,i)\in M\times S$. For $T>0$, define the probability measure

$$\mu_T(E\times\{j\}) := \frac{1}{T}\cdot\int_0^T \mathsf{P}^t_{\xi,i}(E\times\{j\})\ dt,\quad E\in\mathcal{B}(M), j\in S.$$

The family of measures $(\mu_T)_{T>0}$ is tight because $M\times S$ is compact. By Prokhorov's theorem, tightness of $(\mu_T)_{T>0}$ implies that the family is relatively compact with respect to the topology induced by weak convergence. This means there is a monotone increasing sequence of times $(T_l)_{l\geq 1}$ that diverges to $+\infty$ as well as a probability measure μ on $M\times S$ such that

$$\lim_{l\to\infty}\sum_{j\in S}\int_M f(\eta,j)\ \mu_{T_l}(d\eta\times\{j\}) = \sum_{j\in S}\int_M f(\eta,j)\ \mu(d\eta\times\{j\}) \tag{26}$$

for all continuous and bounded functions f on $M\times S$. For a fixed positive integer l, we have

$$\begin{aligned}\sum_{j\in S}\int_M f(\eta,j)\ \mu_{T_l}(d\eta\times\{j\}) &= \sum_{j\in S}\int_M f(\eta,j)\frac{1}{T_l}\int_0^{T_l} \mathsf{P}^t_{\xi,i}(d\eta\times\{j\})\ dt\\ &= \frac{1}{T_l}\int_0^{T_l}\sum_{j\in S}\int_M f(\eta,j)\ \mathsf{P}^t_{\xi,i}(d\eta\times\{j\})\ dt\\ &= \frac{1}{T_l}\int_0^{T_l}\mathsf{P}^t f(\xi,i)\ dt.\end{aligned}$$

Thus, we can write (26) as

$$\lim_{l\to\infty}\frac{1}{T_l}\int_0^{T_l}\mathsf{P}^t f(\xi,i)\ dt = \sum_{j\in S}\int_M f(\eta,j)\ \mu(d\eta\times\{j\}). \tag{27}$$

We want to show that μ is an invariant measure for $(\mathsf{P}^t)_{t\geq 0}$. To this end, it suffices to verify identity (4) from Section 1.3 for a bounded and continuous function f. Fix a bounded and continuous function f on $M\times S$ and let $r>0$. Since $(\mathsf{P}^t)_{t\geq 0}$ is Feller, the function $g(\eta,j) := \mathsf{P}^r f(\eta,j)$ is also continuous and bounded. Replacing f with g in (26) and using the semigroup property, we obtain

$$\lim_{l\to\infty}\frac{1}{T_l}\int_0^{T_l}\mathsf{P}^{t+r} f(\xi,i)\ dt = \sum_{j\in S}\int_M \mathsf{P}^r f(\eta,j)\ \mu(d\eta\times\{j\}). \tag{28}$$

For any positive integer l, the substitution $s=t+r$ yields

$$\begin{aligned}\frac{1}{T_l}\int_0^{T_l}\mathsf{P}^{t+r}f(\xi,i)\ dt &= \frac{1}{T_l}\int_r^{T_l+r}\mathsf{P}^s f(\xi,i)\ ds\\ &= \frac{1}{T_l}\int_0^{T_l}\mathsf{P}^s f(\xi,i)\ ds + \frac{1}{T_l}\int_{T_l}^{T_l+r}\mathsf{P}^s f(\xi,i)\ ds - \frac{1}{T_l}\int_0^{r}\mathsf{P}^s f(\xi,i)\ ds.\end{aligned}$$

If we let r go to 0, the term on the right converges to

$$\frac{1}{T_l}\int_0^{T_l}\mathsf{P}^s f(\xi,i)\ ds$$

by dominated convergence. And since $(\mathsf{P}^t)_{t\geq 0}$ is stochastically continuous, the term to the right of the equality sign in (28) converges to

$$\sum_{j\in S}\int_M f(\eta,j)\mu(d\eta\times\{j\})$$

as r tends to 0. We have thus shown that

$$\lim_{l\to\infty}\frac{1}{T_l}\int_0^{T_l} \mathsf{P}^s f(\xi,i)\; ds = \sum_{j\in S}\int_M f(\eta,j)\mu(d\eta\times\{j\}).$$

Applying (27) to the term on the left, we obtain (4). This shows that μ is indeed an invariant measure.

The total-variation distance of two probability measures μ and ν on $(M\times S, \mathcal{B}(M)\otimes \mathcal{P}(S))$ is defined as

$$tv(\mu,\nu) := \frac{1}{2}\sup_{\|f\|_\infty\le 1}\left(\sum_{i\in S}\int_M f(\xi,i)\;\mu(d\xi\times\{i\}) - \sum_{i\in S}\int_M f(\xi,i)\;\nu(d\xi\times\{i\})\right), \quad (29)$$

where the supremum is taken over the set of measurable functions on $M\times S$ that are bounded by 1 (see [10, page 2]). Alternatively, one can define the total-variation distance in terms of couplings. A coupling of two probability measures μ and ν on $M\times S$ is a measure Γ on the product space $(M\times S)\times(M\times S)$, endowed with the σ-algebra $(\mathcal{B}(M)\otimes\mathcal{P}(S))\otimes(\mathcal{B}(M)\otimes\mathcal{P}(S))$, such that μ and ν are the marginals of Γ. For example, the product measure $\mu\otimes\nu$ is a coupling of μ and ν. Let $C(\mu,\nu)$ denote the set of couplings of μ and ν. Then,

$$\begin{aligned} tv(\mu,\nu) &= \inf_{\Gamma\in C(\mu,\nu)} \Gamma(\{((\xi,i),(\eta,j))\in(M\times S)\times(M\times S):\; (\xi,i)\neq(\eta,j)\}) \\ &= \inf_{Y,Z:\mathcal{L}(Y,Z)\in C(\mu,\nu)} Pr(Y\neq Z). \end{aligned} \quad (30)$$

In (30), the infimum is taken over all random variables Y and Z whose joint distribution is in $C(\mu,\nu)$. We assume that Y and Z are defined on a common probability space with probability measure Pr. The equivalence of the two definitions of the total-variation distance follows from the Kantorovich–Rubinstein formula, see [36, Particular Case 5.16]. We will need the following property of the total-variation distance in the proof of Theorem 9.

Lemma 8 *Let μ and ν be probability measures and let P be a transition probability kernel on $(M\times S,\mathcal{B}(M)\otimes\mathcal{P}(S))$. Then,*

$$tv(\mu\mathsf{P},\nu\mathsf{P})\le tv(\mu,\nu).$$

PROOF: Let f be a measurable function on $M \times S$ that is bounded by 1. Then, $\mathsf{P}f$ is also measurable and

$$\begin{aligned}
|\mathsf{P}f(\xi,i)| &= \left|\sum_{j\in S}\int_M f(\eta,j)\ \mathsf{P}_{\xi,i}(d\eta\times\{j\})\right| \\
&\leq \sum_{j\in S}\int_M |f(\eta,j)|\ \mathsf{P}_{\xi,i}(d\eta\times\{j\}) \\
&\leq \mathsf{P}_{\xi,i}(M\times S) = 1.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{i\in S}\int_M f(\xi,i)\ \mu\mathsf{P}(d\xi\times\{i\}) - \sum_{i\in S}\int_M f(\xi,i)\ \nu\mathsf{P}(d\xi\times\{i\}) \\
=&\sum_{i\in S}\int_M \mathsf{P}f(\xi,i)\ \mu(d\xi\times\{i\}) - \sum_{i\in S}\int_M \mathsf{P}f(\xi,i)\ \nu(d\xi\times\{i\}) \\
\leq&\sup_{\|g\|_\infty\leq 1}\left(\sum_{i\in S}\int_M g(\xi,i)\ \mu(d\xi\times\{i\}) - \sum_{i\in S}\int_M g(\xi,i)\ \nu(d\xi\times\{i\})\right) \\
=&2tv(\mu,\nu).
\end{aligned}$$

Taking the supremum over all measurable functions f with $\|f\|_\infty \leq 1$ yields Lemma 8.

□

In addition to compactness of M, we assume that the strong hypoellipticity condition holds at a D-accessible point in M. By Theorem 2, the invariant measure is unique. The next theorem asserts that the distribution of $(X,A)_t$ converges in total-variation distance to this unique invariant measure at an exponential rate.

Theorem 9 *Let μ denote the unique invariant measure of $(\mathsf{P}^t)_{t\geq 0}$. There exist constants $c > 1$ and $\alpha > 0$ such that*

$$tv(\pi\mathsf{P}^t,\mu) \leq c\cdot e^{-\alpha t}$$

for any probability measure π on $M \times S$ and for any $t \geq 0$.

Theorem 9 is due to Benaïm, Le Borgne, Malrieu and Zitt (see [6, Theorem 4.6]). Given Lemma 7, the proof is a standard exercise in the use of the coupling method. The proof we give was suggested to us by Jonathan Mattingly. See also [29, Chapter 5]. When moving from discrete to continuous time, we follow [28, Section III.20].

PROOF: Fix a probability measure π on $M \times S$. By Lemma 7, there exist a probability measure ν on $M \times S$, a constant $\bar{c} \in (0,1)$ and a time $s > 0$ such that

$$\inf_{\eta \in M, i \in S} \mathsf{P}^s_{\eta,i}(E \times \{l\}) \geq \bar{c} \cdot \nu(E \times \{l\})$$

for all measurable sets $E \subset M$ and for all $l \in S$. For this s, we define the transition probability kernel $\bar{\mathsf{P}} := \mathsf{P}^s$. We will proceed according to the following strategy. For each integer $N \geq 1$, we construct a coupling of the measures $\pi\bar{\mathsf{P}}^N$ and $\mu\bar{\mathsf{P}}^N$. These couplings will be constructed in such a way that the tails of the associated coupling times decay exponentially as N goes to infinity. With (30), we obtain an exponentially decaying upper bound for $tv(\pi\bar{\mathsf{P}}^N, \mu)$.

Let $(Y_m)_{m\geq 0}$ and $(Z_m)_{m\geq 0}$ be Markov chains on $M \times S$ with transition probability kernel $\bar{\mathsf{P}}$ and initial distributions π and μ, respectively. For $N \geq 1$, we define four additional Markov chains $(\hat{Y}^{(N)}_m)_{m\geq 0}, (\tilde{Y}^{(N)}_m)_{m\geq 1}, (\hat{Z}^{(N)}_m)_{m\geq 0}$ and $(\tilde{Z}^{(N)}_m)_{m\geq 1}$ on $M \times S$ as follows: Let $(U_m)_{m\geq 1}$ be a sequence of independent, identically distributed random variables with values in $(M \times S, \mathcal{B}(M) \otimes \mathcal{P}(S))$, and assume that U_1 is distributed according to ν. Let $(\beta_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, and assume that β_1 is Bernoulli distributed, with $P(\beta_1 = 1) = \bar{c}$ and $P(\beta_1 = 0) = 1 - \bar{c}$. For each $(\eta, i) \in M \times S$, we define the measure

$$\tilde{\mathsf{P}}_{\eta,i}(E \times \{l\}) := \frac{1}{1-\bar{c}} \cdot (\bar{\mathsf{P}}_{\eta,i}(E \times \{l\}) - \bar{c}\nu(E \times \{l\})), \quad E \in \mathcal{B}(M), l \in S.$$

Let $\hat{Y}^{(N)}_0$ be distributed according to π and let $\hat{Z}^{(N)}_0$ be distributed according to μ. For $0 \leq l \leq N$, let $\tilde{Y}^{(N)}_{l+1}$ be distributed according to $\tilde{\mathsf{P}}_{\eta,i}$, provided that $\hat{Y}^{(N)}_l = (\eta, i)$, and let $\tilde{Z}^{(N)}_{l+1}$ be distributed according to $\tilde{\mathsf{P}}_{\eta,i}$, provided that $\hat{Z}^{(N)}_l = (\eta, i)$. Then, let

$$\hat{Y}^{(N)}_{l+1} := \beta_{l+1} U_{l+1} + (1 - \beta_{l+1})\tilde{Y}^{(N)}_{l+1}$$

and

$$\hat{Z}_{l+1}^{(N)} := \begin{cases} \beta_{l+1}U_{l+1} + (1-\beta_{l+1})\tilde{Z}_{l+1}^{(N)}, & \hat{Y}_l^{(N)} \neq \hat{Z}_l^{(N)} \\ \hat{Y}_{l+1}^{(N)}, & \hat{Y}_l^{(N)} = \hat{Z}_l^{(N)}. \end{cases}$$

For $l > N$, let $\hat{Y}_{l+1}^{(N)} := Y_{l+1}$ and let $\hat{Z}_{l+1}^{(N)} := Z_{l+1}$.

Let us show that the processes $(Y_m)_{m\geq 0}$ and $(\hat{Y}_m^{(N)})_{m\geq 0}$ are identically distributed. From its construction, it is clear that $(\hat{Y}_m^{(N)})_{m\geq 0}$ is also a Markov chain with initial distribution π. For $0 \leq l \leq N$ and for $(\eta, i) \in M \times S$, we have

$$\begin{aligned}
&Pr(\hat{Y}_{l+1}^{(N)} \in E \times \{j\} | \hat{Y}_l^{(N)} = (\eta, i)) \\
=&Pr((\beta_{l+1}U_{l+1} + (1-\beta_{l+1})\tilde{Y}_{l+1}^{(N)}) \in E \times \{j\} | \hat{Y}_l^{(N)} = (\eta, i)) \\
=&Pr(\beta_{l+1} = 1) \cdot Pr(U_{l+1} \in E \times \{j\}) + Pr(\beta_{l+1} = 0) \cdot Pr(\tilde{Y}_{l+1}^{(N)} \in E \times \{j\} | \hat{Y}_l^{(N)} = (\eta, i)) \\
=&\bar{c} \cdot \nu(E \times \{j\}) + (1-\bar{c}) \cdot \tilde{\mathsf{P}}_{\eta,i}(E \times \{j\}) \\
=&\bar{c}\nu(E \times \{j\}) + \bar{\mathsf{P}}_{\eta,i}(E \times \{j\}) - \bar{c}\nu(E \times \{j\}) \\
=&\bar{\mathsf{P}}_{\eta,i}(E \times \{j\}), \quad E \in \mathcal{B}(M), j \in S.
\end{aligned}$$

For $l > N$,

$$Pr(\hat{Y}_{l+1}^{(N)} \in E \times \{j\} | \hat{Y}_l^{(N)} = (\eta, i)) = \bar{\mathsf{P}}_{\eta,i}(E \times \{j\}), \quad E \in \mathcal{B}(M), j \in S$$

follows immediately from the construction of $\hat{Y}^{(N)}$. Along the same lines, one can show that $(Z_m)_{m\geq 0}$ and $(\hat{Z}_m^{(N)})_{m\geq 0}$ are identically distributed. Let τ be the coupling time of the processes $(\hat{Y}_m^{(N)})_{m\geq 0}$ and $(\hat{Z}_m^{(N)})_{m\geq 0}$, i.e.

$$\tau := \inf\{m \geq 0 : \ \hat{Y}_m^{(N)} = \hat{Z}_m^{(N)}\}.$$

If $\inf\{m \geq 0 : \ \hat{Y}_m^{(N)} = \hat{Z}_m^{(N)}\} > N$, we have $\beta_1 = \ldots = \beta_N = 0$. To see this, assume the statement doesn't hold. Then, there is an $l \in \{0, \ldots, N-1\}$ for which $\beta_{l+1} = 1$. It follows that $\hat{Y}_{l+1}^{(N)} = U_{l+1}$. Since $\inf\{m \geq 0 : \ \hat{Y}_m^{(N)} = \hat{Z}_m^{(N)}\} > N$ and since $l < N$, we have $\hat{Y}_l^{(N)} \neq \hat{Z}_l^{(N)}$. Thus, by definition, $\hat{Z}_{l+1}^{(N)} = U_{l+1} = \hat{Y}_{l+1}^{(N)}$, a contradiction.

Therefore,

$$Pr(\tau > N) \leq Pr(\beta_1 = \ldots = \beta_N = 0) = (1 - \bar{c})^N.$$

Next, we derive an upper bound on $tv(\pi\bar{\mathsf{P}}^N, \mu)$. Since μ is invariant with respect to $(\mathsf{P}^t)_{t\geq 0}$, it is also $\bar{\mathsf{P}}$-invariant. Since $\pi\bar{\mathsf{P}}^N$ is the distribution of $\hat{Y}_N^{(N)}$ and since $\mu\bar{\mathsf{P}}^N$ is the distribution of $\hat{Z}_N^{(N)}$, the joint distribution of $(\hat{Y}_N^{(N)}, \hat{Z}_N^{(N)})$ is a coupling of $\pi\bar{\mathsf{P}}^N$ and $\mu\bar{\mathsf{P}}^N$. With (30) and using the definition of $\hat{Z}^{(N)}$, we obtain

$$tv(\pi\bar{\mathsf{P}}^N, \mu) = tv(\pi\bar{\mathsf{P}}^N, \mu\bar{\mathsf{P}}^N) \leq Pr(\hat{Y}_N^{(N)} \neq \hat{Z}_N^{(N)}) = Pr(\tau > N) \leq (1 - \bar{c})^N. \quad (31)$$

With $\alpha = -\frac{\ln(1-\bar{c})}{s}$, we can rewrite (31) as

$$tv(\pi\mathsf{P}^{Ns}, \mu) \leq e^{-\alpha N s}. \quad (32)$$

Since N was arbitrarily chosen, inequality (32) holds for any integer $N \geq 1$. It remains to extend (32) to times that are not of the form Ns for some integer $N \geq 1$. Fix an arbitrary time $t \geq 0$. There is a unique integer $N \geq 1$ such that $(N-1)s \leq t < Ns$. Using Lemma 8 and the semigroup property, we obtain

$$tv(\pi\mathsf{P}^t, \mu) \leq tv(\pi\mathsf{P}^{(N-1)s}, \mu) \leq e^{-\alpha(N-1)s}.$$

Theorem 9 then follows with $c = e^{\alpha s}$. □

In the case of a noncompact M, Harris's ergodic theorem (see [10, Theorem 2.10]), together with Lemma 7, implies a similar result if the semigroup $(\mathsf{P}^t)_{t\geq 0}$ admits a Lyapunov function. Exponential convergence to the invariant measure holds with respect to a weighted version of the total-variation distance, with the weight depending on the Lyapunov function.

3.3 Convergence in Wasserstein distance

Another sufficient condition for exponential convergence in a noncompact setting has been provided by Benaïm, Le Borgne, Malrieu and Zitt in [7, Assumption 1.8]. We present the version in [10], adapted to our more restrictive setting of switching between deterministic trajectories (as opposed to switching between Markov processes).

Given a Polish space (E, d), the Wasserstein distance of two probability measures μ and ν on E is defined by

$$\mathcal{W}_d(\mu, \nu) := \inf_{\Gamma \in C(\mu,\nu)} \int_{E \times E} d(x, y)\ \Gamma(dx, dy),$$

where one should recall from Section 3.2 that $C(\mu, \nu)$ denotes the set of couplings of μ and ν. For a real number $p \geq 1$, the Wasserstein distance of order p is defined as

$$\mathcal{W}_d^{(p)}(\mu, \nu) := \left(\inf_{\Gamma \in C(\mu,\nu)} \int_{E \times E} d(x, y)^p\ \Gamma(dx, dy) \right)^{\frac{1}{p}},$$

see for instance [36, Definition 6.1].

For our switching system, assume that the continuous component X lives on a Polish space (E, d). Let $(\Phi_i)_{i \in S}$ be the flow functions associated to the vector fields $(u_i)_{i \in S}$, and assume that Φ_i^t is globally Lipschitz continuous with Lipschitz constant L_i^t for any $i \in S$ and for any $t \geq 0$. This means that

$$d(\Phi_i^t(\xi), \Phi_i^t(\eta)) \leq L_i^t \cdot d(\xi, \eta)$$

for any $\xi, \eta \in E$. Furthermore, suppose that

$$\kappa_i := \inf_{t > 0} \left(-\frac{\ln(L_i^t)}{t} \right) \tag{33}$$

is a well-defined real number for any $i \in S$. If one assumes, as we do, that the switching rates are independent of the position of X, the stochastic process A on S is Markov and has an invariant measure ν. The condition

$$\sum_{i \in S} \nu(\{i\}) \cdot \kappa_i > 0 \tag{34}$$

then implies convergence to an invariant measure in a certain Wasserstein distance defined in terms of d. Condition (34) can be interpreted as (X, A) contracting in mean, see [10, page 5].

Recall that $\mu_i(\cdot) := \mu(\cdot \times \{i\})$.

Theorem 10 *Under condition* (34)*, there exist an invariant measure* μ *for* $(\mathsf{P}^t)_{t\geq 0}$ *and constants* $c, \alpha, T > 0$ *such that*

$$\mathcal{W}_{\hat{d}}(\delta_{(\xi,i)}\mathsf{P}^t, \mu) \leq ce^{-\alpha t} \cdot (1 + \mathcal{W}_d(\delta_\xi, \mu_i)), \quad t \geq T,$$

where

$$\hat{d}((\xi, i), (\eta, j)) := \mathbb{1}_{i\neq j} + \mathbb{1}_{i=j} \cdot \min\{1, d(\xi, \eta)\}.$$

In [7, Theorem 1.10], the authors establish a slightly different convergence result. Under a moment condition, they show convergence with respect to a mixture of the pth Wasserstein distance for the continuous component X and the total variation distance for the discrete component A.

3.4 Examples

In Example 1, the associated flow functions are $\Phi_1^t(\eta) = \eta e^{-t}$, $\Phi_2^t(\eta) = \eta + t$ and $\Phi_3^t(\eta) = \eta - t$, with global Lipschitz constants $L_1^t = e^{-t}$ and $L_2^t = L_3^t = 1$ for any $t \geq 0$. If we define $\kappa_1, \kappa_2, \kappa_3$ as in (33), we have $\kappa_1 = 1$ and $\kappa_2 = \kappa_3 = 0$. Since we allow switching from any vector field to any other vector field, criterion (34) implies existence of an invariant measure. Theorem 2 implies that the invariant measure is unique and absolutely continuous.

In Example 2, the dynamics are eventually confined to the set $(0, 1)$, so existence of an invariant measure follows using the Krylov–Bogoliubov method. Uniqueness and absolute continuity follow again from Theorem 2. Since u_1 and u_2 point in opposite directions, the strong hypoellipticity condition holds at every point in $(0, 1)$. Theorem 9 then implies exponential convergence to the invariant measure in total-variation distance.

The situation in Example 3 is uniformly elliptic, i.e. the tangent space at any point on the torus is spanned by the vectors obtained through evaluating the vector fields in D at this point. In particular, the weak hypoellipticity condition holds at every point. However, we have already pointed out in Section 2.5 that there is no point on the torus where the strong hypoellipticity condition is satisfied, so Theorem 9 does not apply. Given that the transition probabilities (and thus measures of the form $\delta_{(\xi,i)}\mathsf{P}^t$) are singular in this example, we do not have convergence to the invariant measure in total-variation distance.

A quick computation shows that the flows Φ_1 and Φ_2 associated to the linear vector fields in Example 6 are given by

$$\Phi_1^t(\xi_1,\xi_2) = e^{-at}\begin{pmatrix} 1 & ct \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$$

and

$$\Phi_2^t(\xi_1,\xi_2) = e^{-at}\begin{pmatrix} 1 & 0 \\ -ct & 1 \end{pmatrix}\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}.$$

We have

$$\begin{aligned} \Phi_1^t(0,\xi_2) - \Phi_1^t(0,\eta_2) &= e^{-at}\begin{pmatrix} 1 & ct \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ \xi_2-\eta_2 \end{pmatrix} \\ &= (\xi_2-\eta_2)\cdot e^{-at}\begin{pmatrix} ct \\ 1 \end{pmatrix}. \end{aligned}$$

Taking the Euclidean norm on both sides, we obtain

$$\|\Phi_1^t(0,\xi_2) - \Phi_1^t(0,\eta_2)\|_2 = |\xi_2-\eta_2|\cdot e^{-at}\cdot\sqrt{c^2t^2+1}.$$

Thus,

$$L_1^t \geq e^{-at}\cdot\sqrt{c^2t^2+1} > cte^{-at}.$$

Similarly, one shows that $L_2^t > cte^{-at}$. Then, for $i=1,2$,

$$-\frac{\ln(L_i^t)}{t} \leq a - \frac{\ln(ct)}{t}.$$

The term on the right attains its minimum on $(0, \infty)$ at $t = \frac{e}{c}$. Hence,

$$\kappa_i \leq a - \frac{c}{e}.$$

If $\lambda_1 = \lambda_2 = \lambda$ and if $\frac{c}{a} \geq e$, we have

$$\nu(\{1\}) \cdot \kappa_1 + \nu(\{2\}) \cdot \kappa_2 \leq a - \frac{c}{e} \leq 0,$$

so condition (34) does not hold in this case. Indeed, Lawley, Mattingly and Reed show in [27, Lemma 3.3] that the norm of X_t diverges to ∞ almost surely if the ratio $\frac{c}{a}$ lies above a certain threshold value that depends on the ratio of c and the switching rate λ. See also [8] for an earlier result of this type where some convex combination of the matrices U_1 and U_2 has a positive eigenvalue. In Example 6, for $\lambda \in (0, 1)$, the matrix $\lambda U_1 + (1 - \lambda)U_2$ has eigenvalues $-a - i\sqrt{\lambda \cdot (1 - \lambda)}$ and $-a + i\sqrt{\lambda \cdot (1 - \lambda)}$, which still have negative real part.

CHAPTER IV

REGULARITY OF INVARIANT DENSITIES

In this chapter, we study the regularity theory for switching systems whose continuous component X lives on $\mathbb{R}$. In Section 4.1, we state our main result: Smoothness of the vector fields in D translates into smoothness of invariant densities away from critical points of the vector fields (Theorem 11). In Section 4.2, we state two integral equations that are satisfied by invariant densities. For differentiable densities, these equations can be derived from the Fokker–Planck equations (see Appendix B), but since we intend to use the equations to show differentiability of invariant densities, we need to come up with a proof that does not rely on the Fokker–Planck equations. The integral equation stated in Lemma 9 plays an important role in the proof of Theorem 11. This proof is developed in Section 4.3. The integral equation in Lemma 10 will figure prominently in Chapter 6. Section 4.4 contains the proofs of Lemmas 9 and 10.

Throughout this chapter, we assume that $M = \mathbb{R}$. We assume that the vector fields in D are continuously differentiable and forward-complete. Recall from Section 1.3 that a probability measure μ on $\mathbb{R} \times S$ is an invariant measure with respect to the Markov semigroup $(\mathsf{P}^t)_{t \geq 0}$ if

$$\mu_i(E) = \sum_{j \in S} \int_{\mathbb{R}} \mathsf{P}^t_{\xi,j}(E \times \{i\})\ \mu_j(d\xi)$$

for any Borel set $E \subset \mathbb{R}$, for any $i \in S$ and for any $t \geq 0$. Here, μ_i denotes the marginal $\mu(\cdot \times \{i\})$ on $\mathbb{R}$. Also recall that we use the term "invariant density" for the probability density function of an absolutely continuous invariant measure. An invariant density ρ is an L^1-function on $\mathbb{R} \times S$, and we will usually consider the projections $(\rho_i)_{i \in S}$ that are defined on $\mathbb{R}$ by $\rho_i(\xi) := \rho(\xi, i)$. In what is maybe in abuse of terminology, we refer to these projections as invariant densities of the invariant measure. These invariant densities are then elements of $L^1(\mathbb{R})$ and whenever we state a regularity property of ρ_i, we mean to say that the equivalence class ρ_i has a representative with this regularity property.

We call a point $\xi \in \mathbb{R}$ critical if $u_i(\xi) = 0$ for some $i \in S$. We call $\xi \in \mathbb{R}$ noncritical if $u_i(\xi) \neq 0$ for all $i \in S$ and we call it uniformly critical if $u_i(\xi) = 0$ for all $i \in S$. Throughout this chapter, we assume that the set of critical points of the vector fields in D has no accumulation points. If ξ is a critical point of a vector field u_i for some $i \in S$, we write that u_i is positive to the right of ξ if there is an open interval with left endpoint ξ on which u_i is positive. In this definition, "right" can be replaced with "left" and "positive" with "negative".

4.1 Smoothness of invariant densities at noncritical points

Let μ be an invariant measure of $(\mathsf{P}^t)_{t\geq 0}$ that is absolutely continuous with respect to Lebesgue measure. Let $(\rho_i)_{i\in S}$ denote the invariant densities associated to μ. If n is a positive integer, we call a function $\mathscr{C}^n$ on a set I or at a point ξ if the function is n times continuously differentiable on I or at ξ. Being $\mathscr{C}^0$ means being continuous.

Theorem 11 *Let $\xi \in \mathbb{R}$ be noncritical, and assume that there exist an integer $n \geq 1$ and a closed interval I containing ξ in its interior on which all vector fields in D are $\mathscr{C}^{n+1}$. Then, the invariant densities $(\rho_i)_{i\in S}$ are $\mathscr{C}^n$ at ξ.*

Theorem 11 is proved in Section 4.3. The following statement is an immediate consequence of Theorem 11: If $\xi \in \mathbb{R}$ is noncritical and if all vector fields in D are $\mathscr{C}^\infty$ on a closed interval I containing ξ in its interior, then the invariant densities are $\mathscr{C}^\infty$ at ξ.

In Theorem 11, we assume that there is some absolutely continuous invariant measure that is not necessarily unique. By Theorem 2, the invariant measure of $(\mathsf{P}^t)_{t\geq 0}$ is absolutely continuous (and unique) if there exists a point $\xi \in \mathbb{R}$ that is not uniformly critical and that is D-accessible from any starting point $\eta \in \mathbb{R}$. See Chapter 3 for conditions guaranteeing existence of an invariant measure.

4.2 Integral equations for invariant densities

In this section, we establish two integral equations satisfied almost everywhere by invariant densities of $(\mathsf{P}^t)_{t\geq 0}$. Loosely stated, the equations illustrate how mass with respect to an invariant density ρ_i accumulates at a point $\eta \in \mathbb{R}$. At some point in time, there is a switch from a vector field in $D \setminus \{u_i\}$ to u_i, and the flow associated to u_i transports mass to η. In a sense, we condition on the time and nature of this last switch to u_i. The family of equations in Lemma 9 describe the mass transport for a finite history of the process. In this case, there is a positive probability of having no switch. Lemma 9 will be the basic tool in the proof of Theorem 11. The equation in Lemma 10 describes the transport mechanism for an infinite history. This guarantees that with probability 1, there is at least one switch. Lemma 10 will play an important role in the proofs of Theorem 12 in Section 6.1 and Theorem 13 in Section 6.2.

Let μ be an absolutely continuous invariant measure of $(\mathsf{P}^t)_{t\geq 0}$, with invariant densities $(\rho_i)_{i\in S}$. Since we do not assume backward completeness of the vector fields in D, we have to be careful when studying the history of a switching trajectory. It could happen that the backward flow associated to a vector field goes off to $-\infty$ or ∞ in finite time. For any $\eta \in \mathbb{R}$ and for any $i \in S$, let $\tau_i(\eta)$ denote the supremum over the set of times $t \geq 0$ for which $t \mapsto \Phi_i^{-t}(\eta)$ is well-defined. With this definition, we introduce the shorthand

$$\Phi_i^t \# h(\eta) := \begin{cases} \frac{h(\Phi_i^{-t}(\eta))}{D\Phi_i^t(\Phi_i^{-t}(\eta))}, & t < \tau_i(\eta) \\ 0, & t \geq \tau_i(\eta) \end{cases} \tag{35}$$

for the pushforward of the function h under the flow map Φ_i^t. We think of h as a density function on the real line. Note that $D\Phi_i^t > 0$ in dimension one, so there is no need for absolute value in the denominator. Since u_i is assumed to be $\mathscr{C}^1$, so is

$\eta \mapsto \Phi_i^t(\eta)$, and the differential $D\Phi_i^t$ is well-defined.

Let $L^1_+(\mathbb{R})$ be the set of L^1-functions on the real line that have a nonnegative representative. In other words, $L^1_+(\mathbb{R})$ is the space of densities for finite measures on $\mathbb{R}$. For any $h \in L^1_+(\mathbb{R})$ and for any $T > 0$, define the Perron–Frobenius operators

$$\bar{\mathsf{P}}_i^T h(\eta) := \frac{1}{T} \cdot \int_0^T e^{-\lambda_i t} \cdot \Phi_i^t \# h(\eta)\; dt \tag{36}$$

and

$$\hat{\mathsf{P}}_i^T h(\eta) := \frac{1}{T} \int_0^T e^{-\lambda_i t} \cdot (T-t) \cdot \Phi_i^t \# h(\eta)\; dt. \tag{37}$$

We can now state the truncated version of the integral equation.

Lemma 9 *For any $i \in S$ and for any $T > 0$,*

$$\rho_i \equiv \bar{\mathsf{P}}_i^T \rho_i + \sum_{j \neq i} \lambda_{j,i} \cdot \hat{\mathsf{P}}_i^T \rho_j.$$

To state the integral equation over an infinite time horizon, we define the operators

$$\bar{\mathsf{P}}_i h(\eta) := \int_0^\infty e^{-\lambda_i t} \cdot \Phi_i^t \# h(\eta)\; dt, \quad i \in S$$

for densities $h \in L^1_+(\mathbb{R})$.

Lemma 10 *For any $i \in S$,*

$$\rho_i = \sum_{j \neq i} \lambda_{j,i} \cdot \bar{\mathsf{P}}_i \rho_j.$$

Lemmas 9 and 10 are proved in Section 4.4. As will become apparent from these proofs, the lemmas continue to hold if the state space $\mathbb{R}$ of the continuous component X is replaced with a finite-dimensional smooth manifold.

4.3 Proof of Theorem 11

In this section, we prove Theorem 11, which was stated in Section 4.1. By assumption, there exist an integer $n \geq 1$ and a closed interval I with ξ in its interior. Since ξ is noncritical and since for each vector field in D, the set of critical points has no accumulation point, we may assume without loss of generality that I does not contain any critical points. Let $I_0 \subset I$ be another compact interval containing ξ in its interior, whose endpoints are a positive distance away from the endpoints of I. As the trajectories of the X-component of the switching process have bounded speed on compact subsets of $\mathbb{R}$, there is a small time $T_0 > 0$ such that $(\Phi_{\mathbf{i}}^{\mathbf{s}})^{-1}(\eta) \in I$ for any finite index sequence $\mathbf{i}$, any corresponding sequence of nonnegative switching times $\mathbf{s}$ with sum of components less than or equal to T_0 and for any $\eta \in I_0$.

We define the integration kernels

$$\mathcal{K}_i(\zeta, \eta) := \frac{\exp\left(\lambda_i \cdot \int_\eta^\zeta \frac{dx}{u_i(x)}\right)}{u_i(\eta)} \tag{38}$$

and

$$\hat{\mathcal{K}}_i^{T_0}(\zeta, \eta) := \left(T_0 + \int_\eta^\zeta \frac{dx}{u_i(x)}\right) \cdot \mathcal{K}_i(\zeta, \eta) \tag{39}$$

for $i \in S$ and $(\zeta, \eta) \in I \times I_0$. With these definitions, we have the following representations of $\bar{\mathsf{P}}_i^{T_0}\rho_i$ and $\hat{\mathsf{P}}_i^{T_0}\rho_j$. See (36) and (37) for the definitions of $\bar{\mathsf{P}}_i^{T_0}$ and $\hat{\mathsf{P}}_i^{T_0}$.

Lemma 11 *For any $i \in S$ and for any $\eta \in I_0$,*

$$\bar{\mathsf{P}}_i^{T_0}\rho_i(\eta) = \frac{1}{T_0} \cdot \int_{\Phi_i^{-T_0}(\eta)}^{\eta} \rho_i(\zeta) \cdot \mathcal{K}_i(\zeta, \eta)\, d\zeta. \tag{40}$$

For any $i, j \in S$, $i \neq j$, and for any $\eta \in I_0$,

$$\hat{\mathsf{P}}_i^{T_0}\rho_j(\eta) = \frac{1}{T_0} \cdot \int_{\Phi_i^{-T_0}(\eta)}^{\eta} \rho_j(\zeta) \cdot \hat{\mathcal{K}}_i^{T_0}(\zeta, \eta)\, d\zeta. \tag{41}$$

Our definition of T_0 implies that the interval $[\Phi_i^{-T_0}(\eta), \eta]$ (or $[\eta, \Phi_i^{-T_0}(\eta)]$ if $u_i(\xi) < 0$) is contained in I, so the integrals on the right are well-defined. Notice in particular that this reasoning still holds if u_i is not backward complete.

Proof: Fix an $\eta \in I_0$ and recall the definition of $\Phi_i^t \# \rho_i$ in (35). Linearity of the Jacobi flow gives

$$D\Phi_i^t(\Phi_i^{-t}(\eta)) = \frac{u_i(\eta)}{u_i(\Phi_i^{-t}(\eta))},$$

hence

$$\Phi_i^t \# \rho_i(\eta) = \rho_i(\Phi_i^{-t}(\eta)) \cdot \frac{u_i(\Phi_i^{-t}(\eta))}{u_i(\eta)}$$

for any $t \in [0, T_0]$. The change of variables $\zeta = \Phi_i^{-t}(\eta)$ then yields (40). Formula (41) is proved similarly. $\square$

In (40) and (41), the expressions on the right still make sense if $\mathcal{K}_i$ and $\hat{\mathcal{K}}_i^{T_0}$ are replaced with arbitrary kernels on $I \times I_0$. For any such kernel $\mathcal{H}$ and for any $i, j \in S$, set

$$\mathcal{H}_i^{T_0} \rho_j(\eta) := \frac{1}{T_0} \cdot \int_{\Phi_i^{-T_0}(\eta)}^{\eta} \rho_j(\zeta) \cdot \mathcal{H}(\zeta, \eta) \, d\zeta. \tag{42}$$

The following lemma addresses regularity of the integration kernels $(\mathcal{K}_i)_{i \in S}$ and $(\hat{\mathcal{K}}_i^{T_0})_{i \in S}$.

Lemma 12 *The kernels $(\mathcal{K}_i)_{i \in S}$ and $(\hat{\mathcal{K}}_i^{T_0})_{i \in S}$ are $\mathscr{C}^{n+1}$ on $I \times I_0$.*

Proof: This follows from our assumption that u_i is $\mathscr{C}^{n+1}$ and nonzero on I. $\square$

The following lemmas illustrate the smoothing effect of the operators $(\bar{\mathsf{P}}_i^{T_0})_{i \in S}$ and $(\hat{\mathsf{P}}_i^{T_0})_{i \in S}$. We begin by showing that, away from critical points, the densities $(\rho_i)_{i \in S}$ are bounded.

Lemma 13 *The densities $(\rho_i)_{i \in S}$ are bounded on the interval I_0.*

Proof: Fix an $i \in S$. By Lemma 9, it is enough to show that $\bar{\mathsf{P}}_i^{T_0} \rho_i$ and $(\hat{\mathsf{P}}_i^{T_0} \rho_j)_{j \neq i}$ are bounded on I_0. Since $\mathcal{K}_i$ and $\hat{\mathcal{K}}_i^{T_0}$ are $\mathscr{C}^1$ on the compact set $I \times I_0$ (Lemma 12), they are also bounded on $I \times I_0$ by constants k_i and k_{i,T_0}. For $j \in S$, let $\|\rho_j\|_1$ denote the L^1-norm of ρ_j on $\mathbb{R}$. Using integral representation (40),

$$\bar{\mathsf{P}}_i^{T_0} \rho_i(\eta) \leq \frac{k_i}{T_0} \cdot \int_{\Phi_i^{-T_0}(\eta)}^{\eta} \rho_i(\zeta) \, d\zeta \leq \frac{k_i \cdot \|\rho_i\|_1}{T_0}$$

for any $\eta \in I_0$. And using integral representation (41),

$$\hat{\mathsf{P}}_i^{T_0} \rho_j(\eta) \leq \frac{k_{i,T_0}}{T_0} \cdot \int_{\Phi_i^{-T_0}(\eta)}^{\eta} \rho_j(\zeta) \, d\zeta \leq \frac{k_{i,T_0} \cdot \|\rho_j\|_1}{T_0}$$

for any $j \neq i$, $\eta \in I_0$. $\square$

Remark 1 In the proof of Lemma 13, we did not use any concrete information about $\mathcal{K}_i$ or $\hat{\mathcal{K}}_i^{T_0}$ other than boundedness on $I \times I_0$. The result still holds if $\mathcal{K}_i$ and $\hat{\mathcal{K}}_i^{T_0}$ are replaced with arbitrary kernels that are bounded on $I \times I_0$. Besides, the time T_0 can be replaced with any time $T \in (0, T_0)$.

The following corollary will be useful in Section 6.3 when we derive asymptotics for invariant densities at critical points.

Corollary 1 *Let $i \in S$ and assume that $\xi \in \mathbb{R}$ is not a critical point of u_i. Then, there is a compact interval I with ξ in its interior, such that ρ_i is bounded on I.*

In Lemma 13, we assumed that ξ is noncritical. Here, the point ξ may be critical for some of the vector fields in D, just not for the particular vector field u_i whose corresponding density function we are interested in.

PROOF: Since $u_i(\xi) \neq 0$ and since the set of critical points of u_i has no accumulation points, there is a compact interval I that has ξ in its interior, but does not contain any critical points of u_i. Let $I_0 \subset I$ be another compact interval with ξ in its interior such that the endpoints of I_0 are a positive distance away from the endpoints of I. Choose $T > 0$ so small that $\Phi_i^{-t}(\eta) \in I$ for any $\eta \in I_0$ and for any $t \in [0, T]$. Define the kernels $\mathcal{K}_i$ and $\hat{\mathcal{K}}_i^T$ according to (38) and (39). These kernels are bounded on $I \times I_0$, and we can repeat the proof of Lemma 13 to finish the argument. $\square$

Let $I_1 \subset I_0$ be a compact interval that contains ξ in its interior and whose endpoints are a positive distance away from the endpoints of I_0. Let $T_1 \in (0, T_0]$ be so small that $(\Phi_{\mathbf{i}}^{\mathbf{s}})^{-1}(\eta) \in I_0$ for any index sequence $\mathbf{i}$, any corresponding sequence of nonnegative switching times $\mathbf{s}$ with l^1-norm less than or equal to T_1, and for any $\eta \in I_1$.

Lemma 14 *The densities $(\rho_i)_{i \in S}$ are Lipschitz continuous on I_1.*

Proof: Fix an $i \in S$. By Lemma 9, it is enough to show that $\bar{\mathsf{P}}_i^{T_1}\rho_i$ and $(\hat{\mathsf{P}}_i^{T_1}\rho_j)_{j\neq i}$ are Lipschitz continuous on I_1. By Lemma 13, ρ_i is bounded on I_0 by some constant r_i. Let L be a Lipschitz constant of $\mathcal{K}_i$ on $I \times I_0$ and let $\tilde{L}$ be a Lipschitz constant of the flow function Φ_i on $[-T_1, 0] \times I_1$. The constant k_i is defined as in the proof of Lemma 13 and k_{i,T_1} is defined in analogy to k_{i,T_0}. Fix two points $\eta, \vartheta \in I_1$. As $\Phi_i^{-T_1}(\eta)$ and $\Phi_t^{-T_1}(\vartheta)$ are both contained in I_0, we obtain the estimate

$$\begin{aligned}
&|\bar{\mathsf{P}}_i^{T_1}\rho_i(\eta) - \bar{\mathsf{P}}_i^{T_1}\rho_i(\vartheta)| \\
=&\frac{1}{T_1} \cdot \left| \int_{\Phi_i^{-T_1}(\eta)}^{\eta} \rho_i(\zeta) \cdot \mathcal{K}_i(\zeta,\eta)\, d\zeta - \int_{\Phi_i^{-T_1}(\vartheta)}^{\vartheta} \rho_i(\zeta) \cdot \mathcal{K}_i(\zeta,\vartheta)\, d\zeta \right| \\
\leq&\frac{1}{T_1} \cdot \left(\left| \int_{\Phi_i^{-T_1}(\eta)}^{\Phi_i^{-T_1}(\vartheta)} \rho_i(\zeta) \cdot \mathcal{K}_i(\zeta,\eta)\, d\zeta \right| + \left| \int_{\eta}^{\vartheta} \rho_i(\zeta) \cdot \mathcal{K}_i(\zeta,\vartheta)\, d\zeta \right| \right. \\
&\left. + \left| \int_{\Phi_i^{-T_1}(\vartheta)}^{\eta} \rho_i(\zeta) \cdot (\mathcal{K}_i(\zeta,\eta) - \mathcal{K}_i(\zeta,\vartheta))\, d\zeta \right| \right) \\
\leq&\|\vartheta - \eta\| \cdot \frac{1}{T_1} \cdot (r_i k_i \cdot (1 + \tilde{L}) + L \cdot \|\rho_i\|_1).
\end{aligned}$$

Let $\hat{L}$ be a Lipschitz constant of $\hat{\mathcal{K}}_i^{T_1}$ on $I \times I_0$. For a fixed $j \neq i$, the density ρ_j is bounded on I_0 by a constant r_j, and

$$|\hat{\mathsf{P}}_i^{T_1}\rho_j(\eta) - \hat{\mathsf{P}}_i^{T_1}\rho_j(\vartheta)| \leq |\vartheta - \eta| \cdot \frac{1}{T_1} \cdot (r_j k_{i,T_1} \cdot (1 + \tilde{L}) + \hat{L} \cdot \|\rho_j\|_1). \tag{43}$$

□

Remark 2 Lemma 14 continues to hold if $\mathcal{K}_i$ and $\hat{\mathcal{K}}_i^{T_1}$ are replaced with arbitrary kernels that are Lipschitz continuous on $I \times I_0$ and if T_1 is replaced with an arbitrary time $T \in (0, T_1)$.

Remark 3 Lemma 14 implies the following: If an open interval I does not contain any critical points, then all invariant densities ρ_i are Lipschitz continuous on I. Slightly modifying the proof of Lemma 14, one can show a related statement: If an open interval I does not contain any critical points of a particular vector field u_i (but possibly critical points of other vector fields), the invariant density ρ_i is continuous on I.

Notice that we can only guarantee continuity, not Lipschitz continuity, of ρ_i. Since we allow for critical points of the other vector fields $(u_j)_{j\neq i}$ on I, we can no longer ascertain boundedness of the corresponding densities $(\rho_j)_{j\neq i}$. Instead of (43), we obtain the weaker estimate

$$\begin{aligned}|\hat{\mathsf{P}}_i^{T_1}\rho_j(\eta) - \hat{\mathsf{P}}_i^{T_1}\rho_j(\vartheta)| \leq &\frac{k_{i,T_1}}{T_1} \cdot \left(\left| \int_{\Phi_i^{-T_1}(\vartheta)}^{\Phi_i^{-T_1}(\eta)} \rho_j(\zeta)\ d\zeta \right| + \left| \int_{\vartheta}^{\eta} \rho_j(\zeta)\ d\zeta \right| \right) \\ &+ \frac{\|\rho_j\|_1}{T_1} \cdot \hat{L} \cdot |\vartheta - \eta|.\end{aligned}$$

Lemma 15 illustrates the actual smoothing mechanism.

Lemma 15 *For any integer $k \in \{0, \ldots, n-1\}$, the following statement holds. If the densities $(\rho_i)_{i\in S}$ are $\mathscr{C}^k$ on a compact interval $\tilde{I} \subset I_1$ that contains ξ in its interior, there exist a compact interval $\tilde{I}' \subset \tilde{I}$ with ξ in its interior and a time $T \in (0, T_1]$ such that for any $\mathscr{C}^{k+2}$-kernel $\mathcal{H}$ on $I \times I_0$, the functions $(\mathcal{H}_i^T\rho_j)_{i,j\in S}$ are $\mathscr{C}^{k+1}$ on $\tilde{I}'$.*

Recall that we defined $\mathcal{H}_i^T\rho_j$ in (42).

Proof: We prove Lemma 15 by induction on k. In the base case, assume that the densities $(\rho_i)_{i\in S}$ are continuous on $\tilde{I} \subset I_1$. Let $\tilde{I}' \subset \tilde{I}$ be a compact interval that contains ξ in its interior and whose endpoints are a positive distance away from the endpoints of $\tilde{I}$. Let $T \in (0, T_1]$ be so small that $(\Phi_{\mathbf{i}}^{\mathbf{s}})^{-1}(\eta) \in \tilde{I}$ for any index sequence $\mathbf{i}$, any corresponding sequences of nonnegative switching times $\mathbf{s}$ with l^1-norm less than or equal to T, and for any $\eta \in \tilde{I}'$. For any $\mathscr{C}^2$-kernel $\mathcal{H}$ on $I \times I_0$, for any $\eta \in \tilde{I}'$

and for any $i, j \in S$,

$$
\begin{aligned}
\frac{d}{d\eta}\mathcal{H}_i^T\rho_j(\eta) =& \frac{1}{T}\cdot\left(\rho_j(\eta)\cdot\mathcal{H}(\eta,\eta) - \rho_j(\Phi_i^{-T}(\eta))\cdot\mathcal{H}(\Phi_i^{-T}(\eta),\eta)\cdot\frac{d}{d\eta}\Phi_i^{-T}(\eta)\right)\\
&+\frac{1}{T}\cdot\int_{\Phi_i^{-T}(\eta)}^{\eta}\rho_j(\zeta)\cdot\partial_2\mathcal{H}(\zeta,\eta)\,d\zeta\\
=&\frac{1}{T}\cdot\left(\rho_j(\eta)\cdot\mathcal{H}(\eta,\eta) - \rho_j(\Phi_i^{-T}(\eta))\cdot\mathcal{H}(\Phi_i^{-T}(\eta),\eta)\cdot\frac{d}{d\eta}\Phi_i^{-T}(\eta)\right)\\
&+(\partial_2\mathcal{H})_i^T\rho_j(\eta).
\end{aligned}
\tag{44}
$$

Here, $\partial_2\mathcal{H}$ denotes the partial derivative of $\mathcal{H}$ with respect to its second component. Since ρ_j is assumed to be $\mathscr{C}^0$ on $\tilde{I}$, since $\mathcal{H}$ is $\mathscr{C}^2$ on $I\times I_0$ and since u_i is $\mathscr{C}^{n+1}$ on I, the first term in (44) is $\mathscr{C}^0$ on $\tilde{I}'$.

It remains to show that $(\partial_2\mathcal{H})_i^T\rho_j$ is $\mathscr{C}^0$, but this follows along the lines of Lemma 14, keeping in mind that $\partial_2\mathcal{H}$ is Lipschitz continuous on $I\times I_0$ and that $T\le T_1$ (see Remark 2). Since $\frac{d}{d\eta}\mathcal{H}_i^T\rho_j(\eta)$ is $\mathscr{C}^0$ on $\tilde{I}'$, it follows that $\mathcal{H}_i^T\rho_j$ is $\mathscr{C}^1$ on $\tilde{I}'$. This completes the base case.

In the induction step, let k be a fixed integer in $\{1,\dots,n-1\}$ and assume that the statement holds for $k-1$. Assume that the densities $(\rho_i)_{i\in S}$ are $\mathscr{C}^k$ on $\tilde{I}\subset I_1$. The densities $(\rho_i)_{i\in S}$ are then also $\mathscr{C}^{k-1}$ on $\tilde{I}$. By induction hypothesis, there exist a compact interval $\tilde{I}'\subset\tilde{I}$ with ξ in its interior and a time $T\in(0,T_1]$ such that for any $\mathscr{C}^{k+1}$-kernel $\mathcal{H}$ on $I\times I_0$, the functions $(\mathcal{H}_i^T\rho_j)_{i,j\in S}$ are $\mathscr{C}^k$ on $\tilde{I}'$. Without loss of generality, we can assume that the endpoints of $\tilde{I}'$ are a positive distance away from the endpoints of $\tilde{I}$ and that T is so small that $(\Phi_{\mathbf{i}}^{\mathbf{s}})^{-1}(\eta)\in\tilde{I}$ for any index sequence $\mathbf{i}$, any corresponding sequence of nonnegative switching times $\mathbf{s}$ with l^1-norm less than or equal to T, and for any $\eta\in\tilde{I}'$. Let $\mathcal{H}$ be a $\mathscr{C}^{k+2}$-kernel on $I\times I_0$. Then, (44) holds for any $\eta\in\tilde{I}'$ and for any $i,j\in S$.

Since ρ_j is by assumption $\mathscr{C}^k$ on $\tilde{I}$, since $\mathcal{H}$ is $\mathscr{C}^{k+2}$ on $I\times I_0$ and since u_i is $\mathscr{C}^{n+1}$ on I, the first term in (44) is $\mathscr{C}^k$ on $\tilde{I}'$. In addition, $\partial_2\mathcal{H}$ is a $\mathscr{C}^{k+1}$-kernel on $I\times I_0$. By induction hypothesis, $(\partial_2\mathcal{H})_i^T\rho_j$ is $\mathscr{C}^k$ on $\tilde{I}'$, so $\frac{d}{d\eta}\mathcal{H}_i^T\rho_j(\eta)$ is $\mathscr{C}^k$ on $\tilde{I}'$. From this, it follows that $\mathcal{H}_i^T \rho_j$ is $\mathscr{C}^{k+1}$ on $\tilde{I}'$. $\square$

Proof of Theorem 11: In order to prove Theorem 11, it suffices to show the following statement:

For any integer $k \in \{0, \ldots, n\}$, there is a compact interval I_{k+1} with ξ in its interior such that the densities $(\rho_i)_{i \in S}$ are $\mathscr{C}^k$ on I_{k+1}.

We prove this statement by induction on k. By Lemma 14, the densities $(\rho_i)_{i \in S}$ are Lipschitz continuous on I_1. This takes care of the base case. In the induction step, let k be an integer in $\{1, \ldots, n\}$ and assume that the densities $(\rho_i)_{i \in S}$ are $\mathscr{C}^{k-1}$ on a compact interval $I_k \subset I_1$ with ξ in its interior. By Lemma 15, there exist a compact interval $I_{k+1} \subset I_k$ with ξ in its interior and a time $T \in (0, T_1]$ such that for any $\mathscr{C}^{k+1}$-kernel $\mathcal{H}$ on $I \times I_0$, the functions $(\mathcal{H}_i^T \rho_j)_{i,j \in S}$ are $\mathscr{C}^k$ on I_{k+1}. Fix an $i \in S$. Lemma 15 applied to the integration kernel $\mathcal{K}_i$ yields that $\bar{\mathsf{P}}_i^T \rho_i$ is $\mathscr{C}^k$ on I_{k+1}. And applying Lemma 15 to $\hat{\mathcal{K}}_i^T$ yields that for any $j \neq i$, $\hat{\mathsf{P}}_i^T \rho_j$ is $\mathscr{C}^k$ on I_{k+1}. By Lemma 9, ρ_i is $\mathscr{C}^k$ on I_{k+1}. $\square$

In Malliavin calculus and many other areas of mathematics, regularity statements for densities and for functions in general are typically proved using integration by parts (see for instance [5]). Such an approach might also work for invariant densities of switching systems, as suggested by Jonathan Mattingly. For a one-dimensional continuous component, the main ingredients in this approach are Stroock's lemma (see [31, Lemma 3.1]) and the following control-theory lemma that we state without proof.

Lemma 16 *For any $\xi \in \mathbb{R}$, there exist a neighborhood V of ξ, a time $t > 0$ and an $\epsilon > 0$ such that the following holds: For any index sequence $\mathbf{i}$ of finite length m that includes indices i and j with $u_i(\xi) \neq u_j(\xi)$, there is an open set $W \subset \mathbb{R}^{m-1}$ with the following features.*

(a) The closure of $\Delta_{t,m-1}$ is contained in W.

(b) For any $\eta \in V$, the differential of

$$f^\eta : \mathbb{R}^{m-1} \to \mathbb{R}, (s_1, \ldots, s_{m-1}) \mapsto \Phi_{\mathbf{i}}^{(s_1,\ldots,s_{m-1},t-\sum_{l=1}^{m-1} s_l)}(\eta)$$

has full rank in W.

(c) For any $\mathbf{s} \in W$ and for any $\eta \in V$,

$$\|\nabla f^\eta(\mathbf{s})\| \geq \epsilon.$$

Lemma 16 provides conditions under which the map assigning to each sequence of switching times with fixed terminal time t the corresponding terminal point on $\mathbb{R}$ is guaranteed to be locally regular. It is very similar to the first part of Theorem 5 by Chow, Jurdjevic–Sussmann and Krener in the case of a one-dimensional manifold. Notice in particular that $u_i(\xi) \neq u_j(\xi)$ for some $i, j \in S$ means that the strong hypoellipticity condition holds at ξ. In addition to the conclusion of Theorem 5, we obtain a lower bound on the norm of the gradient of f^η that is uniform in the switching sequence.

4.4 Proof of Lemmas 9 and 10

For $t \geq 0$, as defined in (2) let $\mu\mathsf{P}^t$ denote the distribution of $(X, A)_t$ starting from the distribution μ. conditioned on the initial distribution μ. Since μ is invariant under $(\mathsf{P}^t)_{t\geq 0}$,

$$\mu_i(\cdot) = \int_{T^{(1)}}^{T^{(2)}} \pi(t) \cdot \mu\mathsf{P}^t(\cdot \times \{i\})\, dt \tag{45}$$

for any $T^{(1)} < T^{(2)}$ in $[0, \infty]$ and for any probability density $\pi(t)$ on $(T^{(1)}, T^{(2)})$. We will expand the expression on the right with respect to the sequences of driving vector fields and will ultimately see how ρ_i gets transformed through the action of the Markov semigroup and through time-averaging.

The following formula is the key to Lemmas 9 and 10.

Lemma 17 *Let $E \subset \mathbb{R}$ be a Borel set and let $i \in S$. For any $T^{(1)} < T^{(2)}$ in $[0, \infty]$ and for any probability density $\pi(t)$ on $(T^{(1)}, T^{(2)})$,*

$$\mu_i(E) = \int_E \Big(\int_{T^{(1)}}^{T^{(2)}} \pi(t) e^{-\lambda_i t} \, \Phi_i^t \# \rho_i(\eta) \, dt$$
$$+ \sum_{j \neq i} \lambda_{j,i} \Big(\int_0^{T^{(1)}} e^{-\lambda_i t} \, \Phi_i^t \# \rho_j(\eta) \, dt + \int_{T^{(1)}}^{T^{(2)}} c(t) e^{-\lambda_i t} \, \Phi_i^t \# \rho_j(\eta) \, dt \Big) \Big) \, d\eta,$$

where $c(t) := \int_0^{T^{(2)} - t} \pi(s + t) \, ds$.

Given this representation for μ_i, we first show Lemma 9 and then Lemma 10. Finally, we prove the representation itself.

4.4.1 Proof of Lemma 9

When we set $T^{(1)} := 0$, $T^{(2)} := T$ and $\pi(t) := \frac{1}{T}$, the identity in Lemma 17 becomes

$$\mu_i(E) = \int_E \Big(\frac{1}{T} \int_0^T e^{-\lambda_i t} \, \Phi_i^t \# \rho_i(\eta) \, dt + \sum_{j \neq i} \lambda_{j,i} \frac{1}{T} \int_0^T (T - t) e^{-\lambda_i t} \, \Phi_i^t \# \rho_j(\eta) \, dt \Big) \, d\eta.$$

This implies Lemma 9.

4.4.2 Proof of Lemma 10

When we set $T^{(1)} := T$ for some time $T > 0$, $T^{(2)} := \infty$ and $\pi(t) := e^{T-t}$, the identity in Lemma 17 becomes

$$\mu_i(E) = \int_E \Big(\int_T^\infty e^{T-t} e^{-\lambda_i t} \, \Phi_i^t \# \rho_i(\eta) \, dt$$
$$+ \sum_{j \neq i} \lambda_{j,i} \Big(\int_0^T e^{-\lambda_i t} \, \Phi_i^t \# \rho_j(\eta) \, dt + \int_T^\infty e^{T-t} e^{-\lambda_i t} \, \Phi_i^t \# \rho_j(\eta) \, dt \Big) \Big) \, d\eta.$$

Since μ is a probability measure,

$$\begin{aligned} \int_E \int_T^\infty e^{T-t} e^{-\lambda_i t} \, \Phi_i^t \# \rho_i(\eta) \, dt \, d\eta = & \int_T^\infty e^{T-t} e^{-\lambda_i t} \mu_i((\Phi_i^t)^{-1}(E)) \, dt \\ \leq & e^{-\lambda_i T} \int_T^\infty e^{T-t} \, dt = e^{-\lambda_i T}, \end{aligned}$$

where one should observe that the set $(\Phi_i^t)^{-1}(E)$ is well-defined even if $\Phi_i^{-t}(\eta)$ is undefined for some $\eta \in E$. Similarly,

$$\sum_{j\neq i} \lambda_{j,i} \int_E \int_T^\infty e^{T-t} e^{-\lambda_i t}\, \Phi_i^t \# \rho_j(\eta)\, dt\, d\eta \leq \sum_{j\neq i} \lambda_{j,i} e^{-\lambda_i T}.$$

Letting T go to infinity, we obtain

$$\mu_i(E) = \int_E \sum_{j\neq i} \lambda_{j,i} \int_0^\infty e^{-\lambda_i t}\, \Phi_i^t \# \rho_j(\eta)\, dt\, d\eta$$

and Lemma 10 follows.

4.4.3 Proof of Lemma 17

Fix an $i \in S$. We introduce some notation. For any $t > 0$ and for any index sequence $\mathbf{i}$ with terminal index i, let $C_{\mathbf{i}}^t$ denote the event that the driving vector fields up to time t appear in the order given by $\mathbf{i}$. For any index sequence $\mathbf{i} = (i_1, \ldots, i_{m-1}, i)$ of length $m \geq 2$, let $\mathsf{P}_{\mathbf{i}}$ be the probability that the first m driving vector fields appear in the order given by $\mathbf{i}$, conditioned on u_{i_1} being the first driving vector field. For $T > 0$ and $m \in \mathbb{N}$, we define the simplex $\Delta_{T,m}$ as the interior of the convex hull of the vectors $0, Te_1, \ldots, Te_m$ in $\mathbb{R}^m$. For any vector v with m components, no matter whether v is a vector of indices, switching times or switching rates, let $v^{(m-1)}$ denote the projection of v onto its first $(m-1)$ coordinates. Moreover, let $\|v\|_1$ be the sum of the coordinates of v and let $\langle \cdot, \cdot \rangle$ denote the Euclidean inner product on the space that fits the context (usually $\mathbb{R}^{m-1}$ or $\mathbb{R}^m$).

Lemma 18 *For any $T^{(1)} < T^{(2)}$ in $[0, \infty]$ and for any function $\pi(t)$ that is nonnegative and integrable on $(T^{(1)}, T^{(2)})$,*

$$\int_{T^{(1)}}^{T^{(2)}} \pi(t) \mathsf{P}_{\xi,i}(C_{(i)}^t) \int_{\mathbb{R}} \mathsf{P}_{\xi,i}(X_t \in E | C_{(i)}^t)\, \mu_i(d\xi)\, dt$$
$$= \int_E \int_{T^{(1)}}^{T^{(2)}} \pi(t) e^{-\lambda_i t}\, \Phi_i^t \# \rho_i(\eta)\, dt\, d\eta.$$

Proof: This is immediate. $\square$

Lemma 19 *For any index sequence* $\mathbf{i} = (i_1, \ldots, i_{m-1}, i)$ *of length* $m \geq 2$*, for any* $T^{(1)} < T^{(2)}$ *in* $[0, \infty]$ *and for any function* $\pi(t)$ *that is nonnegative and integrable on* $(T^{(1)}, T^{(2)})$*,*

$$\int_{T^{(1)}}^{T^{(2)}} \pi(t) \mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}}) \int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_t \in E | C^t_{\mathbf{i}})\, \mu_{i_1}(d\xi)\, dt$$
$$= \mathsf{P}_{\mathbf{i}} \prod_{l=1}^{m-1} \lambda_{i_l} \int_{\Delta_{T^{(2)},m} \setminus \Delta_{T^{(1)},m}} \pi(\|\mathbf{s}\|_1) e^{-\langle \lambda^{(m-1)}, \mathbf{s}^{(m-1)} \rangle} e^{-\lambda_i s_m} \mu_{i_1}((\Phi^{\mathbf{s}}_{\mathbf{i}})^{-1}(E))\, d\mathbf{s},$$

where $\lambda^{(m-1)} := (\lambda_{i_1}, \ldots, \lambda_{i_{m-1}})^T$*.*

Proof: Fix an index sequence $\mathbf{i} = (i_1, \ldots, i_{m-1}, i)$ of length m, $T^{(1)} < T^{(2)} \in [0, \infty]$ and a nonnegative integrable function π on $(T^{(1)}, T^{(2)})$. Let $T_1, \ldots, T_m$ be independent, exponentially distributed random variables such that T_l has parameter λ_{i_l} for $1 \leq l \leq m-1$ and T_m has parameter λ_i. For any $t \geq T^{(1)}$,

$$\int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_t \in E | C^t_{\mathbf{i}})\, \mu_{i_1}(d\xi)$$
$$= \frac{1}{\mathsf{P}(R^t_{\mathbf{i}})} \int_{\mathbb{R}} \mathsf{P}\left(\Phi_{\mathbf{i}}^{(T_1, \ldots, T_{m-1}, t - \sum_{l=1}^{m-1} T_l)}(\xi) \in E, R^t_{\mathbf{i}}\right)\ \mu_{i_1}(d\xi), \quad (46)$$

where

$$R^t_{\mathbf{i}} := \left\{ \sum_{l=1}^{m-1} T_l < t \leq \sum_{l=1}^{m} T_l \right\}.$$

As a notational shorthand, we introduce the functions

$$f^{\xi}_{t,\mathbf{i}} :\ \mathbb{R}^{m-1} \to \mathbb{R},\ (s_1, \ldots, s_{m-1}) \mapsto \Phi_{\mathbf{i}}^{(s_1, \ldots, s_{m-1}, t - \sum_{l=1}^{m-1} s_l)}(\xi).$$

Then,

$$\mathsf{P}\left(\Phi_{\mathbf{i}}^{(T_1, \ldots, T_{m-1}, t - \sum_{l=1}^{m-1} T_l)}(\xi) \in E, R^t_{\mathbf{i}}\right)$$
$$= \int_{\Delta_{t,m-1}} \mathbb{1}_{\{f^{\xi}_{t,\mathbf{i}}(\mathbf{s}) \in E\}}(\mathbf{s}) \prod_{l=1}^{m-1} \lambda_{i_l} e^{-\lambda_{i_l} s_l} e^{-\lambda_i (t - \|\mathbf{s}\|_1)}\, d\mathbf{s},$$

which implies that (46) can be written as

$$\frac{1}{\mathsf{P}(R^t_{\mathbf{i}})} \int_{\mathbb{R}} \int_{\Delta_{t,m-1}} \mathbb{1}_{\{f^{\xi}_{t,\mathbf{i}}(\mathbf{s}) \in E\}}(\mathbf{s}) \prod_{l=1}^{m-1} \lambda_{i_l} e^{-\lambda_{i_l} s_l} e^{-\lambda_i (t - \|\mathbf{s}\|_1)}\, d\mathbf{s}\, \mu_{i_1}(d\xi). \quad (47)$$

Interchanging the order of integration, (47) becomes

$$\frac{1}{\mathsf{P}(R^t_{\mathbf{i}})}\int_{\Delta_{t,m-1}}\prod_{l=1}^{m-1}\lambda_{i_l}e^{-\lambda_{i_l}s_l-\lambda_i(t-\|\mathbf{s}\|_1)}\mu_{i_1}((f^{\cdot}_{t,\mathbf{i}}(\mathbf{s}))^{-1}(E))\,d\mathbf{s}.$$

We have thus shown that

$$\int_{T^{(1)}}^{T^{(2)}}\pi(t)\mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}})\int_{\mathbb{R}}\mathsf{P}_{\xi,i_1}(X_t\in E|C^t_{\mathbf{i}})\,\mu_{i_1}(d\xi)\,dt=\int_{T^{(1)}}^{T^{(2)}}\pi(t)\frac{\mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}})}{\mathsf{P}(R^t_{\mathbf{i}})}$$
$$\cdot\int_{\Delta_{t,m-1}}\prod_{l=1}^{m-1}\lambda_{i_l}e^{-\lambda_{i_l}s_l-\lambda_i(t-\|\mathbf{s}\|_1)}\mu_{i_1}((f^{\cdot}_{t,\mathbf{i}}(\mathbf{s}))^{-1}(E))\,d\mathbf{s}\,dt.$$

The term $\frac{\mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}})}{\mathsf{P}(R^t_{\mathbf{i}})}$ gives the probability that the first m driving vector fields appear according to index sequence $\mathbf{i}$, conditioned on u_{i_1} being the first driving vector field. It is therefore equal to $\mathsf{P}_{\mathbf{i}}$. Interchanging the order of integration and substituting $s_m=t-\|\mathbf{s}\|_1$, Lemma 19 follows. □

Lemma 20 *For any index sequence $\mathbf{i}=(i_1,\ldots,i_{m-1},i)$ of length $m\geq 2$, for any $T^{(1)}<T^{(2)}$ in $[0,\infty]$ and for any function $\pi(t)$ that is nonnegative and integrable on $(T^{(1)},T^{(2)})$,*

$$\int_{T^{(1)}}^{T^{(2)}}\pi(t)\mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}})\int_{\mathbb{R}}\mathsf{P}_{\xi,i_1}(X_t\in E|C^t_{\mathbf{i}})\,\mu_{i_1}(d\xi)\,dt=$$
$$\int_{\Delta_{T^{(2)},2}\setminus\Delta_{T^{(1)},2}}\lambda_{i_{m-1},i}e^{-\lambda_i t}\pi(s+t)\mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}^{(m-1)}})$$
$$\cdot\int_{\mathbb{R}}\mathsf{P}_{\xi,i_1}(X_s\in(\Phi^t_i)^{-1}(E)|C^s_{\mathbf{i}^{(m-1)}})\,\mu_{i_1}(d\xi)\,d(s,t).$$

Proof: For notational compactness, we momentarily introduce the notation $\tilde{\Delta}_i(m,t):=\Delta_{T^{(i)}-t,m-1}$. By Tonelli's theorem, the term to the right of the equality sign in Lemma 19 can be written as

$$\int_0^{T^{(1)}} \mathsf{P}_{\mathbf{i}} \prod_{l=1}^{m-1} \lambda_{i_l} \int_{(\tilde\Delta_2\setminus\tilde\Delta_1)(m,t)} \pi(\|\mathbf{s}\|_1+t)e^{-\langle\lambda^{(m-1)},\mathbf{s}\rangle-\lambda_i t}\mu_{i_1}((\Phi_{\mathbf{i}}^{(\mathbf{s},t)})^{-1}(E))\,d\mathbf{s}\,dt$$
$$+\int_{T^{(1)}}^{T^{(2)}} \mathsf{P}_{\mathbf{i}} \prod_{l=1}^{m-1} \lambda_{i_l} \int_{\tilde\Delta_2(m,t)} \pi(\|\mathbf{s}\|_1+t)e^{-\langle\lambda^{(m-1)},\mathbf{s}\rangle}e^{-\lambda_i t}$$
$$\cdot\,\mu_{i_1}((\Phi_{\mathbf{i}}^{(\mathbf{s},t)})^{-1}(E))\,d\mathbf{s}\,dt$$
$$=\int_0^{T^{(1)}} \lambda_{i_{m-1},i}e^{-\lambda_i t}\mathsf{P}_{\mathbf{i}^{(m-1)}} \prod_{l=1}^{m-2} \lambda_{i_l} \int_{(\tilde\Delta_2\setminus\tilde\Delta_1)(m,t)} \pi_t(\|\mathbf{s}\|_1)e^{-\langle\lambda^{(m-1)},\mathbf{s}\rangle}$$
$$\cdot\,\mu_{i_1}((\Phi_{\mathbf{i}^{(m-1)}}^{\mathbf{s}})^{-1}((\Phi_i^t)^{-1}(E)))\,d\mathbf{s}\,dt$$
$$+\int_{T^{(1)}}^{T^{(2)}} \lambda_{i_{m-1},i}e^{-\lambda_i t}\mathsf{P}_{\mathbf{i}^{(m-1)}} \prod_{l=1}^{m-2} \lambda_{i_l} \int_{\tilde\Delta_2(m,t)} \pi_t(\|\mathbf{s}\|_1)e^{-\langle\lambda^{(m-1)},\mathbf{s}\rangle}$$
$$\cdot\,\mu_{i_1}((\Phi_{\mathbf{i}^{(m-1)}}^{\mathbf{s}})^{-1}((\Phi_i^t)^{-1}(E)))\,d\mathbf{s}\,dt,$$

where the function $\pi_t(s) := \pi(s+t)$ is nonnegative and integrable on $(T^{(1)}-t, T^{(2)}-t)$ if $t < T^{(1)}$, and is nonnegative and integrable on $(0, T^{(2)}-t)$ if $t > T^{(1)}$.

By another application of Lemma 19 (if $m > 2$) or of Lemma 18 (if $m = 2$), the previous term becomes

$$\int_0^{T^{(1)}} \lambda_{i_{m-1},i}e^{-\lambda_i t}\int_{T^{(1)}-t}^{T^{(2)}-t} \pi_t(s)\mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}^{(m-1)}})$$
$$\cdot\int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s\in(\Phi_i^t)^{-1}(E)|C^s_{\mathbf{i}^{(m-1)}})\,\mu_{i_1}(d\xi)\,ds\,dt$$
$$+\int_{T^{(1)}}^{T^{(2)}} \lambda_{i_{m-1},i}e^{-\lambda_i t}\int_0^{T^{(2)}-t} \pi_t(s)\mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}^{(m-1)}})$$
$$\cdot\int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s\in(\Phi_i^t)^{-1}(E)|C^s_{\mathbf{i}^{(m-1)}})\,\mu_{i_1}(d\xi)\,ds\,dt$$
$$=\int_{\Delta_{T^{(2)},2}\setminus\Delta_{T^{(1)},2}} \lambda_{i_{m-1},i}e^{-\lambda_i t}\pi(s+t)\mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}^{(m-1)}})$$
$$\cdot\int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s\in(\Phi_i^t)^{-1}(E)|C^s_{\mathbf{i}^{(m-1)}})\,\mu_{i_1}(d\xi)\,d(s,t).$$

□

Proof of Lemma 17: Fix a Borel set E, $T^{(1)} < T^{(2)} \in [0,\infty]$ and a probability density π on $(T^{(1)}, T^{(2)})$. Expanding the term to the right of the equality sign in (45) by conditioning on the sequences of driving vector fields, we obtain

$$\mu_i(E) = \int_{T^{(1)}}^{T^{(2)}} \pi(t)\mathsf{P}_{\xi,i}(C^t_{(i)}) \int_{\mathbb{R}} \mathsf{P}_{\xi,i}(X_t \in E | C^t_{(i)})\, \mu_i(d\xi)\, dt$$

$$+ \sum_{\mathbf{i}:|\mathbf{i}|\geq 2}^{(i)} \int_{T^{(1)}}^{T^{(2)}} \pi(t)\mathsf{P}_{\xi,i_1}(C^t_{\mathbf{i}}) \int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_t \in E | C^t_{\mathbf{i}})\, \mu_{i_1}(d\xi)\, dt, \tag{48}$$

where $\sum^{(i)}_{\mathbf{i}:|\mathbf{i}|\geq 2}$ extends over all index sequences $\mathbf{i} = (i_1, \ldots, i_{m-1}, i)$ with terminal index i and length ≥ 2.

By Lemma 18, it is enough to show that the term in (48) equals

$$\sum_{j\neq i} \lambda_{j,i} \int_E \Big(\int_0^{T^{(1)}} e^{-\lambda_i t}\, \Phi^t_i \# \rho_j(\eta)\, dt + \int_{T^{(1)}}^{T^{(2)}} c(t) e^{-\lambda_i t} \Phi^t_i \# \rho_j(\eta)\, dt \Big)\, d\eta,$$

where $c(t)$ is defined as in Lemma 17. For any $m \geq 2$, let $\sum^{(i)}_{\mathbf{i}:|\mathbf{i}|=m}$ be the sum over all index sequences of length m with terminal index i. For any $j \in S$, let $\sum^{(j)}_{\mathbf{i}}$ be the sum over all index sequences $\mathbf{i}$ with terminal index j. By Lemma 20, the term in (48) can be written as

$$\sum_{m=2}^{\infty} \sum_{\mathbf{i}:|\mathbf{i}|=m}^{(i)} \int_{\Delta_{T^{(2)},2} \setminus \Delta_{T^{(1)},2}} \lambda_{i_{m-1},i} e^{-\lambda_i t} \pi(s+t) \mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}^{(m-1)}})$$

$$\cdot \int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s \in (\Phi^t_i)^{-1}(E) | C^s_{\mathbf{i}^{(m-1)}})\, \mu_{i_1}(d\xi)\, d(s,t)$$

$$= \sum_{j\neq i} \int_{\Delta_{T^{(2)},2} \setminus \Delta_{T^{(1)},2}} \lambda_{j,i} e^{-\lambda_i t} \pi(s+t)$$

$$\cdot \sum_{\mathbf{i}}^{(j)} \mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}}) \int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s \in (\Phi^t_i)^{-1}(E) | C^s_{\mathbf{i}})\ \mu_{i_1}(d\xi)\, d(s,t). \tag{49}$$

Moreover, for any fixed s,

$$\sum_{\mathbf{i}}^{(j)} \mathsf{P}_{\xi,i_1}(C^s_{\mathbf{i}}) \int_{\mathbb{R}} \mathsf{P}_{\xi,i_1}(X_s \in (\Phi^t_i)^{-1}(E) | C^s_{\mathbf{i}})\ \mu_{i_1}(d\xi) = \mu \mathsf{P}^s((\Phi^t_i)^{-1}(E) \times \{j\}).$$

Since μ is invariant, $\mu\mathsf{P}^s((\Phi^t_i)^{-1}(E) \times \{j\})$ equals $\mu_j((\Phi^t_i)^{-1}(E))$ and is thus independent of s.

As a result, the right side of (49) equals

$$\sum_{j\neq i}\int_{\Delta_{T^{(2)},2}\setminus\Delta_{T^{(1)},2}} \lambda_{j,i}e^{-\lambda_i t}\pi(s+t)\mu_j((\Phi_i^t)^{-1}(E))\ d(s,t)$$
$$= \sum_{j\neq i}\int_0^{T^{(1)}} \lambda_{j,i}e^{-\lambda_i t}\mu_j((\Phi_i^t)^{-1}(E))\int_{T^{(1)}-t}^{T^{(2)}-t}\pi(s+t)\ ds\,dt$$
$$+\sum_{j\neq i}\int_{T^{(1)}}^{T^{(2)}} \lambda_{j,i}e^{-\lambda_i t}\mu_j((\Phi_i^t)^{-1}(E))\int_0^{T^{(2)}-t}\pi(s+t)\,ds\,dt,$$

and Lemma 17 follows. □

CHAPTER V

THE SUPPORT OF INVARIANT MEASURES ON $\mathbb{R}$

Let μ be an invariant measure of the Markov semigroup $(\mathsf{P}^t)_{t\geq 0}$. In this chapter, we describe the support of the measures $(\mu_i)_{i\in S}$, which are measures on the real line. In Section 5.1, we introduce the notion of a minimal invariant set and present an algorithm that allows us to identify the minimal invariant sets of a switching system with one-dimensional continuous component. The only information required is the critical points and signs of the driving vector fields. In Section 5.2, we relate minimal invariant sets to the support of invariant measures and ultimately arrive at a description of the support in terms of minimal invariant sets.

A point $\xi \in \mathbb{R}$ lies in the support of μ_i if and only if $\mu_i(U) > 0$ for any open neighborhood U of ξ. Recall from Section 2.1 that a point $\xi \in \mathbb{R}$ is called D-reachable from a point $\eta \in \mathbb{R}$ if there exist a finite index sequence $\mathbf{i}$ and a corresponding sequence of nonnegative switching times $\mathbf{t}$ such that

$$\Phi_{\mathbf{i}}^{\mathbf{t}}(\eta) = \xi.$$

For any $\xi \in \mathbb{R}$, we define $L(\xi)$ as the set of points that are D-reachable from ξ. We call a point $\xi \in \mathbb{R}$ D-accessible from $\eta \in \mathbb{R}$ if for any open neighborhood U of ξ there exist a finite index sequence $\mathbf{i}$ and a corresponding sequence of nonnegative switching times $\mathbf{t}$ such that

$$\Phi_{\mathbf{i}}^{\mathbf{t}}(\eta) \in U.$$

Let L denote the set of points on the real line that are D-accessible from any point in $\mathbb{R}$.

5.1 *Minimal invariant sets*

A nonempty set $I \subset \mathbb{R}$ is called invariant if

$$\Phi_{\mathbf{i}}^{\mathbf{t}}(\xi) \in I$$

for any $\xi \in I$, any finite index sequence $\mathbf{i}$ and any corresponding sequence of nonnegative switching times $\mathbf{t}$. A minimal invariant set is an invariant set for which any nonempty strict subset is not invariant. Alternatively, a minimal invariant set is a nonempty set I with the property that

$$L(\xi) = I \tag{50}$$

for any $\xi \in I$.

The following algorithm yields exactly the minimal invariant sets of our switching system.

1. Mark $-\infty$ with the label "l" and mark $+\infty$ with the label "r".

2. Mark those critical points where all vector fields in D are nonnegative with an "l" and mark those critical points where all vector fields in D are nonpositive with an "r". If a critical point has both labels "l" and "r", it is uniformly critical. All uniformly critical points form minimal invariant sets.

3. Consider all points, including $-\infty$, with the label "l". This includes those points that carry both labels. As $+\infty$ doesn't have label "l", each of these points has a closest labeled point to its right. If this point has label "r", the open, possibly infinite, interval with the "l"-labeled and the "r"-labeled points as its endpoints is a candidate for a minimal invariant set. It is indeed a minimal invariant set if and only if it contains two not necessarily distinct points ξ and η for which there are vector fields $u, v \in D$ with $u(\xi) > 0$ and $v(\eta) < 0$.

Proposition 1 *The algorithm above characterizes the minimal invariant sets of the switching system completely. Minimal invariant sets are thus either open intervals or point sets with exactly one element.*

PROOF: We first show that any set identified by the algorithm as a minimal invariant set is indeed a minimal invariant set. Let $\mathcal{S}$ be a set identified by the algorithm as a minimal invariant set. Then, either $\mathcal{S} = \{\xi\}$, where ξ is a uniformly critical point, or $\mathcal{S}$ is an open interval (l, r), where $l < r$ are elements of the extended real line such that

(a) $l = -\infty$ or $u_i(l) \geq 0$ for any $i \in S$

(b) $r = \infty$ or $u_i(r) \leq 0$ for any $i \in S$

(c) for any critical point ξ in (l, r) there exist indices $i, j \in S$ with $u_i(\xi) < 0 < u_j(\xi)$

(d) if there are no critical points in (l, r), there are at least points $\xi, \eta \in (l, r)$ and indices $i, j \in S$ with $u_i(\xi) < 0 < u_j(\eta)$.

If $\mathcal{S} = \{\xi\}$, it is clear that $\mathcal{S}$ is a minimal invariant set: The only strict subset of $\mathcal{S}$ is the empty set, and $\mathcal{S}$ is invariant because ξ is uniformly critical.

If $\mathcal{S} = (l, r)$, no switching trajectory starting in $\mathcal{S}$ can get to the left of l or to the right of r. This is obvious if $l = -\infty$ or $r = \infty$. If l or r are finite, it is guaranteed by Conditions a and b, respectively. Hence, $\mathcal{S}$ is invariant. Next, we show that $\mathcal{S}$ is also minimal. Assume that $\mathcal{S}$ is not minimal. Then, there is a nonempty strict subset $\mathcal{R}$ of $\mathcal{S}$ that is invariant. In addition, there is a point $\zeta \in \mathcal{S}$ with $u_i(\zeta) \leq 0$ for any $i \in S$. To see this, fix a point $\eta \in \mathcal{S} \setminus \mathcal{R}$ and a point $\xi \in \mathcal{R}$. We can assume without loss of generality that $\eta > \xi$. Since $\mathcal{R}$ is invariant, η is not D-reachable from ξ. Thus, there is a point $\zeta \in [\xi, \eta]$ with $u_i(\zeta) \leq 0$ for any $i \in S$. In light of Condition c, ζ is not critical. On the other hand, Condition d implies that there exist a $\tilde{\zeta} \in \mathcal{S}$ and a

$j \in S$ with $u_j(\tilde{\zeta}) > 0$. Assume without loss of generality that $\tilde{\zeta} > \zeta$, and let

$$\hat{\zeta} := \sup\{\theta \in [\zeta, \tilde{\zeta}] : \; u_i(\theta) < 0 \; \forall i \in S\}.$$

The point $\hat{\zeta}$ is a critical point in $\mathcal{S}$ with $u_i(\hat{\zeta}) \leq 0$ for any $i \in S$. This violates Condition c.

Conversely, let I be a minimal invariant set. We need to show that the algorithm correctly identifies I as a minimal invariant set. Due to the minimality assumption, I is an interval. If I contains exactly one point, this point is uniformly critical, for otherwise, I would not be invariant.

If I has at least two elements, it is an interval with distinct endpoints. We show that if an endpoint of I is finite, it must be a critical point: Let ξ be a finite endpoint of I, say its left endpoint, and assume that ξ is noncritical. Since I is invariant, $u_i(\xi) > 0$ for any $i \in S$. By continuity of the vector fields, there is an $\epsilon > 0$ such that $u_i(\eta) > 0$ for any $i \in S$ and for any $\eta \in [\xi, \xi + \epsilon]$. By choosing ϵ sufficiently small, we can then ensure that $I \setminus [\xi, \xi + \epsilon]$ is a nonempty strict subset of I that is invariant. This contradicts the minimality assumption on I. Invariance of I also implies that the endpoints of I are not D-reachable from a starting point in the interior of I. Hence, I is an open interval (l, r), where l and r may be finite or infinite.

It remains to show that Conditions c and d are satisfied. Let $\xi \in I$ be a critical point. If $u_i(\xi) \geq 0$ for any $i \in S$, the interval $(\xi, r) \subset (l, r)$ is invariant as well – a contradiction. Similarly, $u_i(\xi)$ cannot be nonpositive for all $i \in S$, so we can find $i, j \in S$ with $u_i(\xi) < 0 < u_j(\xi)$. To show that Condition d holds, assume that $u_i(\eta) \geq 0$ for all $\eta \in I$ and for all $i \in S$. For $\xi \in I$, the interval (ξ, r) is invariant, which contradicts the minimality assumption. $\square$

Proposition 2 *Minimal invariant sets are pairwise disjoint.*

Proof: Let I and J be minimal invariant sets with $I \cap J \neq \emptyset$. As the intersection of invariant sets, $I \cap J$ is invariant. Since I and J are minimal, it follows that $I = I \cap J = J$. □

5.2 How minimal invariant sets relate to the support of invariant measures

We begin by relating invariant measures of the global dynamics on $\mathbb{R} \times S$ to invariant measures of the switching dynamics confined to a minimal invariant set.

Let $I \subset \mathbb{R}$ be a minimal invariant set. On $I \times S$, we define the semigroup $(\mathsf{p}^t)_{t \geq 0}$ by

$$\mathsf{p}^t_{\xi,i}(E \times \{j\}) := \mathsf{P}^t_{\xi,i}(E \times \{j\})$$

for any $(\xi, i) \in I \times S$, for any set E in the Borel σ-algebra on I and for any $j \in S$. Hence, $(\mathsf{p}^t)_{t \geq 0}$ can be thought of as the restriction of $(\mathsf{P}^t)_{t \geq 0}$ to $I \times S$. It is well-defined because I is invariant.

Proposition 3 *There is a one-to-one correspondence between invariant measures of $(\mathsf{p}^t)_{t \geq 0}$ and those invariant measures of $(\mathsf{P}^t)_{t \geq 0}$ that assign mass 1 to $I \times S$.*

Proof: Let ν be an invariant measure for $(\mathsf{p}^t)_{t \geq 0}$. By setting

$$\mu(E \times \{j\}) := \nu(E \cap I \times \{j\})$$

for Borel sets $E \subset \mathbb{R}$ and $j \in S$, we define a probability measure μ on $\mathbb{R} \times S$. This measure is invariant with respect to $(\mathsf{P}^t)_{t \geq 0}$, for

$$\begin{aligned}
\mu(E \times \{j\}) &= \nu(E \cap I \times \{j\}) \\
&= \sum_{i \in S} \int_I \mathsf{p}^t_{\xi,i}(E \cap I \times \{j\})\ \nu(d\xi \times \{i\}) \\
&= \sum_{i \in S} \int_I \mathsf{P}^t_{\xi,i}(E \cap I \times \{j\})\ \nu(d\xi \times \{i\}) \\
&= \sum_{i \in S} \int_I \mathsf{P}^t_{\xi,i}(E \times \{j\})\ \mu(d\xi \times \{i\})
\end{aligned}$$

by invariance of ν and I.

Conversely, let μ be invariant with respect to $(\mathsf{P}^t)_{t\geq 0}$ with $\mu(I \times S) = 1$. We define the measure ν as the restriction of μ to $I \times S$. As $\mu(I \times S) = 1$, ν is then also a probability measure, and for Borel sets $E \subset I$,

$$\begin{aligned}\nu(E \times \{j\}) &= \mu(E \times \{j\}) \\ &= \sum_{i\in S} \int_{\mathbb{R}} \mathsf{P}^t_{\xi,i}(E \times \{j\})\ \mu(d\xi \times \{i\}) \\ &= \sum_{i\in S} \int_{I} \mathsf{P}^t_{\xi,i}(E \times \{j\})\ \mu(d\xi \times \{i\}) \\ &= \sum_{i\in S} \int_{I} \mathsf{p}^t_{\xi,i}(E \times \{j\})\ \nu(d\xi \times \{i\}).\end{aligned}$$

□

Next, we show that the support of the measure $\mu(\cdot \times S)$ does not contain points outside of the closure of minimal invariant sets.

Proposition 4 *Let $\xi \in \mathbb{R}$ be a point that is not contained in the closure of a minimal invariant set. Then, ξ is not contained in the support of $\mu(\cdot \times S)$.*

To prove Proposition 4, we need to establish several lemmas. We begin with a simple criterion for membership in a minimal invariant set.

Lemma 21 *A point $\xi \in \mathbb{R}$ does not belong to any minimal invariant set if and only if there is a point $\eta \in L(\xi)$ with $\xi \notin L(\eta)$.*

PROOF: By characterization (50) of minimal invariant sets, a point $\xi \in \mathbb{R}$ is contained in a minimal invariant set if and only if $L(\xi)$ is a minimal invariant set. This is in turn equivalent to

$$L(\xi) = \bigcap_{\eta\in L(\xi)} L(\eta).$$

The inclusion $\bigcap_{\eta\in L(\xi)} L(\eta) \subset L(\xi)$ always holds, so a point $\xi \in \mathbb{R}$ is *not* contained in a minimal invariant set if and only if there exist points $\eta, \zeta \in L(\xi)$ with $\zeta \notin L(\eta)$.

If such points η and ζ exist, the point ξ is not contained in $L(\eta)$ either because ζ is D-reachable from ξ. And if there is a point $\eta \in L(\xi)$ with $\xi \notin L(\eta)$, we can choose $\zeta = \xi$. $\square$

Recall that for $T > 0$ and for a positive integer m, $\Delta_{T,m-1}$ denotes the simplex

$$\left\{ \mathbf{s} \in (0,\infty)^{m-1} : \sum_{l=1}^{m-1} s_l < T \right\}.$$

Lemma 22 *Let $\xi \in \mathbb{R}$ be a point that is not contained in the closure of a minimal invariant set. Then, there exist an open interval I containing ξ, an open set $U \subset \mathbb{R}$, a time $T > 0$, an index sequence $\mathbf{i} = (i_1, \ldots, i_m)$ and an open set $\Delta \subset \Delta_{T,m-1}$ such that*

$$\Phi_{\mathbf{i}}^{(\mathbf{s}, T - \sum_{l=1}^{m-1} s_l)}(\eta) \in U, \quad \eta \in I, \mathbf{s} \in \Delta, \tag{51}$$

and

$$I \cap L(\vartheta) = \emptyset, \quad \vartheta \in U. \tag{52}$$

Proof: Since ξ is not contained in the closure of a minimal invariant set, there is an $\epsilon > 0$ such that

(i) none of the points in $[\xi - \epsilon, \xi + \epsilon]$ belong to a minimal invariant set,

(ii) there is an $i \in S$ such that $u_i(\eta) \neq 0$ for all $\eta \in [\xi - \epsilon, \xi + \epsilon]$.

Property (ii) follows from the fact that uniformly critical points form minimal invariant sets. Assume without loss of generality that $u_i(\eta) > 0$ for all $\eta \in [\xi - \epsilon, \xi + \epsilon]$. This implies that the right endpoint $\xi + \epsilon$ is D-reachable from any point in $[\xi - \epsilon, \xi + \epsilon]$. Since $\xi + \epsilon$ is not contained in a minimal invariant set, Lemma 21 implies existence of a $\zeta \in L(\xi + \epsilon)$ with $\xi + \epsilon \notin L(\zeta)$. Then, $\eta \notin L(\zeta)$ for all $\eta \in [\xi - \epsilon, \xi + \epsilon]$.

Let U denote the interior of $L(\zeta)$. Since ζ is D-reachable from $\xi + \epsilon$, it is not uniformly critical. Therefore, U is not the empty set. Moreover, $U \subset L(\eta)$ for any $\eta \in [\xi - \epsilon, \xi + \epsilon]$ and $L(\vartheta) \cap [\xi - \epsilon, \xi + \epsilon] = \emptyset$ for all $\vartheta \in U$. Fix an arbitrary point

$\hat{\vartheta} \in U$. Then, $\hat{\vartheta}$ is D-reachable, i.e. there exist an index sequence $\mathbf{i} = (i_1, \ldots, i_m)$ and a corresponding time sequence $\mathbf{t}$ with $\hat{\vartheta} = \Phi_{\mathbf{i}}^{\mathbf{t}}(\xi)$. Let T be the sum of all components of $\mathbf{t}$. Since the map

$$(\eta, \mathbf{s}) \mapsto \Phi_{\mathbf{i}}^{\mathbf{s}, T - \sum_{l=1}^{m-1} s_l)}(\eta)$$

is continuous on $[\xi - \epsilon, \xi + \epsilon] \times \Delta_{T,m-1}$, there exist an open interval I containing ξ and an open set $\Delta \subset \Delta_{T,m-1}$ such that (51) holds. We can assume without loss of generality that $I \subset [\xi - \epsilon, \xi + \epsilon]$. Then, (52) holds as well. $\square$

Next, we record a simple consequence of the fact that the speed of the process X is bounded on bounded sets.

Lemma 23 *Let $I \subset \mathbb{R}$ be a nonempty and bounded open set. Then, there exist constants $\epsilon', c' > 0$ and an open set $I' \subset I$ such that*

$$\inf_{\eta \in I', i, j \in S} \mathsf{P}_{\eta,i}^{\epsilon'}(I \times \{j\}) \geq c'. \tag{53}$$

See [6] for a proof of Lemma 23.

Proof of Proposition 4: To derive a contradiction, we assume that ξ belongs to the support of $\mu(\cdot \times S)$. By Lemma 22, there exist an open interval I containing ξ, an open set $U \subset \mathbb{R}$, a time $T > 0$, an index sequence $\mathbf{i} = (i_1, \ldots, i_m)$ and an open set $\Delta \subset \Delta_{T,m-1}$ such that (51) and (52) hold.

By Lemma 23, there are constants $\epsilon', c' > 0$ and an open set $I' \subset I$ such that (53) is satisfied. As I' is an open interval containing ξ,

$$c := \mu(I' \times S) > 0.$$

Therefore,

$$\begin{aligned}
\sum_{i \in S} \int_{I'} \mathsf{P}_{\theta,i}^{T+\epsilon'}(U \times S)\ \mu_i(d\theta) =& \sum_{i \in S} \int_{I'} \sum_{l \in S} \int_{\mathbb{R}} \mathsf{P}_{\eta,l}^{T}(U \times S) \cdot \mathsf{P}_{\theta,i}^{\epsilon'}(d\eta \times \{l\})\ \mu_i(d\theta) \\
\geq& \sum_{i \in S} \int_{I'} \int_{I} \mathsf{P}_{\eta,i_1}^{T}(U \times S) \cdot \mathsf{P}_{\theta,i}^{\epsilon'}(d\eta \times \{i_1\})\ \mu_i(d\theta) \\
\geq& c \cdot c' \cdot \inf_{\eta \in I} \mathsf{P}_{\eta,i_1}^{T}(U \times S).
\end{aligned}$$

Next, we show that $\inf_{\eta \in I} \mathsf{P}^T_{\eta,i_1}(U \times S) > 0$. Fix a point $\eta \in I$ and let $C_{\mathbf{i}}$ denote the event that the driving vector fields up to time T appear in the order given by $\mathbf{i}$. Let $\mathsf{P}_{\mathbf{i}}$ be the probability that the first m driving vector fields appear in the order given by $\mathbf{i}$, conditioned on u_{i_1} being the first driving vector field. Similarly to Lemma 19 in Section 4.4, we have

$$\begin{aligned} \mathsf{P}^T_{\eta,i_1}(U \times S) \geq & \mathsf{P}_{\eta,i_1}(X_T \in U, C_{\mathbf{i}}) \\ \geq & \mathsf{P}_{\mathbf{i}} \cdot \int_\Delta \prod_{l=1}^{m-1} \lambda_{i_l} \cdot e^{-\lambda_{i_l} s_l} \cdot e^{-\lambda_{i_m}(T-(s_1+\ldots+s_{m-1}))} d\mathbf{s}. \end{aligned} \tag{54}$$

The term in (54) is strictly positive and does not depend on η. We conclude that

$$a := \sum_{i \in S} \int_{I'} \mathsf{P}^{T+\epsilon'}_{\theta,i}(U \times S)\ \mu_i(d\theta) > 0.$$

Hence, there is a positive integer N with $N \cdot a > 1$. Let $\mu\mathsf{P}$ denote the distribution of the Markov process (X, A) with initial distribution μ. For $0 \leq k \leq N-1$, define the event

$$E_k := \{X_{k\cdot(T+\epsilon')} \in I', X_{(k+1)\cdot(T+\epsilon')} \in U, X_{j\cdot(T+\epsilon')} \in I'^c \text{ for } k+2 \leq j \leq N\}.$$

Since the sets $(E_k)_{0 \leq k \leq N-1}$ are pairwise disjoint,

$$\mu\mathsf{P}(X_{N\cdot(T+\epsilon')} \in I'^c) \geq \sum_{k=0}^{N-1} \mu\mathsf{P}(E_k).$$

Since I' cannot be reached from any point in U, we have

$$E_k = \{X_{k\cdot(T+\epsilon')} \in I', X_{(k+1)\cdot(T+\epsilon')} \in U\}.$$

Then, for $0 \leq k \leq N-1$,

$$\mu\mathsf{P}(E_k) = \sum_{i \in S} \int_{\mathbb{R}} \sum_{j \in S} \int_{I'} \mathsf{P}^{T+\epsilon'}_{\eta,j}(U \times S) \cdot \mathsf{P}^{k\cdot(T+\epsilon')}_{\theta,i}(d\eta \times \{j\})\ \mu_i(d\theta) = a$$

because μ is an invariant measure. We infer that

$$\mu\mathsf{P}(X_{N\cdot(T+\epsilon')} \in I'^c) \geq N \cdot a > 1,$$

which is impossible. Hence, ξ is not contained in the support of $\mu(\cdot \times S)$. $\square$

In Proposition 3, we saw that invariant measures on minimal invariant sets correspond to invariant measures on $\mathbb{R}$ that are supported on a minimal invariant set. In the following proposition, we show uniqueness of the invariant measure on a given minimal invariant set.

Proposition 5 *Any minimal invariant set admits at most one invariant measure.*

Proof: Let I be a minimal invariant set. If $I = \{\xi\}$ for some uniformly critical point ξ, uniqueness of the invariant measure is clear.

If I is an open interval, it does not contain any uniformly critical points by Proposition 1. By the alternative characterization of minimal invariant sets in (50), $I = L(\eta)$ for any $\eta \in I$. Thus, any point in I is D-reachable from all starting points in I. By [1, Theorem 1], this implies uniqueness of the invariant measure of the restricted semigroup $(\mathsf{p}^t)_{t\geq 0}$. $\square$

Now, assume that the invariant measure μ is ergodic. If I is a minimal invariant set, ergodicity of μ implies that $\mu(I \times S)$ is either 0 or 1. It is then natural to ask whether we can assign a unique minimal invariant set I to μ for which $\mu(I \times S) = 1$. The following proposition shows that this can be done.

Proposition 6 *If μ is ergodic, there is a unique minimal invariant set I with $\mu(I \times S) = 1$.*

Proof: Let us begin by showing that such a minimal invariant set exists. Since μ is ergodic, it is enough to show that $\mu(I \times S) > 0$ for some minimal invariant set I. We denote the set of points not contained in the closure of a minimal invariant set by $\mathcal{T}$. According to Proposition 4, the intersection of $\mathcal{T}$ and of the support of $\mu(\cdot \times S)$ is empty, so there is a point $\xi \in \mathcal{T}^c$ that also lies in the support of $\mu(\cdot \times S)$. As $\xi \in \mathcal{T}^c$, there is a minimal invariant set I whose closure contains ξ. We distinguish between several cases.

First, assume that $I = \{\xi\}$. Then, ξ is uniformly critical and may or may not be an endpoint of one or two additional minimal invariant sets. If there are no minimal invariant sets adjacent to $\{\xi\}$, we can find an open neighborhood U of ξ such that $U \setminus \{\xi\} \subset \mathcal{T}$. Since the complement of the support of $\mu(\cdot \times S)$ has measure 0, it follows that $\mu(U \setminus \{\xi\} \times S) = 0$. Therefore, $\mu(\{\xi\} \times S) > 0$.

If there is at least one open minimal invariant set adjacent to $\{\xi\}$, we have $\mu(\{\xi\} \times S) > 0$, or at least one of the adjacent minimal invariant sets has strictly positive $\mu(\cdot \times S)$-measure.

Now, assume that $I = (l, r)$. If $\xi \in I$, it is immediate from the definition of the support that $\mu(I \times S) > 0$. If ξ is an endpoint of I, assume without loss of generality that $\xi = l$. We have already dealt with the case where ξ is uniformly critical. If ξ is critical but not uniformly critical, we still have $\mu(\{\xi\} \times S) > 0$ or $\mu(I \times S) > 0$ or $\mu(J \times S) > 0$, provided that J is an open minimal invariant set with ξ as its right endpoint. We only need to exclude the case that $\mu(\{\xi\} \times S) > 0$. This can be done similarly to the proof of Proposition 4.

Uniqueness of the minimal invariant set follows from Proposition 2. $\square$

Proposition 7 *If μ is ergodic, there is a unique minimal invariant set I such that the support of the measures $(\mu_i)_{i \in S}$ equals the closure of I.*

PROOF: Let I be the unique minimal invariant set for which $\mu(I \times S) = 1$ and whose existence is postulated in Proposition 6. By characterization (50) of minimal invariant sets, every point in I is D-reachable from any other point in I. This implies that I is contained in the support of $(\mu_i)_{i \in S}$. With $\mu(I \times S) = 1$, the statement follows. $\square$

Corollary 2 *For any minimal invariant set I, there is at most one invariant measure μ with $\mu(I \times S) = 1$.*

Proof: This follows from Propositions 3 and 5. $\square$

Corollary 3 *Let μ be an invariant measure, not necessarily ergodic. Then, there exist minimal invariant sets $I_1, \ldots, I_N$ such that the support of μ_j equals the closure of $\bigcup_{i=1}^{N} I_i$ for any $j \in S$. If μ is absolutely continuous, each of the minimal invariant sets I_i is an open interval.*

Proof: This follows from Proposition 7 and from the Ergodic Decomposition Theorem, see, e.g., [19]. $\square$

In particular, the support of μ_j only depends on the invariant measure μ and not on the index j. If μ is absolutely continuous, all of the minimal invariant sets I_i in Corollary 3 are open intervals.

Finally, we show that invariant densities are positive in the interior of the support of $(\mu_i)_{i \in S}$. We will need this result in Chapter 6. Suppose that μ is absolutely continuous with respect to the product of Lebesgue measure on $\mathbb{R}$ and counting measure on S, and let $(\rho_i)_{i \in S}$ be the invariant densities associated to μ.

Lemma 24 *Let I be an open interval that is contained in the support of $(\mu_i)_{i \in S}$. If the vector field u_i does not have any critical points in I, then $\rho_i(\eta) > 0$ for any $\eta \in I$.*

Proof: Fix a point $\eta \in I$. Let $\tilde{I}$ be a closed subinterval of I, with η contained in the interior of $\tilde{I}$. Let $T > 0$ be so small that

$$\Phi_{\mathbf{i}}^{\mathbf{s}}(\eta) \in \tilde{I}$$

for any finite index sequence $\mathbf{i}$ and any corresponding sequence of switching times $\mathbf{s}$ with $\|\mathbf{s}\|_1 \leq T$.

Since u_i does not have any critical points in I, the function

$$\zeta \mapsto \exp\left(-\lambda_i \cdot \int_\zeta^\eta \frac{dx}{u_i(x)}\right)$$

is bounded below on $[\Phi_i^{-T}(\eta), \eta]$ by a constant $c > 0$. Using Lemma 9 and (40), we obtain the estimate

$$\begin{aligned}
\rho_i(\eta) \geq & \bar{\mathsf{P}}_i^T \rho_i(\eta) = \frac{1}{u_i(\eta)} \cdot \frac{1}{T} \cdot \int_{\Phi_i^{-T}(\eta)}^\eta \rho_i(\zeta) \cdot \exp\left(-\lambda_i \cdot \int_\zeta^\eta \frac{dx}{u_i(x)}\right) \, d\zeta \\
\geq & \frac{c}{T} \cdot \frac{1}{|u_i(\eta)|} \cdot \left| \int_{\Phi_i^{-T}(\eta)}^\eta \rho_i(\zeta) \, d\zeta \right| \\
= & \frac{c}{T} \cdot \frac{1}{|u_i(\eta)|} \cdot \mu_i((\Phi_i^{-T}(\eta), \eta)) > 0. \qquad (55)
\end{aligned}$$

For (55), we used that $(\Phi_i^{-T}(\eta), \eta)$ is contained in the support of μ_i. $\square$

CHAPTER VI

ASYMPTOTICS FOR INVARIANT DENSITIES AT A CRITICAL POINT

In this chapter, we derive the asymptotically dominant term of an invariant density as its argument approaches a critical point of the corresponding vector field. As in Chapters 4 and 5, we assume that $M = \mathbb{R}$. In fact, all assumptions made at the beginning of Chapter 4 remain in place. In Section 6.1, we state our results for the nonanalytic case. These results are proved in Section 6.3. In Section 6.2, we obtain slightly stronger asymptotics under the additional assumption that all vector fields in D are analytic in a neighborhood of the critical point. These results are proved in Section 6.4.

6.1 Asymptotics for nonanalytic vector fields

Let μ be an invariant measure of $(\mathsf{P}^t)_{t\geq 0}$ that is absolutely continuous with respect to the product of Lebesgue measure on $\mathbb{R}$ and counting measure on S. Let $(\rho_i)_{i\in S}$ denote the invariant densities associated to μ. In this section, we study the asymptotic behavior of $\rho_i(\eta)$, $i \in S$, as η approaches a critical point of the corresponding vector field u_i.

Let ξ be a critical point of u_i for some $i \in S$, and assume that none of the other vector fields in D have ξ as a critical point. Without loss of generality, let $u = u_1$ and let $\xi = 0$. Recall our standing assumption that for all vector fields $u_j \in D$, the set of critical points of u_j has no accumulation point (see Chapter 4). Then, there is a $\delta > 0$ such that none of the vector fields in $D \setminus \{u_1\}$ have a critical point in $[0, \delta]$ and u_1 has no critical point in $(0, \delta]$. To simplify the analysis further, we assume that there is a constant $a \neq 0$ such that

$$u_1(\eta) = -a\eta + O(\eta^2)$$

as η approaches 0 from the right, i.e. u_1 behaves almost linearly near 0. The constant a can be thought of as the contraction or expansion coefficient of u_1 near 0. If u_1 was of order $O(\eta^\alpha)$ for $\alpha < 1$, it would not be Lipschitz continuous. If u_1 was of order $O(\eta^\alpha)$ for $\alpha > 1$, identifying the asymptotically dominant term would become more complicated. Under these assumptions, we study the asymptotic behavior of ρ_1 as η approaches 0 from the right. Due to the symmetric nature of the problem, there is no need to investigate the case of η approaching 0 from the left separately.

In Section 5.2, we showed that the support of the measures $(\mu_i)_{i \in S}$ can be represented as a finite union of closed intervals of positive length (see Corollary 3). Let $\mathscr{I}$ denote the collection of these intervals. If μ is ergodic, $\mathscr{I}$ contains only one interval. Exactly one of the following statements holds:

(A) 0 is the left endpoint of an open interval that does not contain any points from the support of $(\mu_i)_{i \in S}$.

(B) 0 is contained in the interior of an interval $I \in \mathscr{I}$.

(C) 0 is the left endpoint of an interval $I \in \mathscr{I}$.

Although these statements are not formulated in terms of the given vector fields, it is easy to see which of them holds by using the algorithm in Section 5.1.

In case A, ρ_1 is constantly equal to zero on an open interval with left endpoint 0. Cases B and C are more intricate and are dealt with in Theorems 13 and 12. In case C, either 0 is the right endpoint of an open interval that does not contain any points from the support of $(\mu_i)_{i \in S}$, or 0 is the right endpoint of an interval $J \in \mathscr{I}$. But if 0 is both left endpoint of an interval $I \in \mathscr{I}$ and right endpoint of an interval $J \in \mathscr{I}$, it is uniformly critical (see Section 5.1). Since we assume that 0 is only critical for u_1, this second scenario cannot occur.

In Example 1, the projections $(\mu_i)_{i\in S}$ of the unique invariant measure are supported on $\mathbb{R}$ (see also Chapter 5). This is then an example of case B. In Example 2, the support of the measures $(\mu_i)_{i\in S}$ is the closed interval $[0,1]$, so this is an example of case C.

To state our result on the asymptotically dominant term of ρ_1 in the nonanalytic case, we introduce the function

$$\bar{\rho}(\eta) := \sum_{i>1} \lambda_{i,1} \cdot \rho_i(\eta).$$

Theorem 12 *Under the assumptions above, the following statements hold.*

1. *Let $\lambda_1 < a$. In cases B and C, there is a constant $c > 0$ such that*

$$\rho_1(\eta) = c\eta^{\frac{\lambda_1}{a}-1} + o(\eta^{\frac{\lambda_1}{a}-1})$$

as η approaches 0 from the right.

2. *Let $\lambda_1 > a > 0$. In case B,*

$$\lim_{\eta\downarrow 0} \rho_1(\eta) = \frac{\bar{\rho}(0)}{\lambda_1 - a} > 0.$$

In case C,

$$\lim_{\eta\downarrow 0} \rho_1(\eta) = \frac{\bar{\rho}(0)}{\lambda_1 - a} = 0.$$

3. *Let $\lambda_1 = a$. In case B, there exist constants $c', c > 0$ such that*

$$-c' \cdot \ln(\eta) \le \rho_1(\eta) \le -c \cdot \ln(\eta)$$

for η sufficiently small. In case C, there is a constant $c > 0$ such that

$$\rho_1(\eta) \le -c \cdot \ln(\eta).$$

Theorem 12 is proved in Section 6.3.

Remark 4 If $a < 0$, i.e. if 0 is a repelling critical point of u_1, case C is not possible (see Chapter 5). In case B,

$$\lim_{\eta \downarrow 0} \rho_1(\eta) = \frac{\bar{\rho}(0)}{\lambda_1 - a} > 0.$$

The proof of this statement is similar to the proof of Theorem 12 and we omit it.

Theorem 12 implies the following conditions for boundedness of ρ_1 to the right of 0.

Corollary 4 1. *If $\lambda_1 < a$, ρ_1 is unbounded to the right of 0 in cases B and C.*

2. *If $\lambda_1 > a > 0$, ρ_1 is bounded to the right of 0 in cases B and C.*

3. *If $\lambda_1 = a$, ρ_1 is unbounded to the right of 0 in case B. In case C, our analysis is inconclusive.*

Remark 5 The conditions in Corollary 4 align with intuition. If $\lambda_1 < a$, the rate of switching away from u_1 is lower than the rate at which u_1 contracts to its critical point 0. In this case, the rate at which mass accumulates in the vicinity of 0 is high, which results in a singularity of the invariant density at 0. If $\lambda_1 > a > 0$, the rate of switching away from u_1 is higher than the rate of contracting to 0. The rate at which mass accumulates at 0 is low and ρ_1 is bounded near 0 (see [2, Theorem 1, part c]).

6.2 Asymptotic analysis for analytic vector fields

In this section, we assume that all vector fields in D are analytic in an open interval around 0. All other assumptions from Section 6.1 are kept in place. In particular, u_1 has a critical point at 0 and

$$u_1(\eta) = -a\eta + O(\eta^2)$$

as η approaches 0 from the right. Recall from Section 6.1 that

$$\bar{\rho}(\eta) := \sum_{i>1} \lambda_{i,1} \cdot \rho_i(\eta).$$

Theorem 13 *Under the assumptions above, the following statements hold.*

1. *Let $\lambda_1 < a$. In cases B and C, there is a constant $c > 0$ such that*

$$\rho_1(\eta) = c\eta^{\frac{\lambda_1}{a}-1} + o(\eta^{\frac{\lambda_1}{a}-1})$$

   as η approaches 0 from the right.

2. *Let $\lambda_1 > a > 0$. In case B,*

$$\lim_{\eta \downarrow 0} \rho_1(\eta) = \frac{\bar{\rho}(0)}{\lambda_1 - a} > 0.$$

   In case C, there is a constant $c > 0$ such that

$$\rho_1(\eta) = c\eta^{\frac{\lambda_1}{a}-1} + o(\eta^{\frac{\lambda_1}{a}-1})$$

   as η approaches 0 from the right.

3. *Let $\lambda_1 = a$. In case B, there is a constant $c > 0$ such that*

$$\rho_1(\eta) = -c\ln(\eta) + o(\ln(\eta))$$

   as η approaches 0 from the right. In case C, $\rho_1(\eta)$ converges to a positive constant as η approaches 0 from the right.

Theorem 13 is proved in Section 6.4. Note that in the critical case $\lambda_1 = a$, the density ρ_1 is unbounded to the right of $\eta = 0$ in case B and bounded in case C.

For $\lambda_1 < a$, the conclusions of Theorems 12 and 13 are identical. For $\lambda_1 > a > 0$, Theorem 13 has a stronger conclusion in case C. And for $\lambda_1 = a$, the conclusions in Theorem 13 are stronger in both cases B and C. It is natural to ask whether the analyticity assumption in Theorem 13 is essential, or whether one can recover exactly the same conclusions under the weaker assumptions of Theorem 12. We tried to get rid of the analyticity assumption, but have not had any success so far.

6.3 Proof of Theorem 12

In this section, we prove Theorem 12.

In both cases B and C, there is an open interval I with left endpoint 0 such that

$$\rho_i(\eta) > 0$$

for any $\eta \in I$ and for any $i \in S$. This is an immediate consequence of Lemma 24. Let $\delta > 0$ be so small that none of the vector fields $(u_i)_{i>1}$ have a critical point in $[0, \delta]$, and that u_1 has no critical point in $(0, \delta]$. Let $a > 0$. The vector field u_1 is then strictly negative on $(0, \delta]$. For $\eta \in (0, \delta)$, define $\vartheta := \lim_{t \to \tau_1(\eta)} \Phi_1^{-t}(\eta)$, where $\tau_1(\eta)$ was introduced in Section 4.2. This limit is independent of the concrete choice of η.

By Lemma 24, there is a constant $c > 0$ such that $\bar{\rho}(\eta) \geq c$ for any $\eta \in [\frac{\delta}{2}, \delta]$. In case B, we can even assume that $\bar{\rho}(\eta) \geq c$ for any $\eta \in [0, \delta]$. And by Remark 3, $\bar{\rho}$ is continuous on $[0, \delta]$, which implies that $\bar{\rho}$ is bounded from above on $[0, \delta]$ by some constant $\bar{\rho}_\infty$.

Set

$$r(\eta) := -\frac{1}{u_1(\eta)} - \frac{1}{a\eta}, \quad \eta \in (0, \vartheta).$$

It is not hard to see that $r(\eta)$ is bounded on $(0, \delta]$ by a constant $r_\infty > 0$. Furthermore, as $u_1 < 0$ on $(0, \vartheta)$, we have $r(\eta) \geq -\frac{1}{a\eta}$ for any $\eta \in (0, \vartheta)$. For $\eta, \zeta \in [0, \vartheta]$, define

$$E(\eta, \zeta) := \exp\left(-\lambda_1 \cdot \int_\eta^\zeta r(x)\, dx\right).$$

Lemma 25 *The function $\zeta \mapsto \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \zeta)$ is integrable on (δ, ϑ) for any $\eta \in [0, \delta]$.*

Proof: For $\eta \in [0, \delta]$ and $\zeta \in (\delta, \vartheta)$,

$$\begin{aligned} \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \zeta) =& \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \delta) \cdot E(\delta, \zeta) \\ \leq& \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot e^{\lambda_1 \delta r_\infty} \cdot \exp\left(\lambda_1 \cdot \int_\delta^\zeta \frac{dx}{ax}\right) \\ =& \bar{\rho}(\zeta) \cdot e^{\lambda_1 \delta r_\infty} \cdot \delta^{-\frac{\lambda_1}{a}}. \end{aligned}$$

The fact that $\bar{\rho}$ is integrable implies the statement. $\square$

In analogy to Lemma 11, we have the following representation for ρ_1.

Lemma 26 *For any $\eta \in (0, \delta)$,*

$$\rho_1(\eta) = \left(\frac{\eta^{\frac{\lambda_1}{a} - 1}}{a} + r(\eta)\eta^{\frac{\lambda_1}{a}} \right) \cdot \int_\eta^\vartheta \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \zeta) \ d\zeta.$$

PROOF: Fix an $\eta \in (0, \delta)$. Using Lemma 10 and the change of variables $\zeta = \Phi_1^{-t}(\eta)$, we obtain

$$\rho_1(\eta) = -\frac{1}{u_1(\eta)} \cdot \int_\eta^\vartheta \bar{\rho}(\zeta) \cdot \exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)} \right) \ d\zeta. \tag{56}$$

Since

$$\exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)} \right) = \exp\left(-\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{ax} \right) \cdot E(\eta, \zeta) = \eta^{\frac{\lambda_1}{a}} \cdot \zeta^{-\frac{\lambda_1}{a}} \cdot E(\eta, \zeta)$$

for any $\zeta \in (\eta, \vartheta)$, and since $\zeta \mapsto \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \zeta)$ is integrable by Lemma 25, the statement follows. $\square$

PROOF OF THEOREM 12: Fix an $\eta \in (0, \delta)$. Throughout the proof, we work with the formula for ρ_1 provided in Lemma 26.

First, let $\lambda_1 < a$. Observe that $\zeta \mapsto \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(0, \zeta)$ is integrable on $(0, \delta)$ because

$$\zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(0, \zeta) \leq \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}_\infty e^{\lambda_1 \delta r_\infty}.$$

Together with Lemma 25, we see that this function is integrable on $(0, \vartheta)$, which implies that

$$\lim_{\eta \downarrow 0} \left(\int_\eta^\vartheta \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(\eta, \zeta) \ d\zeta \right) = \int_0^\vartheta \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(0, \zeta) \ d\zeta < \infty$$

by dominated convergence. In addition,

$$\begin{aligned} \int_0^\vartheta \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(0, \zeta) \ d\zeta &\geq \int_{\frac{\delta}{2}}^\delta \zeta^{-\frac{\lambda_1}{a}} \cdot \bar{\rho}(\zeta) \cdot E(0, \zeta) \ d\zeta \\ &\geq \frac{\delta^{1-\frac{\lambda_1}{a}}}{2} \cdot c \cdot e^{-\lambda_1 \delta r_\infty} > 0. \end{aligned}$$

And since $r(\eta)$ is bounded on $(0,\delta)$, $\lim_{\eta\downarrow 0}(r(\eta)\cdot\eta^{\frac{\lambda_1}{a}}) = 0$. Part 1 of Theorem 12 follows then from Lemma 26.

Now, let $\lambda_1 > a > 0$. In case B, $\bar{\rho}(0) > 0$ by Lemma 24. In case C, $\bar{\rho}(0) = 0$ because 0 is the right endpoint of an open interval that does not contain any points from the support of $(\mu_i)_{i\in S}$.

Since $\lambda_1 > a > 0$, there is a small $\alpha > 0$ such that

$$\frac{\lambda_1}{a}\cdot(1-\alpha) > 1.$$

Let $\eta \in (0,\delta)$ so that $\eta < \eta^\alpha < \delta$. Then,

$$\rho_1(\eta) = \left(\frac{\eta^{\frac{\lambda_1}{a}-1}}{a} + r(\eta)\eta^{\frac{\lambda_1}{a}}\right)\cdot\int_\eta^{\eta^\alpha}\zeta^{-\frac{\lambda_1}{a}}\cdot\bar{\rho}(\zeta)\cdot E(\eta,\zeta)\,d\zeta \tag{57}$$

$$+\left(\frac{\eta^{\frac{\lambda_1}{a}-1}}{a} + r(\eta)\eta^{\frac{\lambda_1}{a}}\right)\cdot\int_{\eta^\alpha}^{\vartheta}\zeta^{-\frac{\lambda_1}{a}}\cdot\bar{\rho}(\zeta)\cdot E(\eta,\zeta)\,d\zeta. \tag{58}$$

The term in (58) is bounded from above by

$$\begin{aligned}
&\left(\frac{\eta^{\frac{\lambda_1}{a}-1}}{a} + r_\infty\eta^{\frac{\lambda_1}{a}}\right)\cdot\left(\int_{\eta^\alpha}^{\delta}\zeta^{-\frac{\lambda_1}{a}}\cdot\bar{\rho}(\zeta)\cdot E(\eta,\zeta)\,d\zeta + \int_\delta^\vartheta\zeta^{-\frac{\lambda_1}{a}}\cdot\bar{\rho}(\zeta)\cdot E(\eta,\zeta)\,d\zeta\right)\\
\le&\left(\frac{\eta^{\frac{\lambda_1}{a}-1}}{a} + r_\infty\eta^{\frac{\lambda_1}{a}}\right)\cdot(\eta^{-\frac{\lambda_1}{a}\alpha}\cdot e^{\lambda_1\delta r_\infty}\cdot\|\bar{\rho}\|_1 + \delta^{-\frac{\lambda_1}{a}}\cdot e^{\lambda_1\delta r_\infty}\cdot\|\bar{\rho}\|_1)\\
=&e^{\lambda_1\delta r_\infty}\cdot\|\bar{\rho}\|_1\cdot\left(\frac{\eta^{\frac{\lambda_1}{a}\cdot(1-\alpha)-1}}{a} + r_\infty\eta^{\frac{\lambda_1}{a}\cdot(1-\alpha)} + \frac{\delta^{-\frac{\lambda_1}{a}}}{a}\eta^{\frac{\lambda_1}{a}-1} + r_\infty\delta^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\right),
\end{aligned}$$

which converges to 0 as η approaches 0 from the right.

Since $\eta^\alpha < \delta$, the function $\zeta\mapsto\bar{\rho}(\zeta)\cdot E(\eta,\zeta)$ is continuous on $[\eta,\eta^\alpha]$. By the mean-value theorem for integration, there exists $\zeta_\eta \in (\eta,\eta^\alpha)$ such that the term to the right of the equality sign in (57) equals

$$\begin{aligned}
&\left(\frac{\eta^{\frac{\lambda_1}{a}-1}}{a} + r(\eta)\eta^{\frac{\lambda_1}{a}}\right)\cdot\int_\eta^{\eta^\alpha}\zeta^{-\frac{\lambda_1}{a}}\,d\zeta\cdot\bar{\rho}(\zeta_\eta)\cdot E(\eta,\zeta_\eta)\\
=&\left(\frac{1}{a}\cdot(1-\eta^{(1-\alpha)\cdot(\frac{\lambda_1}{a}-1)}) + r(\eta)\cdot(\eta-\eta^{\alpha+(1-\alpha)\cdot\frac{\lambda_1}{a}})\right)\cdot\frac{a}{\lambda_1-a}\cdot\bar{\rho}(\zeta_\eta)\cdot E(\eta,\zeta_\eta).
\end{aligned}$$

Since $\zeta_\eta \in (\eta,\eta^\alpha)$ for any η, it is clear that $\lim_{\eta\downarrow 0}\zeta_\eta = 0$. Continuity of $\bar{\rho}$ at $\eta = 0$ and integrability of $r(x)$ on $(0,\delta)$ then imply that

$$\lim_{\eta\downarrow 0}\left(\bar{\rho}(\zeta_\eta)\cdot E(\eta,\zeta_\eta)\cdot\frac{a}{\lambda_1-a}\right) = a\cdot\frac{\bar{\rho}(0)}{\lambda_1-a}.$$

Furthermore,

$$\lim_{\eta\downarrow 0}\left(\frac{1}{a}\cdot(1-\eta^{(1-\alpha)\cdot(\frac{\lambda_1}{a}-1)})\right)=\frac{1}{a}.$$

Finally, for small $\eta>0$, we have $\eta>\eta^{\alpha+(1-\alpha)\cdot\frac{\lambda_1}{a}}$. It follows that

$$|r(\eta)|\cdot(\eta-\eta^{\alpha+(1-\alpha)\cdot\frac{\lambda_1}{a}})\le r_\infty\cdot(\eta-\eta^{\alpha+(1-\alpha)\cdot\frac{\lambda_1}{a}}),$$

which converges to 0 as η approaches 0 from the right. This completes the proof of part 2 of Theorem 12.

Finally, assume that $\lambda_1=a$. For $\eta\in(0,\delta)$,

$$\begin{aligned}\rho_1(\eta)=&\left(\frac{1}{a}+r(\eta)\eta\right)\cdot\int_\eta^\delta\zeta^{-1}\cdot\bar\rho(\zeta)\cdot E(\eta,\zeta)\;d\zeta\\&+\left(\frac{1}{a}+r(\eta)\eta\right)\cdot\int_\delta^\vartheta\zeta^{-1}\cdot\bar\rho(\zeta)\cdot E(\eta,\zeta)\;d\zeta.\end{aligned}\tag{59}$$

By Lemma 25, the term in (59) is bounded on $(0,\delta)$. In case B, $c\le\bar\rho(\eta)\le\bar\rho_\infty$ for any $\eta\in[0,\delta]$. Therefore,

$$\begin{aligned}&-c\cdot e^{-\lambda_1\delta r_\infty}\cdot\ln(\eta)+c\cdot e^{-\lambda_1\delta r_\infty}\cdot\ln(\delta)\\\le&\int_\eta^\delta\zeta^{-1}\cdot\bar\rho(\zeta)\cdot E(\eta,\zeta)\;d\zeta\\\le&-\bar\rho_\infty\cdot e^{\lambda_1\delta r_\infty}\cdot\ln(\eta)+\bar\rho_\infty\cdot e^{\lambda_1\delta r_\infty}\cdot\ln(\delta)\end{aligned}$$

for $\eta\in(0,\delta)$. As

$$\lim_{\eta\downarrow 0}\left(\frac{1}{a}+r(\eta)\eta\right)=\frac{1}{a},$$

this establishes part 3 of Theorem 12 for case B. In case C, we only have $\bar\rho(\eta)\le\bar\rho_\infty$, which is why we obtain a weaker statement. $\square$

6.4 Proof of Theorem 13

In this section, we prove Theorem 13. The ensuing paragraph follows [2].

For $i\in S$, we introduce the probability flux

$$\varphi_i(\eta):=\rho_i(\eta)\cdot u_i(\eta).$$

The vector of probability fluxes $(\varphi_1(\eta), \ldots, \varphi_n(\eta))^T$ is denoted by $\varphi(\eta)$. As in Section 6.3, we let $\delta > 0$ be so small that the vector fields $(u_i)_{i>1}$ have no critical point in $[0, \delta]$ and u_1 has no critical point in $(0, \delta]$. Since the invariant densities $(\rho_i)_{i\in S}$ are $\mathscr{C}^1$ on $(0, \delta)$, they satisfy the Fokker–Planck equations

$$\rho_i'(\eta)u_i(\eta) + \rho_i(\eta)u_i'(\eta) = -\lambda_i\rho_i(\eta) + \sum_{l\neq i} \lambda_{l,i}\rho_l(\eta), \quad i \in S, \tag{60}$$

on $(0, \delta)$, see [17]. Written in terms of the probability fluxes, (60) becomes

$$\varphi_i'(\eta) = -\frac{\lambda_i}{u_i(\eta)} \cdot \varphi_i(\eta) + \sum_{l\neq i} \frac{\lambda_{l,i}}{u_l(\eta)} \cdot \varphi_l(\eta), \quad i \in S. \tag{61}$$

In Appendix B, we show how Equation (56) can be derived directly from the Fokker–Planck equations if the invariant densities are $\mathscr{C}^1$.

Our approach is to derive the asymptotically dominant term for the probability flux φ_1, which will then immediately give the asymptotically dominant term for ρ_1. We begin by showing that $\lim_{\eta\downarrow 0} \varphi_1(\eta) = 0$.

Lemma 27 *We have* $\lim_{\eta\downarrow 0} \varphi_1(\eta) = 0$.

PROOF: By Remark 3, the limit $\lim_{\eta\downarrow 0} \varphi_i(\eta)$ exists for any $i > 1$. It is an easy corollary of (61) that

$$\sum_{i\in S} \varphi_i'(\eta) = 0$$

for any $\eta \in (0, \delta)$. Thus, the sum of all probability fluxes is equal to a constant k on this interval. Since

$$\varphi_1(\eta) = k - \sum_{i>1} \varphi_i(\eta)$$

for any $\eta \in (0, \delta)$, the limit $l := \lim_{\eta\downarrow 0} \varphi_1(\eta)$ exists as well.

It remains to show that $l = 0$. To obtain a contradiction, assume that $l \neq 0$. Then, there is no loss of generality in assuming that

$$|\varphi_1(\eta)| \geq \frac{|l|}{2}$$

for any $\eta \in (0, \delta)$. Since $u_1(\eta) = -a\eta + o(\eta)$ as η approaches 0 from the right, we may also assume that

$$\left| \frac{u_1(\eta)}{\eta} \right| \leq 2|a|, \quad \eta \in (0, \delta).$$

But this yields

$$\int_0^\delta \rho_1(\eta)\, d\eta = \int_0^\delta \frac{|\varphi_1(\eta)|}{|u_1(\eta)|}\, d\eta \geq \frac{|l|}{4|a|} \cdot \int_0^\delta \frac{d\eta}{\eta} = \infty,$$

which contradicts the fact that ρ_1 is integrable. $\square$

Corollary 5 *In case C,* $\lim_{\eta \downarrow 0} \varphi(\eta) = 0$.

PROOF: In case C, the invariant densities $(\rho_i)_{i \in S}$ vanish to the left of 0. By Remark 3, the densities $(\rho_i)_{i>1}$ are continuous at 0, which implies that $\lim_{\eta \downarrow 0} \rho_i(\eta) = 0$ for any $i > 1$. Hence, $\lim_{\eta \downarrow 0} \varphi_i(\eta) = 0$ for any $i > 1$, and $\lim_{\eta \downarrow 0} \varphi_1(\eta) = 0$ by Lemma 27. $\square$

Recall that $k := |S|$. We introduce the matrix of switching rates

$$\Lambda := \begin{pmatrix} -\lambda_1 & \lambda_{2,1} & \cdots & \lambda_{k,1} \\ \lambda_{1,2} & -\lambda_2 & \cdots & \lambda_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1,k} & \lambda_{2,k} & \cdots & -\lambda_k \end{pmatrix},$$

and let $U(\eta)$ be the diagonal matrix with diagonal entries $\frac{1}{u_1(\eta)}, \ldots, \frac{1}{u_k(\eta)}$.

For a fixed $\epsilon \in (0, \delta)$, we consider the initial-value problem

$$\begin{aligned} \phi'(\eta) &= \Lambda \cdot U(\eta) \cdot \phi(\eta), \\ \phi(\epsilon) &= \varphi(\epsilon), \end{aligned} \tag{62}$$

whose unique solution is $\varphi(\eta)$. Initial-value problem (62) can be written equivalently as

$$\begin{aligned} \phi'(\eta) &= \frac{1}{\eta} B(\eta) \cdot \phi(\eta), \\ \phi(\epsilon) &= \varphi(\epsilon). \end{aligned} \tag{63}$$

Here,

$$B(\eta) := \Lambda \cdot \tilde{U}(\eta),$$

where $\tilde{U}(\eta)$ is the diagonal matrix with diagonal entries $\frac{\eta}{u_1(\eta)}, \dots, \frac{\eta}{u_k(\eta)}$. Note that $B(\eta)$ is analytic at $\eta = 0$. This follows from the fact that the diagonal entries of $\tilde{U}(\eta)$ are analytic at $\eta = 0$, which is easily derived from analyticity of the vector fields. The linear system (63) then has a so-called regular singular point at $\eta = 0$ (see [33, Section 3.11]).

Since $B(\eta)$ is analytic at $\eta = 0$, there exist $\rho \in (0, \delta)$ and a sequence of matrices $(B_l)_{l \geq 0}$ such that

$$B(\eta) = \sum_{l=0}^{\infty} \eta^l \cdot B_l \tag{64}$$

for any $\eta \in (-\rho, \rho)$. There is no loss of generality in assuming that $\rho = \delta$. Since $u_1(\eta) = -a\eta + O(\eta^2)$, and since $u_i(\eta) \neq 0$ for any $i > 1$, the matrix B_0 in (64) has the form

$$B_0 = \begin{pmatrix} \frac{\lambda_1}{a} & 0 & \cdots & 0 \\ -\frac{\lambda_{1,2}}{a} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\lambda_{1,k}}{a} & 0 & \cdots & 0 \end{pmatrix}.$$

It is easy to give a complete description of the eigenvalues and corresponding eigenspaces of B_0.

Lemma 28 *The matrix B_0 has eigenvalues $\frac{\lambda_1}{a}$ and 0. The eigenspace corresponding to $\frac{\lambda_1}{a}$ is spanned by the vector $\lambda := (\lambda_1, -\lambda_{1,2}, -\lambda_{1,3}, \dots, -\lambda_{1,k})^T$. The eigenspace corresponding to 0 is the orthogonal complement to the span of $\{(1, 0, \dots, 0)^T\}$.*

We omit the proof of Lemma 28.

At this point, we need to distinguish between two cases. First, assume that $\frac{\lambda_1}{a}$ is not an integer. Such a condition is sometimes referred to as a nonresonance condition. The following statement is then a reformulation of [33, Proposition 11.2].

Lemma 29 *There is a function*

$$V(\eta) = \mathbb{1} + \sum_{l=1}^{\infty} \eta^l \cdot V_l \tag{65}$$

that satisfies the normal equation

$$\eta V'(\eta) = B(\eta)V(\eta) - V(\eta)B_0, \quad \eta \in (0, \delta) \tag{66}$$

and for which

$$\varphi(\eta) = V(\eta) \cdot \exp\left(\ln\left(\frac{\eta}{\epsilon}\right) B_0\right) V(\epsilon)^{-1}\varphi(\epsilon), \quad \eta \in (0, \delta).$$

Now, we consider the resonance case, i.e we assume that $\frac{\lambda_1}{a}$ is a positive integer. In this case, we may not be able to construct a solution of the form (65) to (66). Instead, we consider the modified version

$$\eta V'(\eta) = B(\eta)V(\eta) - V(\eta)(B_0 + \eta^{\frac{\lambda_1}{a}} Y), \tag{67}$$

where Y is a matrix satisfying

$$B_0 Y = Y\left(B_0 + \frac{\lambda_1}{a}\mathbb{1}\right). \tag{68}$$

In this setting, we have the following reformulation of [33, Proposition 11.5].

Lemma 30 *There exist a function $V(\eta)$ of the form* (65) *and a matrix Y satisfying* (68) *such that $V(\eta)$ satisfies* (67) *with Y and*

$$\varphi(\eta) = V(\eta) \cdot \exp\left(\ln\left(\frac{\eta}{\epsilon}\right) B_0\right) \cdot \exp\left(\ln\left(\frac{\eta}{\epsilon}\right) Y\right) V(\epsilon)^{-1}\varphi(\epsilon), \quad \eta \in (0, \delta).$$

Proof of Theorem 13: Comparing Theorems 13 and 12, we see that we only need to show part 2 for case C and part 3 for both cases.

Let $\nu \in \mathbb{R}$ and let $\tilde{y} \in \mathbb{R}^k$ with first component equal to 0 such that

$$V(\epsilon)^{-1}\varphi(\epsilon) = \nu\lambda + \tilde{y},$$

where λ was defined in Lemma 28. In the nonresonance case, Lemma 28 implies that

$$\begin{aligned}
\exp\left(\ln\left(\frac{\eta}{\epsilon}\right)B_0\right)V(\epsilon)^{-1}\varphi(\epsilon) &= \sum_{l=0}^{\infty}\frac{1}{l!}\cdot\left(\ln\left(\frac{\eta}{\epsilon}\right)\right)^l(\nu B_0^l\lambda + B_0^l\tilde{y}) \\
&= \tilde{y} + \nu\lambda + \sum_{l=1}^{\infty}\frac{1}{l!}\cdot\left(\ln\left(\frac{\eta}{\epsilon}\right)\right)^l\nu\left(\frac{\lambda_1}{a}\right)^l\lambda \\
&= \tilde{y} + \nu\cdot\exp\left(\frac{\lambda_1}{a}\cdot\ln\left(\frac{\eta}{\epsilon}\right)\right)\lambda \\
&= \tilde{y} + \nu\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\lambda, \qquad (69)
\end{aligned}$$

so

$$\varphi(\eta) = \left(\mathbb{1} + \sum_{l=1}^{\infty}\eta^l V_l\right)\cdot(\tilde{y} + \nu\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\lambda), \quad \eta\in(0,\delta) \qquad (70)$$

by Lemma 29. From (70), we infer that

$$\tilde{y} = \lim_{\eta\downarrow 0}\varphi(\eta).$$

In case C, Corollary 5 implies that $\tilde{y} = 0$. If ν was equal to 0, it would then follow that $\varphi \equiv 0$ on $(0,\delta)$. This is impossible in light of Lemma 24. As a result,

$$\varphi(\eta) = \nu\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\lambda + o(\eta^{\frac{\lambda_1}{a}})$$

as η approaches 0 from the right. This establishes part 2 of Theorem 13 for case C and under the assumption that $\frac{\lambda_1}{a}$ is not an integer.

In the resonance case, Proposition 11.6 in [33] implies that $Y^2 = 0$, that $Y\lambda = 0$ and that $Y\tilde{y}$ is an eigenvector of B_0 corresponding to the eigenvalue $\frac{\lambda_1}{a}$. Together with Lemma 30, this yields

$$\begin{aligned}
\varphi(\eta) =& V(\eta)\cdot\exp\left(\ln\left(\frac{\eta}{\epsilon}\right)B_0\right)\cdot\left(\nu\lambda + \tilde{y} + \ln\left(\frac{\eta}{\epsilon}\right)Y(\nu\lambda + \tilde{y})\right) \\
=& V(\eta)\cdot\left(\exp\left(\ln\left(\frac{\eta}{\epsilon}\right)B_0\right)(\nu\lambda + \tilde{y})\right. \qquad (71) \\
&\left. + \ln\left(\frac{\eta}{\epsilon}\right)\cdot\left(Y\tilde{y} + \sum_{l=1}^{\infty}\frac{1}{l!}\cdot\left(\ln\left(\frac{\eta}{\epsilon}\right)\right)^l\left(\frac{\lambda_1}{a}\right)^l Y\tilde{y}\right)\right). \qquad (72)
\end{aligned}$$

Using (69) and (65), the term in (71) and (72) becomes

$$\left(\mathbb{1}+\sum_{l=1}^{\infty}\eta^l V_l\right)\cdot(\tilde{y}+\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\cdot(\nu\lambda-\ln(\epsilon)Y\tilde{y})+\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\ln(\eta)Y\tilde{y}). \tag{73}$$

Let us first consider the situation where $\frac{\lambda_1}{a}>1$. In case C, $\tilde{y}=0$ and we obtain

$$\varphi(\eta)=\nu\epsilon^{-\frac{\lambda_1}{a}}\eta^{\frac{\lambda_1}{a}}\cdot\lambda+o(\eta^{\frac{\lambda_1}{a}})$$

as η approaches 0 from the right. Since $\nu\neq 0$ by Lemma 24, we have established part 2 of Theorem 13 for case C under the assumption that $\frac{\lambda_1}{a}$ is an integer larger than 1.

Now, suppose that $\frac{\lambda_1}{a}=1$. In case C, Representation (73) of $\varphi(\eta)$ implies that

$$\varphi(\eta)=\nu\epsilon^{-1}\eta\cdot\lambda+o(\eta),$$

and part 3 of Theorem 13 follows for case C. In case B, (73) yields

$$\varphi(\eta)=\tilde{y}+\epsilon^{-1}\eta\ln(\eta)Y\tilde{y}+o(\eta\ln(\eta)).$$

Since $Y\tilde{y}$ is an eigenvector of B_0 with corresponding eigenvalue $\frac{\lambda_1}{a}$, Lemma 28 implies that the first component of $Y\tilde{y}$ is nonzero. This yields part 3 of Theorem 13 for case B. $\square$

APPENDIX A

DISCUSSION OF THEOREMS 5 AND 6

The basic idea behind Theorems 5 and 6 is that for a sufficient number of switches, by perturbing the switching time sequences one can generate perturbations to the terminal point in all directions.

The first statement of Theorem 6 corresponds to Theorem 3.1 in [24], which reads as follows: Under the assumptions of Theorem 6, any neighborhood U of ξ contains points that are normally accessible from ξ at arbitrarily small times. A point η in M is called normally accessible from ξ at time $t > 0$ if there exist vectors $\mathbf{i} \in S^{m+1}$ and $(\hat{t}_1, \ldots, \hat{t}_{m+1}) \in \Delta_{t,m+1}$ such that $F_{\mathbf{i}}(\hat{t}_1, \ldots, \hat{t}_{m+1}) = \eta$ and the differential $DF_{\mathbf{i}}(\hat{t}_1, \ldots, \hat{t}_{m+1})$ has full rank. In [24], this is established along the following lines: Fix a neighborhood U of ξ and a time $T > 0$. Since the weak hypoellipticity condition holds in an open neighborhood of ξ, we can assume without loss of generality that the weak hypoellipticity condition holds at every point in U. Recall that n is the dimension of M. Theorem 3.1 in [24] will follow once we show the following statement:

For $1 \leq k \leq n$, there exist an index vector $\mathbf{i} \in S^k$ and an open set $U_k \subset \Delta_{T,k}$ such that the map

$$F_k : \ \Delta_{T,k} \to M, (t_1, \ldots, t_k) \mapsto \Phi_{\mathbf{i}}^{(t_1, \ldots t_k)}(\xi) \tag{74}$$

has the following properties.

(a) The rank of $DF_k(\eta)$ equals k for any $\eta \in U_k$.

(b) The set $F_k(U_k)$ is a k-dimensional submanifold of $M \cap U$.

We use induction. In the base case $k = 1$, there is an index $i \in S$ with $u_i(\xi) \neq 0$, for otherwise we would obtain that $\mathcal{I}(D)(\xi) = \{0\}$, contradicting our assumption that the weak hypoellipticity condition holds at ξ. Since $u_i(\Phi_i^0(\xi)) = u_i(\xi) \neq 0$ and since $t \mapsto u_i(\Phi_i^t(\xi))$ is continuous, there is an $\epsilon \in (0, T)$ such that $u_i(\Phi_i^t(\xi)) \neq 0$ for all $t \in (0, \epsilon)$. Define the map

$$F_1 : \ \Delta_{T,1} \to M, t \mapsto \Phi_i^t(\xi).$$

By the constant-rank theorem (see [24]), there is an open set $U_1 \subset (0, \epsilon)$ such that $F_1(U_1)$ is a one-dimensional submanifold of $M \cap U$.

In the induction step, assume that the statement holds for some $k \in \{1, \ldots, n-1\}$. Then, there exist an index vector $\mathbf{i} \in S^k$ and an open set $U_k \subset \Delta_{T,k}$ such that the map F_k defined according to (74) has properties a and b. Next, we show that there is some $\eta \in F_k(U_k)$ and some index $i \in S$ such that $u_i(\eta)$ is not an element of the tangent space $T_\eta F_k(U_k)$. Under the assumption that $u_i(\eta) \in T_\eta F_k(U_k)$ for all $\eta \in F_k(U_k)$ and for all $i \in S$, one can show that $\mathcal{I}(D)(\eta) \subset T_\eta F_k(U_k)$ for some $\eta \in F_k(U_k)$. But property b states that $T_\eta F_k(U_k)$ has dimension k, which is by assumption strictly less than the dimension n of $\mathcal{I}(D)(\eta)$, a contradiction. In the sequel, we work with this point $\eta \in F_k(U_k)$ and with this index $i \in S$. Since $\eta \in F_k(U_k)$, there is a time vector $\hat{\mathbf{t}} \in U_k$ with $\eta = F_k(\hat{\mathbf{t}})$. Define the map

$$F_{k+1} : \ \Delta_{T,k+1} \to M, (t_1, \ldots, t_{k+1}) \mapsto \Phi_i^{t_{k+1}}(F_k(t_1, \ldots, t_k)).$$

Then,

$$DF_{k+1}(\hat{\mathbf{t}}, 0) = (DF_k(\hat{\mathbf{t}}), u_i(F_k(\hat{\mathbf{t}})) = (DF_k(\hat{\mathbf{t}}), u_i(\eta)).$$

Every column of $DF_k(\hat{\mathbf{t}})$ is in $T_\eta F_k(U_k)$, so the rank of this matrix is k. In addition, $u_i(\eta)$ is not an element of $T_\eta F_k(U_k)$, so $DF_{k+1}(\hat{\mathbf{t}}, 0)$ has full rank $(k+1)$. Then, there is a neighborhood $W \subset \Delta_{T,k+1}$ of $(\hat{\mathbf{t}}, 0)$ where DF_{k+1} has full rank. Notice that one can make sense of this last statement, even though $(\hat{\mathbf{t}}, 0)$ is only a boundary point of $\Delta_{T,k+1}$. By the constant-rank theorem, there is an open set $U_{k+1} \subset W$ such that $F_{k+1}(U_{k+1})$ is a $(k+1)$-dimensional submanifold of $M \cap U$. This finishes the induction step.

It's worth pointing out that only one sequence $\mathbf{i}$ resulting in a map F with a regular point was constructed. But since the flow generated by any vector field is a family of diffeomorphisms, and since the set of points satisying the weak hypoellipticity condition is open, one can append any indices in front or at the back of that sequence without destroying the desired properties, and thus recover this part of Theorem 6 as we state it. The fact that the interior of $L(\xi)$ is nonempty and dense in $L(\xi)$ follows from Theorem 3.2.a in [24]. Theorem 5 follows from applying Theorem 3.1 ([24]) to $\mathbb{R} \times M$ and vector fields $\mathbf{1} \oplus u_i, i \in S$, where

$$(\mathbf{1} \oplus u)(r, \xi) := (1, u(\xi)), \quad (r, \xi) \in \mathbb{R} \times M,$$

and $\mathbf{1}$ is the unit vector field on $\mathbb{R}$ corresponding to the derivation $\partial/\partial r$ and identically equal to 1 in the natural coordinates on $\mathbb{R}$.

APPENDIX B

HOW EQUATION (56) RELATES TO THE FOKKER–PLANCK EQUATIONS

Equation (56) in the proof of Lemma 26 can also be derived from the Fokker–Planck equations, but in order to do this, one needs to assume that the invariant densities are $\mathscr{C}^1$.

It is an immediate consequence of (60) that

$$\bar{\rho}(\zeta) = (\lambda_1 + u_1'(\zeta)) \cdot \rho_1(\zeta) + u_1(\zeta) \cdot \rho_1'(\zeta),$$

see [2]. Hence, the term to the right of the equality sign in (56) equals

$$-\frac{1}{u_1(\eta)} \cdot \int_\eta^\vartheta (\lambda_1 + u_1'(\zeta)) \cdot \rho_1(\zeta) \cdot \exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)}\right) d\zeta \tag{75}$$

$$-\frac{1}{u_1(\eta)} \cdot \int_\eta^\vartheta \rho_1'(\zeta) \cdot u_1(\zeta) \cdot \exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)}\right) d\zeta. \tag{76}$$

As

$$\lim_{\zeta \uparrow \vartheta}\left(\rho_1(\zeta) \cdot u_1(\zeta) \cdot \exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)}\right)\right) = 0$$

if u_1 is smooth and forward-complete, integration by parts implies that the term in (76) equals

$$\rho_1(\eta) + \frac{1}{u_1(\eta)} \cdot \int_\eta^\vartheta (\lambda_1 + u_1'(\zeta)) \cdot \rho_1(\zeta) \cdot \exp\left(\lambda_1 \cdot \int_\eta^\zeta \frac{dx}{u_1(x)}\right) d\zeta. \tag{77}$$

Since the second term in (77) cancels with the term in (75), we obtain (56).

REFERENCES

[1] Bakhtin, Y. and Hurth, T., "Invariant densities for dynamical systems with random switching," *Nonlinearity*, vol. 25, no. 10, pp. 2937–2952, 2012.

[2] Balázs, M., Horváth, G., Kolumbán, S., Kovács, P., and Telek, M., "Fluid level dependent Markov fluid models with continuous zero transition," *Performance Evaluation*, vol. 68, no. 11, pp. 1149 – 1161, 2011. Special Issue: Performance 2011.

[3] Barenblatt, G. I., Iooss, G., and Joseph, D. D., eds., *Nonlinear dynamics and turbulence*. Interaction of Mechanics and Mathematics Series, Boston, MA: Pitman (Advanced Publishing Program), 1983.

[4] Bass, R. F., *Diffusions and elliptic operators*. Probability and its Applications (New York), New York: Springer-Verlag, 1998.

[5] Bell, D. R., *The Malliavin calculus*. Mineola, NY: Dover Publications Inc., 2006. Reprint of the 1987 edition.

[6] Benaïm, M., Le Borgne, S., Malrieu, F., and Zitt, P.-A., "Qualitative properties of certain piecewise deterministic Markov processes," *Available at http://arxiv.org/abs/1204.4143, accepted for publication in Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 2012.

[7] Benaïm, M., Le Borgne, S., Malrieu, F., and Zitt, P.-A., "Quantitative ergodicity for some switched dynamical systems," *Electron. Commun. Probab.*, vol. 17, pp. no. 56, 14, 2012.

[8] Benaïm, M., Le Borgne, S., Malrieu, F., and Zitt, P.-A., "On the stability of planar randomly switched systems," *Ann. Appl. Probab.*, vol. 24, no. 1, pp. 292–311, 2014.

[9] Chow, W.-L., "Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung," *Math. Ann.*, vol. 117, pp. 98–105, 1939.

[10] Cloez, B. and Hairer, M., "Exponential ergodicity for Markov processes with random switching," *Available at http://arxiv.org/abs/1303.6999*, 2013.

[11] Da Prato, G. and Zabczyk, J., *Ergodicity for infinite-dimensional systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[12] Davis, M. H. A., "Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models," *J. Roy. Statist. Soc. Ser. B*, vol. 46, no. 3, pp. 353–388, 1984. With discussion.

[13] Davis, M. H. A., *Markov models and optimization*, vol. 49 of *Monographs on Statistics and Applied Probability.* London: Chapman & Hall, 1993.

[14] Davydov, Y. A., Lifshits, M. A., and Smorodina, N. V., *Local properties of distributions of stochastic functionals*, vol. 173 of *Translations of Mathematical Monographs.* Providence, RI: American Mathematical Society, 1998. Translated from the 1995 Russian original by V. E. Nazaĭkinskiĭ and M. A. Shishkova.

[15] Durrett, R., *Probability: Theory and Examples.* Duxbury Advanced Series, Belmont: Thomson – Brooks/Cole, third ed., 2005.

[16] Ethier, S. N. and Kurtz, T. G., *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.

[17] Faggionato, A., Gabrielli, D., and Ribezzi Crivellari, M., "Non-equilibrium thermodynamics of piecewise deterministic Markov processes," *J. Stat. Phys.*, vol. 137, no. 2, pp. 259–304, 2009.

[18] Goldstein, S., "On diffusion by discontinuous movements, and on the telegraph equation," *Quart. J. Mech. Appl. Math.*, vol. 4, pp. 129–156, 1951.

[19] Hairer, M., "Ergodic properties of Markov processes." Lectures given at the University of Warwick, http://www.hairer.org/notes/Markov.pdf, 2006.

[20] Hairer, M. and Mattingly, J. C., "Yet another look at Harris' ergodic theorem for Markov chains," in *Seminar on Stochastic Analysis, Random Fields and Applications VI*, vol. 63 of *Progr. Probab.*, pp. 109–117, Birkhäuser/Springer Basel AG, Basel, 2011.

[21] Hirsch, M. W., Smale, S., and Devaney, R. L., *Differential equations, dynamical systems, and an introduction to chaos*, vol. 60 of *Pure and Applied Mathematics (Amsterdam).* Elsevier/Academic Press, Amsterdam, second ed., 2004.

[22] Horton, G., Kulkarni, V., Nicol, D., and Trivedi, K., "Fluid stochastic Petri nets: Theory, applications, and solution techniques," *European Journal of Operational Research*, vol. 105, pp. 184–201, 1998.

[23] Hurth, T., "A one-dimensional motion in an attracting environment with random switches," 2010.

[24] Jurdjevic, V., *Geometric control theory*, vol. 52 of *Cambridge Studies in Advanced Mathematics.* Cambridge: Cambridge University Press, 1997.

[25] Kac, M., "A stochastic model related to the telegrapher's equation," *Rocky Mountain J. Math.*, vol. 4, pp. 497–509, 1974. Reprinting of an article published in 1956, Papers arising from a Conference on Stochastic Differential Equations (Univ. Alberta, Edmonton, Alta., 1972).

[26] Krener, A. J., "A generalization of Chow's theorem and the bang-bang theorem to non-linear control problems," *SIAM J. Control*, vol. 12, pp. 43–52, 1974.

[27] Lawley, S. D., Mattingly, J. C., and Reed, M. C., "Sensitivity to switching rates in stochastically switched odes," *Available at http://arxiv.org/abs/1310.2525*, 2013.

[28] Lindvall, T., *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[29] Meyn, S. and Tweedie, R. L., *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second ed., 2009. With a prologue by Peter W. Glynn.

[30] Nualart, D., *The Malliavin calculus and related topics*. Probability and its Applications (New York), Berlin: Springer-Verlag, second ed., 2006.

[31] Stroock, D. W., "The Malliavin calculus and its application to second order parabolic differential equations. I," *Math. Systems Theory*, vol. 14, no. 1, pp. 25–65, 1981.

[32] Sussmann, H. J. and Jurdjevic, V., "Controllability of nonlinear systems," *J. Differential Equations*, vol. 12, pp. 95–116, 1972.

[33] Taylor, M. E., *Introduction to Differential Equations*, vol. 14 of *Pure and Applied Undergraduate Texts*. Providence, RI: American Mathematical Society, 2011.

[34] Taylor, M. E., *Measure theory and integration*, vol. 76 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society, 2006.

[35] Tucker, W., "The Lorenz attractor exists," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 328, no. 12, pp. 1197–1202, 1999.

[36] Villani, C., *Optimal transport*, vol. 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[37] Yin, G. G. and Zhu, C., *Hybrid switching diffusions*, vol. 63 of *Stochastic Modelling and Applied Probability*. New York: Springer, 2010. Properties and applications.

[38] Young, L.-S., "What are SRB measures, and which dynamical systems have them?," *J. Statist. Phys.*, vol. 108, no. 5-6, pp. 733–754, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.