Aspects of Invariant Manifold Theory and Applications

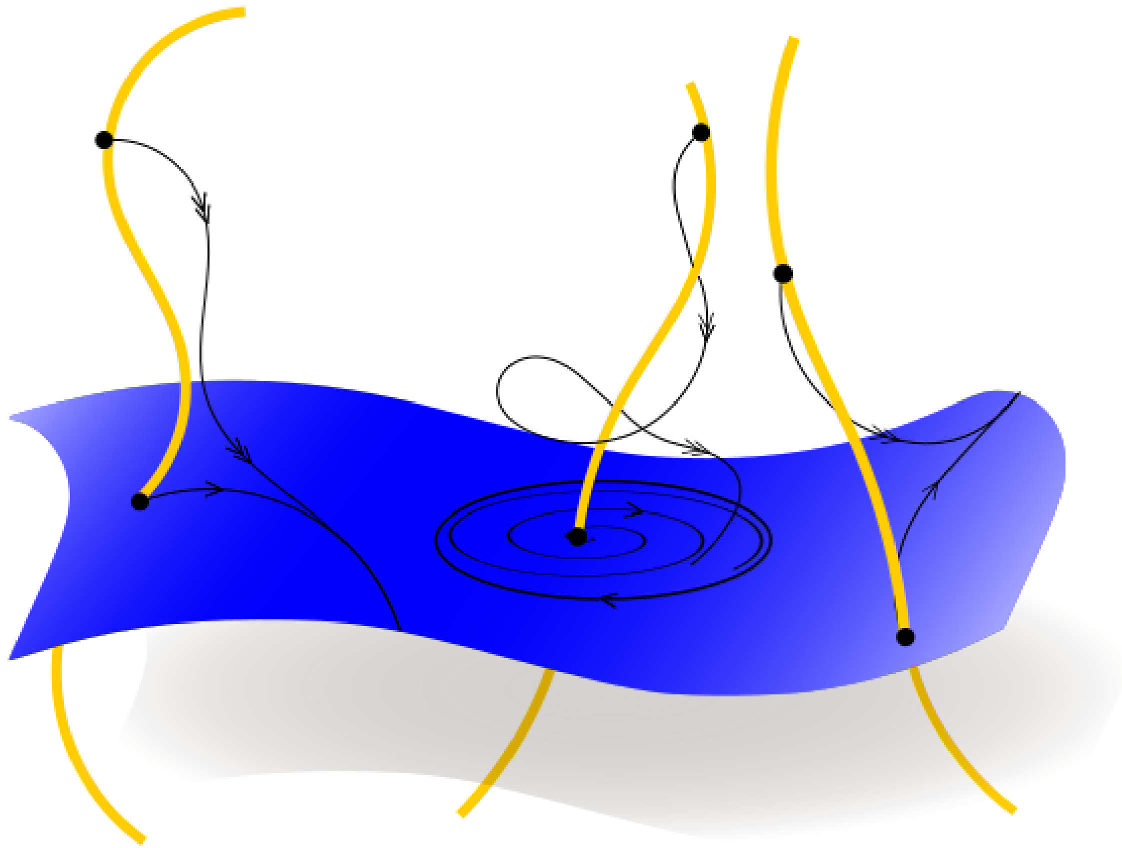
by

Matthew D. Kvalheim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical and Computer Engineering)
in The University of Michigan
2018

Doctoral Committee:

Assistant Professor Shai Revzen, Chair
Professor Anthony M. Bloch
Professor Jessy W. Grizzle
Professor Ralf J. Spatzier



Matthew Kvalheim

kvalheim@umich.edu

ORCID iD: 0000-0002-2662-6760

To my parents

ACKNOWLEDGEMENTS

I would like to begin by thanking my research advisor, Professor Shai Revzen, for taking a chance by investing in me and for the many subsequent hours invested on my behalf. I suspect that I was not always an easy student to advise, and I thank him for his patience, kindness, and restraint in allowing me to pursue my own interests. I also thank him for his many crucial research insights over the years. Throughout the past five years I have learned many valuable academic and personal lessons from him, from his careful approach to science, his ability to see the "big picture" and transcend academic disciplines, his open-minded attitude, and his leadership style. I consider him a role model both academically and personally.

I would like to thank several members of the administrative staff at the University of Michigan. To Becky Turanski, Kevin Calhoun, José-Antonio Rubio, Anne Rhoades, Michelle Feldkamp, Karen Liska, Kathy Austin, and Judi Jones: thank you for helping me.

Next, I would like to thank my collaborators. First and foremost, I thank Dr. Jaap Eldering, with whom I had the pleasure of publishing my first mathematics paper. Without his collaboration, mentorship, careful thinking, and our countless emails and video conferences from across the world, two of the chapters in this thesis would have never come into existence. I thank Dr. Simon Wilshin as well as Professors Clayton Scott and John Guckenheimer for their collaboration on work related to the third chapter of this thesis, and I also thank Professor Peter J. Thomas for helpful suggestions related to the same chapter.

I am fortunate to have interacted with many fantastic professors throughout my education. I thank Professors David Matolak, Jim Zhu, Jerrel Mitchell, Maarten Uijt de Haag, and Douglas Lawrence for their mentorship and generosity with their time spent answering my many questions as an undergraduate. I thank Professors Scott Schneider, Wouter van Limbeek, Sandeep Pradhan, Anthony Bloch, Jessy Grizzle, and Ralf Spatzier for their enthusiasm as educators, for shaping the way I think about solving problems, for their time generously spent on me in (and out of) office hours, and for the career and research advice they have given me during graduate school. I also thank Professors Bloch, Grizzle, and Spatzier for serving on my thesis committee, and for their valuable feedback regarding my work.

Next, I would like to thank a number of my peers. I thank George Council for our many (mathematical, philosophical, cinematic...) stimulating conversations over the years, and for significantly influencing my way of thinking — I feel very fortunate to have begun my Ph.D. journey at the same time and in the same lab as George. I thank Brian Bittner, with whom I had the pleasure of collaborating on work related to this thesis, and my labmates Dan Zhao and Yue Sun for providing an enjoyable working environment. I thank Vikram Sachdeva for his positive attitude and for the substantial amount of entertainment he continually provides me. I thank JJ Lipor, David Hong, and Aniket Deshmukh for many interesting conversations over the years, and I additionally thank JJ for the camaraderie when we took measure theory. I thank Joe Borja, Rob Vandermeulen, Michael Fisher, Ryan Sandberg, and William Clark for conversations about mathematics and helpful feedback on my work. Last but not least, I thank Tim Howell, Grant Fresch, Andrew Orso, Dana Berns, Bradley Keller, and Drew Fisher for keeping me balanced, and for their friendship throughout the good times and bad times.

Finally and most importantly, I thank all of the members of my family. I give my most heartfelt thanks to my parents for raising me, for their unconditional love and support, and for giving me the opportunities which led to my education. I dedicate this thesis to them.

Last but far from least, I thank Emily Joelle for her unconditional love and support, for her patience and understanding, for always believing in me, and for her willingness to drive 125 miles every weekend for two years just to see one another. I am lucky to have you.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

ABSTRACT

We develop tools for the experimental study of biological and robotic systems, based on mathematical concepts from dynamical systems theory. Many of these tools, both theoretical and applied, may be applicable to other classes of systems which are sufficiently complicated that their governing equations defy analytical derivation. Recent years have seen a surge of interest in "data-driven" machine learning approaches to determine such equations of motion, but the high dimensionality of many systems makes direct learning approaches intractable despite modern computing advances. Our approach is to make use of the fact that parsimonious assumptions often lead to significant insights from dynamical systems theory; such insights can be leveraged in learning algorithms to mitigate the "curse of dimensionality" and render them feasible. The dynamical insights we use come in two primary flavors: existence of reduced-order models such as invariant manifolds, and existence of "normal forms" (simple forms of the equations of motion, possibly after a change of coordinates). The former would imply that there are fewer meaningful parameters to compute for a learning algorithm when compared with the number of parameters informed by naive inspection. The latter can directly inform selection of regressors for learning algorithms, reducing a problem in nonparametric estimation to a parametric one, or else inform other targets for estimation.

After the introduction, in the second chapter we study global properties of the global (center-)stable manifold of a normally attracting invariant manifold (NAIM), the special case of a normally hyperbolic invariant manifold (NHIM) with empty unstable bundle. We restrict our attention to continuous-time dynamical systems, or flows. We show that the global stable foliation of a NAIM has the structure of a topological disk bundle, and that

similar statements hold for inflowing NAIMs and for general compact NHIMs. Furthermore, the global stable foliation has a C^k disk bundle structure if the local stable foliation is assumed C^k. We then show that the dynamics restricted to the stable manifold of a compact inflowing NAIM are globally topologically conjugate to the linearized transverse dynamics at the NAIM. Moreover, we give conditions ensuring the existence of a global C^k linearizing conjugacy. We also prove a C^k global linearization result for inflowing NAIMs; we believe that even the local version of this result is new, and may be useful in applications to slow-fast systems. We illustrate the theory by giving applications to geometric singular perturbation theory in the case of an attracting critical manifold: we show that the domain of the Fenichel Normal Form can be extended to the entire global stable manifold, and under additional nonresonance assumptions we derive a smooth global linear normal form.

In the third chapter, we restrict our attention to oscillators, a special case of the NAIMs of the second chapter. Oscillators are ubiquitous in nature, and usually associated with the existence of an "asymptotic phase" which governs the long-term dynamics of the oscillator. We show that asymptotic phase can be expressed as a line integral with respect to a uniquely defined closed differential 1-form, and provide an algorithm for estimating this "Temporal 1-Form" from observational data. Unlike all previously available data-driven phase estimation methods, our algorithm can: (i) use observations that are much shorter than a cycle; (ii) recover phase within the entire region for which data convergent to the limit cycle is available; (iii) recover the phase response curves (PRC-s) that govern weak oscillator coupling; (iv) show isochron curvature, and recover nonlinear features of isochron geometry. Our method may find application wherever models of oscillator dynamics need to be constructed from measured or simulated time-series.

In the fourth chapter, we shift our attention to smooth locomotion (organismal, robotic) such as swimming in viscous media. Many forms of locomotion, both natural and artificial, are dominated by viscous friction in the sense that without power expenditure they quickly

come to a standstill. From geometric mechanics, it is known that in the "Stokesian" (viscous; zero Reynolds number) limit, the motion is governed by a reduced order "connection" model that describes how body shape change produces motion for the body frame with respect to the world. In the "perturbed Stokes regime" where inertial forces are still dominated by viscosity, but are not negligible (low Reynolds number), we show that motion is still governed by a functional relationship between shape velocity and body velocity, but this function is no longer connection-like. We derive this model using results from noncompact NHIM theory in a singular perturbation framework. Using a normal form derived from theoretical properties of this reduced-order model, we develop an algorithm that estimates an approximation to the dynamics near a cyclic body shape change (a "gait") directly from observational data of shape and body motion. This extends our previous work which assumed kinematic "connection" models. To compare the old and new algorithms, we analyze simulated swimmers over a range of inertia to damping ratios. Our new class of models performs well in the Stokesian regime, and over several orders of magnitude outside it into the perturbed Stokes regime, where it gives significantly improved prediction accuracy compared to previous models. This new and more general class of models is of independent interest. Their application to data-driven modeling improves our ability to study the optimality of animal gaits, and our ability to use hardware-in-the-loop optimization to produce gaits for robots.

CHAPTER I

Introduction

1.1 Motivation

The fundamental question motivating this thesis is the following: how can we best study dynamical systems which are sufficiently complicated (e.g., biological systems) that their equations of motion defy direct derivation?

In accordance with the trends of the last decade, one idea is to take a "data-driven" approach and "learn" the equations of motion directly from experimental data. Unfortunately, the sheer volume of data required to naively learn high-fidelity models of sufficiently complex systems renders this idea infeasible in many cases. The primary culprit — the so-called "curse of dimensionality" — obstructs intuition in addition to computation, and is also a common affliction in many other endeavors such as the efficient design of control policies via optimization.

However, dynamical systems theory tells us that parsimonious assumptions about a system often lead to significant theoretical insights into its behavior. These insights can be leveraged to inform data-driven algorithms, thereby mitigating the aforementioned difficulties. We believe that the following are two particularly useful classes of insights:

1. Existence of *reduced-order models* (or briefly *reduced models*) which are lower-dimensional,

and which faithfully approximate the high-dimensional model in an appropriate sense.

2. Existence of *normal forms*.

Suppose that there exists a reduced-order model which captures all or most of the behavior of interest of some system. This means that there are fewer meaningful parameters for a data-driven algorithm to compute, when compared with the number of parameters informed by naive inspection of the system. By converting high-dimensional problems to low-dimensional ones, approaches based on reduced models have enjoyed recent success in areas other than model identification as well, such as the design of control laws for dynamic behavior in bipedal robots [WGK03, HG17, DG17].

Normal forms are a general concept in mathematics, and refer to a "canonical" choice of representative of an equivalence class of mathematical objects, and is typically "simple" in some sense. Examples include the Jordan canonical form and row echelon form for matrices. In dynamics, normal forms concern the existence of coordinate systems in which the equations of motion look particularly "simple." Knowledge that a certain normal form exists is tantamount to explicit knowledge of what qualitative behaviors are possible for a dynamical system, and this knowledge can directly inform targets of estimation for data-driven algorithms (see especially Chapter IV).

This thesis concerns the theory and applications of such normal forms and *model reductions* in the context of smooth dynamical systems theory.

1.2 Background: Reduction and normal forms for dynamical systems

We now introduce some notions from dynamical systems theory which are relevant for this thesis. Consider continuous-time dynamical systems (N, Φ^t) and (M, Ψ^t) given by the flows $\Phi\colon \mathbb{R}\times N \to N$ and $\Psi\colon \mathbb{R}\times M \to M$ on the smooth manifolds M, N.

If $\dim(M) < \dim(N)$, then we consider (M, Ψ^t) to be a reduced model for (N, Φ^t) in either, or both, of the following two situations.

1. There exists a *semiconjugacy* $\pi\colon N \to M$, which means that

$$\forall t \in \mathbb{R} : \Psi^t \circ \pi = \pi \circ \Phi^t. \tag{1.1}$$

2. There exists an embedding $\iota\colon M \to N$ which is also a semiconjugacy:

$$\forall t \in \mathbb{R} : \iota \circ \Psi^t = \Phi^t \circ \iota. \tag{1.2}$$

For us, ι will always be at least $C^{k\geq 1}$. On the other hand, we will consider semiconjugacies π as in (1.1) which are merely continuous, although clearly a smooth π provides a better model reduction.

Given families of continuous-time dynamical systems $(\Psi^t_\lambda)_{\lambda\in\Lambda}$ and $(\Phi^t_\lambda)_{\lambda\in\Lambda}$ on M and N generated by vector fields $(X_\lambda)_{\lambda\in\Lambda}$ and $(Y_\lambda)_{\lambda\in\Lambda}$, we will say that[1] Y_λ is a normal form for X_λ if there exists a homeomorphism (or better, a diffeomorphism) $h\colon M \to N$ which is a conjugacy:

$$\forall t \in \mathbb{R}\colon \Phi^t \circ h = h \circ \Psi^t. \tag{1.3}$$

We next discuss these concepts in the context of invariant manifolds in §1.2.1. As a special but relevant case, we then discuss oscillators in §1.2.2. Finally, in §1.2.3 we discuss model reduction and normal forms for a class of dissipative mechanical systems with symmetry which are relevant for the locomotion of certain animals and robots.

[1]Or perhaps more properly, $(Y_\lambda)_{\lambda\in\Lambda}$ is a normal form for $(X_\lambda)_{\lambda\in\Lambda}$.

1.2.1 Invariant manifolds

An invariant manifold $M \subset N$ is a submanifold of the state space of a dynamical system (N, Φ^t) which is a union of trajectories: $\forall t \in \mathbb{R} : \Phi^t(M) = M$. Invariant manifolds are reduced-order models in the sense of (1.2), with ι the inclusion map. In practical terms, if the system begins in a state belonging to the reduced-order model then it can no longer escape back to exhibiting more complex behaviors. Under additional hypotheses, invariant manifolds are also reduced models in the sense of (1.1). A nice survey of the utility of invariant manifold methods in science and engineering is given in [Wig94, Ch. 1].

One useful class of invariant manifolds are the *normally hyperbolic invariant manifolds (NHIMs)* [HPS77, Fen71, Wig94, Eld13]. NHIMs are generalizations of hyperbolic fixed points and periodic orbits, which are characterized by the property that the dynamics tangent to the invariant manifold are dominated by the dynamics transverse to the invariant manifold (at least to first order). NHIMs are roughly defined to be invariant manifolds sharing this property; a precise definition is given in §2.3. The most relevant class of invariant manifolds for this thesis are NHIMs which are attracting, which we call *normally attracting invariant manifolds (NAIMs)*. With respect to the distance induced by any Riemannian or Finsler metric, a compact NAIM attracts all nearby states at an exponential rate. The theory and applications of compact NAIMs are the focus of this thesis, as described in §1.3. (Hence we focus on NAIMs below, but note that most of the mentioned results have analogues for NHIMs as well.)

NAIMs have a few properties which make them very nice from the perspectives of mathematical modeling and model reduction. (For now, our NAIMs are without boundary.) First, compact NAIMs persist under sufficiently C^1-small perturbations of the vector field defining the dynamical system [HPS77, Fen71], and generalizations exist for the noncompact case [BLZ98, Eld13]. Conversely, any compact invariant manifold which persists under all

C^1-small perturbations are normally hyperbolic [Mañ78], with some qualifications[2]. This is desirable because physical measurements cannot determine parameters of a mathematical model with perfect accuracy, so any physically meaningful feature of a mathematical model should persist under small perturbations. Next, there always exists a (Hölder) continuous map $P^s \colon W^s(M) \to M$ defined on the *global stable manifold* (basin of attraction in the case of a boundaryless NAIM) $W^s(M)$, with the following properties [Fen74, HPS77]:

- P^s is a retraction, i.e., $P^s|_M = \mathrm{id}_M$.
- P^s is a semiconjugacy between $(W^s(M), \Phi^t|_{W^s(M)})$ and $(M, \Phi^t|_M)$.

Hence NAIMs are also reduced models in the sense of (1.1), with P^s playing the role of π. For a compact NAIM without boundary, the map P^s is unique. Any NAIM attracts all nearby states at an exponential rate, but the reason that NAIMs are good reduced-order models is more subtle than this: not only do nearby states approach these invariant manifolds, but they approach *specific trajectories* within the invariant manifold, and they approach these at a faster exponential rate than they approach any other trajectories. This property follows from the existence of the semiconjugacy P^s, and it provides a precise sense in which the restricted dynamics on the NAIM approximate the full model.

The semiconjugacy P^s is sometimes called the "asymptotic phase" map in the literature [BK94], although this terminology is much more common in the special case that M is an exponentially stable periodic orbit for a flow (M is indeed a NAIM in this case). In the dynamics literature, the level sets $(P^s)^{-1}(m)$ of P^s are called the leaves of the "invariant foliation," "stable foliation," or "strong stable foliation" [HPS77]. Assuming that the normal bundle of M is trivial, then it is always possible to complete $m := P^s(x)$ to a set of (continuous) coordinates (m, z) defined on a (possibly small) neighborhood U of M, such that M

[2]Of course if underlying symmetries are present in a physical model, then arbitrary C^1-perturbations may not be possible.

corresponds to $\{z = 0\}$, and in which the dynamics take the normal form

$$\begin{aligned} \dot{z} &= \Lambda(m,z)z \\ \dot{m} &= f(m), \end{aligned} \tag{1.4}$$

and under additional hypotheses these coordinates can be chosen $C^{k\geq 1}$. A classical result [PS70], generalizing the Hartman-Grobman theorem, states that it is actually possible to choose different transverse z coordinates defined on a (possibly small) neighborhood U of M which improve the above normal form to a *linear* normal form

$$\begin{aligned} \dot{z} &= A(m)z \\ \dot{m} &= f(m), \end{aligned}$$

where again there exist additional hypotheses ensuring that these coordinates can be chosen $C^{k\geq 1}$ [Tak71, Sel84, Sel83, BK94, Sak94].

In Chapter II, we will study the topological properties of P^s in detail, and we will prove that these normal forms extend to the entire global stable manifold of M (i.e., the entire basin of attraction in the case of a boundaryless M).

There are also extensions to NHIM theory for invariant manifolds with boundary [Fen71, Fen74, Fen77]. The portion of the definition of such a NHIM involving tangential/transverse flow rates is similar (see §2.3 for the precise definition). However, here the submanifold is merely required to be locally invariant: trajectories in the submanifold need not be confined there for all time, but are required to enter or exit through the boundary only (through an abuse of language, we still call the submanifold an invariant manifold). Depending on the results one wishes to obtain, the vector field is assumed to point inward or outward at the boundary of the invariant manifold; such a manifold is then respectively called *inflowing* or *overflowing* (sometimes outflowing). Two key results are as follows: (i) an inflowing

NAIM M also has a unique continuous asymptotic phase map $P^s\colon W^s(M) \to M$, and (ii) an overflowing NAIM persists under C^1-small perturbations. We study these invariant manifolds in Chapter II, and these invariant manifolds also play a key theoretical role in Chapter IV.

One important class of systems in which NHIMs occur naturally are the so-called *multiple time-scale* or *slow-fast* systems [Kue15]. The invariant manifolds and associated geometric structures of such systems are emphasized in the context of geometric singular perturbation theory, introduced by Fenichel [Fen79]. Here the theory of inflowing and overflowing invariant manifolds is important, since the compact NHIMs occurring in such systems often have boundary. We will study examples of slow-fast systems further in Chapters II and IV.

For completeness, let us relate NHIMs to the possibly more widely-known *center manifolds*. We briefly describe the most basic notion of center manifold relevant for our discussion, and refer the reader to [Car82, GH83] for more details. Given a system of differential equations $\dot{x} = f(x)$ and an equilibrium point x_0 with $f(x_0) = 0$, the eigenvalues of the linearization $\mathsf{D}f(x_0)$ split into collections of eigenvalues having negative, zero, and positive real part. These collections of eigenvalues respectively determine stable, center, and unstable subspaces. The center manifold theorem states that there exist "stable," "center," and "unstable" invariant manifolds respectively tangent to these subspaces. Trajectories in the stable (resp. unstable) manifold approach x_0 exponentially in positive (resp. negative) time. While the stable and unstable manifolds are always unique, in general the center manifold need not be. Center manifolds and NHIMs have somewhat similar "spectral" properties, but they differ in that NHIMs have an instrinsic global definition whereas center manifolds are only defined locally. This local definition manifests itself in the fact that center manifolds are in general nonunique (see [Eld13, Sec. 1.1.2] for more discussion). However, NHIM theory is closely related to center manifold theory, and indeed the existence and properties of center manifolds can be proven as an easy application of the general theory of NHIMs with boundary (see, e.g., [Wig94, Sec. 7.2]).

1.2.2 Oscillators

The image of a periodic orbit of a (continuous-time) dynamical system provides a specific and simple example of an invariant manifold. If the periodic orbit is isolated, we call it a *limit cycle*. Choose any smooth hypersurface intersecting the orbit transversely in a single point. An easy argument using the implicit function theorem shows that there exists a well-defined *Poincaré map* from the hypersurface to itself defined by following the flow of the dynamical system (possibly after shrinking the domain of this map). If the eigenvalues of the linearization of the Poincaré map at the periodic orbit are disjoint from the unit circle, such a periodic orbit is a limit cycle and is called *hyperbolic*. (These eigenvalues of the Poincaré map are conventionally called *Floquet multipliers*, and their logarithms are called *Floquet exponents*.) It is easy to show that the image of such a limit cycle is actually a NHIM. If furthermore the eigenvalues lie inside the open unit disk, then it follows that the periodic orbit is an asymptotically stable limit cycle which attracts all nearby states at an exponential rate. We use the term *oscillator* to refer to the dynamics within the basin of attraction of such an exponentially stable limit cycle. Oscillators are useful models of a variety of phenomena appearing in contexts such as chemical kinetics, electrical circuits, neuroscience, and the locomotion of animals and robots. More examples and details are given in Chapter III, and oscillators appear in the context of locomotion in Chapter IV.

Since the limit cycle's image Γ is a NAIM, it has asymptotic phase: there exists a map $P^s\colon W^s(\Gamma) \to \Gamma$ defined on the basin of attraction $W^s(\Gamma)$ to Γ which is a semiconjugacy. We can assign a "phase coordinate" $\theta \in [0, 2\pi)$ to points on Γ which increases at a constant rate, and we can pull this coordinate back via P^s to construct a phase coordinate θ on all of $W^s(\Gamma)$. This coordinate is as smooth as the underlying dynamical system except at the discontinuity $\theta^{-1}(0)$, although if we identify the endpoints of $[0, 2\pi]$ and view θ as circle-valued, then this coordinate is smooth everywhere. It follows that the level sets of θ are the leaves of the (strong) stable foliation [HPS77] discussed in § 1.2.1, often called *isochrons* in the context of oscillators [Guc75, Win80] and in the physics literature. These isochrons are permuted by the flow, and all points in a single isochron approach each other exponentially fast as they converge toward the limit cycle. Explained differently, since Γ is a NAIM, the normal forms (1.4) and (1.2.1) apply. Furthermore, it can be shown that the linear normal form (1.2.1) is always C^1 for an exponentially stable limit cycle, and application of Floquet theory can further improve this normal form to

$$\dot{z} = Bz$$
$$\dot{\theta} = 1,$$

where B is a constant Hurwitz matrix [BK94]. Hence we see that θ evolves autonomously, as do the remaining transverse coordinates, which decay to zero after a transient phase.

The above shows that the dynamics restricted to the limit cycle are a reduced model in the sense of both (1.1) and (1.2) (again, this is true of all NAIMs). However, for many practical applications, this particular model reduction may be too coarse. The Floquet multipliers of an oscillator will generically be distinct, and in that case normal form theory [BK94] shows that there exist additional attracting invariant "slow manifolds" S corresponding to "slow" Floquet multipliers with large magnitude/modulus[3]. See Figure 1.1. We can alternatively consider such a slow manifold as a reduced model for the the dynamics on $W^s(\Gamma)$, and it can be shown that trajectories in $W^s(\Gamma)$ converge to specific trajectories in S: explicitly, there exists another semiconjugacy $P_1^s\colon W^s(\Gamma) \to S$. Alternatively, we may view the dynamics restricted to Γ as a reduced model for the dynamics restricted to S: all trajectories within S asymptotically converge to Γ in forward time, and there is yet another semiconjugacy

[3]Stronger results on the smoothness of S can be obtained from the pseudo-stable manifold theorem [HPS77] applied to a Poincaré map, which implies that S is always as smooth as the underlying vector field.

$P_2^s\colon S \to \Gamma$. Interestingly, these semiconjugacies compose: $P^s = P_2^s \circ P_1^s$, which illustrates that these model reductions are transitive. As a technical aside, we briefly mention that it can be shown that P^s and P_2^s are automatically as smooth as the underlying dynamics, but P_1^s is a priori merely C^1 in this case, and asymptotic phase maps for general normally hyberbolic invariant manifolds might be merely (Hölder) continuous; thus natural geometric structures for very smooth dynamics can have very poor regularity.

In the physical context of animal or robot locomotion relevant for Chapter IV, the limit cycle may be viewed as representing a perfectly periodic gait subject to no environmental or neuromechanical perturbations. The invariant slow manifold may then be viewed as a collection of states having "slow recovery" when perturbed from the steady gait. Any states not belonging to the slow manifold will quickly return, and may be viewed as "posture errors".

1.2.3 Locomotion of dissipative mechanical systems with symmetry

The study of the locomotion of animals and robots is the study of how bodies move through space by deforming their "shape" to produce reaction forces from the environment that propel the body (e.g., swimming). In this section, we consider the equations of motion for such a mechanical locomotion system which is subject to linear viscous drag forces.

The configuration space of a large class of such locomotion systems can be written as a trivial principle G-bundle $Q = S \times G$, where S is a smooth manifold describing the space of internal body shapes or *shape space*, with G a closed subgroup of $\mathsf{SE}(3)$. If the system is subject to a viscous damping force determined by a Rayleigh dissipation function but not subject to any additional external forces or nonholonomic constraints[4], then under the

[4]In order to be interesting, we must of course allow an external force exerted by the locomoting body. This physically implies that this force takes values in the annihilator of the vertical distribution $\mathbf{0} \times \mathsf{T}G$, with $\mathbf{0}$ the zero section of $\mathsf{T}S$, and therefore it does not affect the derivation of (1.5). See Chapter IV for details.

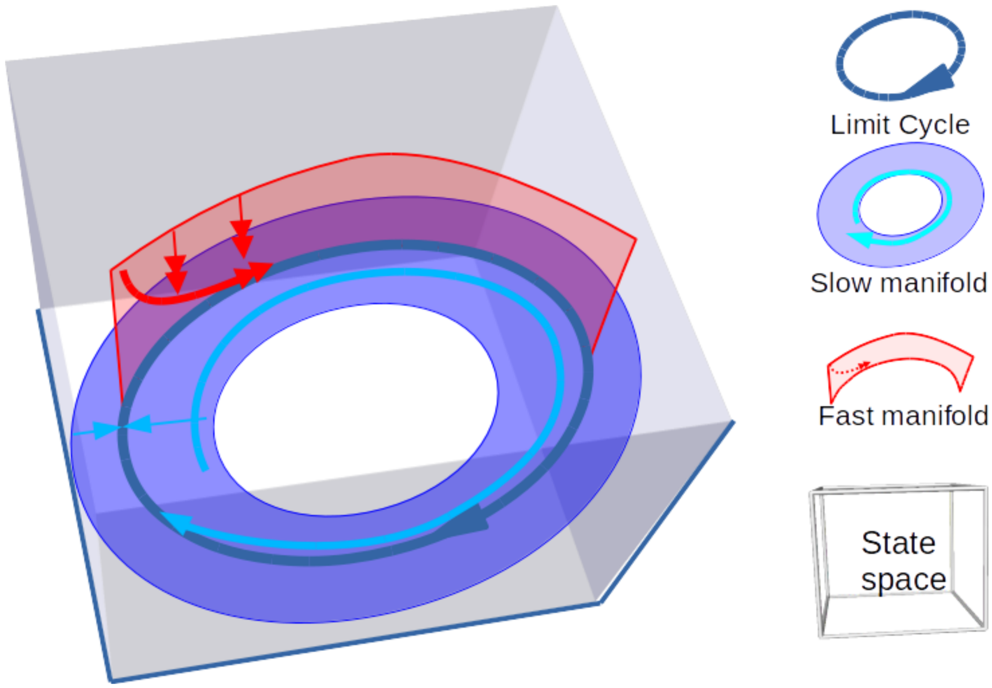


Figure 1.1: A depiction of some invariant manifolds present in the basin of attraction of a generic oscillator. The collection of states corresponding to Floquet multipliers with relatively large magnitude form an invariant "slow manifold." Normal form theory [BK94] shows that there are also "fast manifolds" corresponding to Floquet multipliers with smaller magnitude, on which the rate of attraction to the limit cycle is faster.

assumption that the Lagrangian and viscous force are symmetric under the action of G on $Q = S \times G$ given by left multiplication on the second factor, it can be shown that there always exist coordinates $(r, \dot{r}, g, p)$ for $\mathsf{T}(S \times G)$ such that the equations of motion take the normal form [KM96, Eq. 1]

$$\begin{aligned} g^{-1}\dot{g} &= -A_{\text{mech}}\dot{r} + \mathbb{I}_{\text{loc}}^{-1}p \\ \mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1}\dot{p} &= \mathbb{I}_{\text{loc}}(A_{\text{visc}} - A_{\text{mech}})\dot{r} + p + \mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1}\text{ad}^*_{g^{-1}\dot{g}}p. \end{aligned} \tag{1.5}$$

Here the coordinate $p \in \mathfrak{g}^*$ takes values in the dual space $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$ of G, and $\text{ad}^* \colon \mathfrak{g}^* \to \text{End}(\mathfrak{g}^*)$ is the dual of the adjoint representation $\text{ad} \colon \mathfrak{g} \to \text{End}(\mathfrak{g})$ defined via the Lie bracket $\text{ad}_\xi(\eta) \coloneqq [\xi, \eta]$. The linear maps $A_{\text{mech}}, A_{\text{visc}}, \mathbb{V}_{\text{loc}}, \mathbb{I}_{\text{loc}}$ all depend on r and are respectively called the *local mechanical connection*, the *local viscous connection*, the *local viscous tensor*, and the *local locked inertia tensor*. We will discuss these quantities in more detail in Chapter IV. We mention that (1.5) can be essentially viewed as a special case of a very general reconstruction equation providing a normal form for nonholonomic mechanical systems with symmetry, as derived in [BKMM96].

If viscous damping forces are large relative to the inertia of the locomoting body in a uniform way, then we expect $\mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1}$ to be small (say, with respect to some norm on $\mathfrak{g}^*$). In this case, Kelly and Murray [KM96] formally set $\mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1} = 0$ in (1.5) to obtain $p = \mathbb{I}_{\text{loc}}(A_{\text{mech}} - A_{\text{visc}})$. Substituting this expression into the first equation of (1.5), they obtain

$$g^{-1}\dot{g} = -A_{\text{visc}}\dot{r}. \tag{1.6}$$

In global geometric language, this equation implies that solution curves in Q are horizontal with respect to a certain principal connection on Q (see Definition C.15 in Appendix C) called the *viscous connection*.

Eldering and Jacobs studied this procedure in more detail by working in a singular perturbation framework using NHIM theory [EJ16, App. A]. Roughly speaking, they additionally show that for $\mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1}$ small but nonzero, then under certain additional assumptions there is a NAIM for (1.5) which is close to the viscous connection and attracts all nearby solutions[5]. An (asymptotic) series approximation of this manifold can be explicitly calculated. It is shown in [EJ16, App. A] that these calculations can be used to obtain correction terms for the case that $\mathbb{I}_{\text{loc}}\mathbb{V}_{\text{loc}}^{-1}$ is small but nonzero; these can be added to render (1.5) more accurate. The paper [Eld16] deals with the general problem of realizing nonholonomic constraints as a limit of friction forces in a differential geometric framework, and contains similar computations. The computation of such correction terms (for dynamics restricted to a slow manifold)

[5]The viscous connection is a subbundle of $\mathsf{T}Q$, so in particular it is also a submanifold of $\mathsf{T}Q$, and therefore it makes sense to talk about it being "close" to another submanifold in a suitable topology.

is a basic technique in singular perturbation theory [Jon95], and for example [SKK87] use this technique to compute corrections to the reduced equations of motion for a robot with flexible joints.

1.3 Contributions

This thesis contributes to the theory and applications of normally hyperbolic invariant manifolds (NHIMs).

The first set of theoretical contributions of this work concern the theory of general NHIMs. Restricting our attention to continuous-time dynamical systems, we study global properties of the global (center-)stable manifold $W^s(M)$ of a normally attracting invariant manifold (NAIM) M, the special case of a NHIM which is attracting. We show that if the local stable foliation is C^k, then the global stable foliation has the structure of a C^k disk bundle $P^s\colon W^s(M) \to M$, where P^s is the asymptotic phase map discussed in §1.2.1 (see Appendix C for the definition of a disk bundle). In particular, the global stable manifold is always a topological disk bundle. We also show that similar statements hold for inflowing NAIMs and for general compact NHIMs.

We next show that the linear normal form (1.2.1) discussed in §1.2.1 extends globally to the entire global stable manifold[6]. Moreover, we give conditions ensuring the C^k version of this result, and we also prove a C^k global linearization result for inflowing NAIMs; we believe that even the local version of this inflowing result is new, and may be useful in applications to slow-fast systems.

We illustrate these new results by giving applications to geometric singular perturbation theory in the case of an attracting critical manifold. In particular, we show that the domain

[6]This is true even if the normal bundle of M is nontrivial (this was assumed for simplicity of exposition in §1.2.1), although the statement of the normal form result needs to be modified to refer to a linear flow on a vector bundle. See Chapter II for the precise details.

of the well-known Fenichel Normal Form [JK94, Jon95, Kap99, JT09] can be extended to the entire global stable manifold, and under additional nonresonance assumptions we derive a smooth *global linear* normal form.

The next theoretical contribution of this thesis concerns the dynamics in the basin of attraction of an exponentially stable limit cycle, which we refer to as an oscillator. We introduce and develop the properties of a certain closed differential 1-form which we call the *Temporal 1-Form*. The distribution determined by the kernel of the Temporal 1-Form is integrable and the associated foliation is the (strong-)stable foliation of the oscillator, whose leaves are often called isochrons [Win80, Guc75]. In other words, the Temporal 1-Form is locally the exterior derivative of the asymptotic phase map, when this map is viewed as $\mathbb{R}$-valued as described in §1.2.2. We show that the globally defined Temporal 1-Form is uniquely characterized by two local properties: (i) it is closed, and (ii) its contraction with the underlying vector field is identically one.

We give an application using the Temporal 1-Form: we develop an algorithm to compute the asymptotic phase map of an oscillator which uses only (noisy) time-series data, but does not assume any knowledge of the equations of motion. The algorithm works by computing the Temporal 1-Form via least-squares regression, and is made possible by the fact that this 1-form is globally uniquely defined by the local properties mentioned in the previous paragraph. In contrast, previous methods to compute the asymptotic phase map require explicit knowledge of the equations of motion [Win80, LKO14], and/or require long time series [MM12], and/or are accurate only on the limit cycle itself [RG08]. Either of the first two requirements preclude application of these methods to many experimental data sets, where equations are unknown and successful long trials are rare, and accuracy only on the limit cycle itself is an obvious limitation of an algorithm. We derive theoretical bounds on the performance of our algorithm applied to data afflicted by both measurement and system noise, and in particular data generated by stochastic differential equations. Using a variety of data sets, we test and compare our algorithm to other existing phase estimation methods.

Our final contributions are to the study of locomotion in a regime wherein viscous damping dominates the inertia of the locomoting body. In the limit as the ratio of inertia to damping approaches zero, the motion of the body is governed by the so-called viscous connection, as mentioned in §1.2.3.

Assuming this situation (the so-called "Stokesian limit"), [BHR18] have recently developed a data-driven algorithm to compute the viscous connection — and therefore the equations of motion — over a neighborhood of a periodic orbit in shape space. Their algorithm only requires a noisy ensemble of shape, shape velocity, and body velocity measurements as input. Computation of the equations of motion over a neighborhood of the periodic orbit allows one to test necessary conditions for optimality of the *gait* (periodic orbit) of the locomoting system, and these tests can be nicely formulated in terms of geometric properties of the viscous connection. Additionally, knowledge of a dynamical model enables hardware-in-the-loop optimization for robotics engineering. To understand the value regarding this latter application, note that the infinite dimensionality of appropriate spaces of gaits necessitates a large number of experiments in order to compute the gradient of an appropriate cost functional for use in gait optimization. The cost of running large number of physical experiments — both in terms of time and robot wear — would render hardware-in-the-loop optimization infeasible, but knowledge of a dynamical model allows one to offload these physical experiments to simulation. Using this suggested approach for hardware-in-the-loop optimization, only enough physical experiments to estimate a reasonable local (near a gait) model are first required.

However, for the locomoting systems of interest the inertia/damping ratio can never actually be zero, and is small at best. Hence the Stokesian limit assumed by [BHR18] might not be a reasonable assumption, depending on the specific system. However, in this situation a simplified (nonlinear) connection-like model still exists: our final theoretical contribution extends[7] the results of [EJ16, App. A] to show that if the inertia/damping ratio is small and the shape space is compact, then a NAIM exists which approximates the viscous connection. This observation enables explicit computations of correction terms which, when added to the reduced equations of motion, provably yield a more accurate model than that of the viscous connection model. The corrected equations of motion still involve only the body velocity together with the shape variables and their derivatives. Finally, we give an application of this result: we modify the algorithm of [BHR18] to produce a modified algorithm which approximates the reduced equations of motion for the dynamics restricted to the NAIM. We test this algorithm on a computer simulation of an idealized swimming model of a robot in a viscous fluid, and compare the results to those obtained via the algorithm of [BHR18]. We find that our new algorithm performs offers a significant improvement for a range of inertia-damping ratios.

1.4 Overview of the sequel

The remainder of this thesis is organized as follows.

Chapter II contains our contributions to the general theory of NHIMs, and is based on [EKR18]. Chapter III introduces the Temporal 1-Form for an oscillator, develops the associated theory, and contains work on utilizing the Temporal 1-Form for a data-driven algorithm to compute the asymptotic phase map of an oscillator. Chapter IV contains our theoretical results on the existence and properties of an attracting invariant "slow manifold" for a class of smoothly locomoting mechanical systems, and the use of these results to inform a data-driven algorithm to approximate the dynamics restricted to the slow manifold. Our algorithm can be viewed as an improvement of the algorithm recently developed in [BHR18],

[7] The result in [EJ16, App. A] assumes that configuration space Q is compact. Our extension enables applications to, e.g., configuration spaces of the form $Q = S \times \mathsf{SE}(2)$ with symmetry group $\mathsf{SE}(2)$ the special Euclidean group of planar rigid motions and S compact. Such a situation often arises in practice.

which used regressors determined by the assumption of a kinematic “connection” model for the dynamics.

The “Discussion” sections in each of the following chapters contain discussions of chapter-specific content and suggestions for future work.

Appendix A contains a technical lemma used in Chapter II, which establishes smoothness of a parallel transport map covering a flow on an inflowing invariant manifold; most of the work expended is to obtain one extra degree of differentiability. Appendix B contains a general lemma showing that every inflowing compact NAIM can be embedded in a boundaryless compact NAIM, in such a way that properties like exponential rates are preserved; the proof combines differential topological and dynamical techniques. We use this lemma to prove linearization results for inflowing NAIMs in Chapter II, but it may be of independent interest. Next, Appendix C contains basic definitions and concepts from the theory of fiber bundles which are needed in Chapters II and IV. Appendix D contains variants of the weak law of large numbers (LLN) due to Chebyshev, including an “approximate” version of the law of large numbers. We use this approximate LLN in Chapter III for our analysis of the performance of our phase estimation algorithm on noisy data. Finally, Appendix E contains a proof of Grönwall’s Lemma which we use in our proof of global asymptotic stability of the time-dependent slow manifold, or “integral manifold”, in Chapter IV.

CHAPTER II

Global Linearization and Fiber Bundle Structure of Invariant Manifolds

2.1 Acknowledgements

This chapter is based on joint work with Jaap Eldering and Shai Revzen [EKR18]. Jaap Eldering performed the major share of his contribution to this work while holding a postdoctoral position at ICMC, University of São Paulo, São Carlos, Brazil, supported by FAPESP grant 2015/25947-6. Kvalheim and Revzen were supported by ARO grants W911NF-14-1-0573 and W911NF-17-1-0306 to Revzen. Kvalheim would like to thank Ralf Spatzier for many helpful conversations related to Theorem II.5. We also thank the two anonymous referees and the editors for finding a mistake in a lemma, for bringing the reference [JT09] to our attention, and for their many useful suggestions.

2.2 Introduction

Much of dynamical systems theory pertains to the behavior of points evolving under some smooth flow $\Phi : \mathbb{R} \times Q \to Q$ near an attracting invariant set. One seeks techniques to better understand the behavior of these points. Perhaps the most important method — and the focus of this chapter — is the use of different coordinate systems near the attracting set, with respect to which the dynamics take a simpler form. Particularly strong results in this direction hold in the case that the attracting invariant set is a normally attracting invariant manifold (NAIM). This is a special case of a normally hyperbolic invariant manifold (NHIM), which is roughly defined as follows. A manifold $M \subset Q$ is invariant if $\forall t \in \mathbb{R} : \Phi^t(M) = M$, and normal hyperbolicity means roughly that trajectories converge (or diverge) transversely to M sufficiently faster than they converge (or diverge) within M [Fen71, HPS77, Wig94].

Restricting, for now, to the case that M has no boundary, it is a well-known fundamental result [Fen74, HPS77, Thm 2, Thm 4.1] that a NAIM, as a special case of a NHIM, has an associated "stable foliation": this is a partition of the stability basin of M into submanifolds $W^s(m)$ for $m \in M$ (called "leaves") such that the flow Φ^t maps $W^s(m)$ to $W^s(\Phi^t(m))$, for any $t \geq 0$ and $m \in M$. Furthermore, every x in the stability basin of M has a neighborhood N_x such that N_x is topologically a product of two Euclidean spaces; the first space indexes leaves and the second space locally parametrizes them (see Figure 2.1, left). By using this foliation to define coordinates, one obtains a coordinate system in which the dynamics on M are decoupled from the dynamics transverse to M.

It is also well known [PS70, HPS77, PT77]— and often used in the physical sciences, e.g., in the special case of the Hartman-Grobman theorem [GH83] — that for M a NAIM (or NHIM), there exists an open neighborhood of M in which the flow is topologically conjugate to its partial linearization. For simplicity, we first describe this result in the special case that $M \subset Q$ has a neighborhood diffeomorphic to $M \times \mathbb{R}^n$ via a diffeomorphism which restricts to the identity on $M \times \{0\}$. In this case, we may write the flow Φ^t as (Φ_1^t, Φ_2^t) on $M \times \mathbb{R}^n$. Then this linearization result asserts the existence of a continuous change of coordinates $(p, v) \mapsto (q, w)$ on $M \times \mathbb{R}^n$ — which restricts to the identity on $M \times \{0\}$ — such that for any $(p, v) \in \mathbb{R}^n$, the trajectory $(\Phi_1^t(p, v), \Phi_2^t(p, v))$ is given by $(\Phi_1^t(q, 0), \mathsf{D}\Phi_2^t(q, 0) \cdot w)$ in the new coordinates. In these new coordinates on $M \times \mathbb{R}^n$, not only are the dynamics on M decoupled from the dynamics transverse to M, but the transverse component $w(t) \coloneqq \mathsf{D}\Phi_2^t(q,0)\cdot w$ of a trajectory is the solution of a nonautonomous linear differential equation. Under additional spectral gap assumptions, this coordinate change can be taken to be continuously differentiable [Tak71, Rob71, Sel84, Sel83, Sak94]. Needless to say, many key results in the sciences and engineering rely heavily on linear approximations of this form; this result shows that there exists a coordinate system in which such approximations become exact.

In this chapter we prove several extensions of the familiar local results mentioned above, which we hope to be of both practical and theoretical interest. Our results come in two flavors: (i) we show that the local topological and dynamical structure near the NAIM can be extended (often smoothly) to the entire stability basin, and (ii) we prove new local (and global) linearization results for NAIMs with nonempty boundary, subject to the requirement that the flow is "inward" at the boundary (inflowing NAIMs). The novelty of our results is that, to the best of our knowledge, all previously published work only established versions of our various results either (i) for hyperbolic attracting equilibria and periodic orbits rather than general NAIMs[1], (ii) for NAIMs without boundary, (iii) locally, or on proper subsets of the global stable manifold (in the case of a boundaryless NAIM, this is the stability basin), or some combination thereof. In contrast, our results apply to the entire global stable manifold, and they apply to the even broader class of systems consisting of inflowing NAIMs. Thus our theorems can be used to prove results on compact domains of noncompact attracting manifolds, which can arise (for example) as intersections of a noncompact M with a compact sub-level set of a function. Many noncompact hyperbolically attractive manifolds appear in the sciences and engineering, e.g., in the general context of slow-fast or multiple time scale systems [Kue15] studied using geometric singular perturbation theory

[1] It has recently come to our attention that in a soon-to-be published textbook [Mez17], Igor Mezić gives a very readable proof of a global linearization theorem for the case of arbitrary compact boundaryless NAIMs (see also Remark II.17). In contrast, we also prove a more general result for arbitrary compact inflowing NAIMs, which may have nonempty boundary.

(GSP) [Fen79, Jon95, Kap99]. With the addition of a proper function having a strictly negative Lie derivative on one of its regular level sets, these give rise to compact inflowing NAIMs. We remark that even if a compact domain of a noncompact attracting manifold is not inflowing, useful conclusions about the dynamics can sometimes still be obtained by making local modifications to the flow near the boundary of the domain in order to render it inflowing, and then applying theorems for inflowing NAIMs. We do precisely this in our applications to GSP in §2.6.

2.2.1 Flavor of the key results

We begin by examining the (differential) topology of the global stable manifold, in a form depicted in Figure 2.1 and formulated more precisely in Theorem II.5. We show that the entire global stable manifold of an inflowing NAIM has the structure of a "disk bundle": for M of dimension d in an n-dimensional ambient space, the global stable manifold admits a continuous "projection" onto M, and every point $m \in M$ has a neighborhood $U_m \subset M$ such that the preimage of U_m through the projection is homeomorphic to the product of U_m with $(n-d)$-dimensional Euclidean space $\mathbb{R}^{n-d}$ ("a disk"). Furthermore, projection preimages ("fibers") of points $m \in M$ are mapped via these homeomorphisms to sets of the form $\{m\} \times \mathbb{R}^{n-k}$. We further extend this result by proving that, should the foliation near M be C^k smooth, then the entire global stable manifold has a structure of a C^k disk bundle (for the definition, replace all homeomorphisms with C^k diffeomorphisms above). Anticipating our global linearization results, one can think of this result as a "weak" or differential-topological version of global linearization of the global stable manifold: the global stable manifold always has the (differential) topological structure that one would naively expect from the (differential) topological structure of the local stable manifold.

This result has an application to geometric singular perturbation theory related to the so-called Fenichel Normal Form [JK94, Jon95, Kap99, JT09]; for more details on the relevance of this normal form for slow-fast systems, see §2.6. In the special case that the slow manifold is attracting, we show that our Theorem II.5 implies that the domain of the Fenichel normal form actually extends to the entire *global* stable manifold of the slow manifold.

We then proceed beyond "weak" linearization to the natural follow-up question, and show that in addition to the local topological structure, the local dynamical structure near an inflowing NAIM also extends to the entire global stable manifold: the flow on the global stable manifold is topologically conjugate to its linearization near M, and assuming some conditions on the relative rates of contraction of tangent vectors at M evolving under the linearized flow, the global conjugacy of the flow to its linearization can be taken to be C^k. In addition to this statement being a new global result, to the best of our knowledge, the local version of this linearization result is also new: linearization results previously appearing in the literature [PS70, Rob71, PT77, HPS77, Sel84, Sel83, Sak94] have been stated for *boundaryless* invariant manifolds. This result provides a strong statement regarding how well dynamical systems can be modeled by their transverse linearizations and the dynamics on their attractor. We give an application of this to singular perturbation theory, where the "slow manifold" attractors typically have boundary. Thanks to our results for inflowing NAIMs we show that, under certain spectral conditions, singularly perturbed systems have a global normal form which is *linear* in the fast variables. This normal form is therefore stronger than the Fenichel Normal Form, which is generally (almost) fully nonlinear.

2.2.2 Overview of main results

Restated more technically, in this chapter we prove some results for NHIMs which are of two types.

1. Global versions of well-known local results for compact normally hyperbolic invariant manifolds (NHIMs), and compact, inflowing, normally attracting invariant manifolds

(inflowing NAIMs).

2. New (local and global) linearization results for inflowing NAIMs.

We restrict our attention to the case of flows on a finite-dimensional smooth ambient manifold. We first investigate the structure of the global stable foliation of a compact normally hyperbolic invariant manifold $M \subset Q$ for a flow Φ^t on a smooth manifold Q. We consider the following local-to-global result to be our first major contribution, depicted in Figure 2.1.

Theorem II.5′. *The global stable foliation of a NHIM is a topological disk bundle with fibers coinciding with the leaves of the foliation. If additionally the local stable foliation and the NHIM are assumed C^k, then the global foliation is a C^k disk bundle. This bundle is isomorphic (as a disk bundle) to the stable vector bundle over the NHIM. A similar result holds for the global unstable foliation.*

In particular, if the k-center bunching condition (see Corollary II.10 in §2.4) is assumed, it follows that the global stable foliation is a C^k disk bundle. If both stable and unstable transverse directions are present at M, then $W^s(M) \subset Q$ is generally only an immersed submanifold[2]. Hence our result shows that the global stable manifold is a fiber bundle in its manifold topology, but *not* in the subspace topology. However if only stable transverse directions at M are present, this technicality is avoided and $W^s(M) \subset Q$ is a fiber bundle whose topology coincides with the subspace topology. (Embedded and immersed submanifolds are explained in more detail in [Lee13, Ch. 5].)

We also prove the corresponding fiber bundle result for the global stable foliation of a compact inflowing normally attracting invariant manifold (NAIM) M. I.e., M is a NHIM with empty unstable bundle, but M is allowed to have nonempty boundary, and inflowing

[2]Roughly speaking, this is because — in the case that the unstable bundle is nonempty — $W^s(M)$ can accumulate on itself. This is analogous to the ends of a curve approaching its midpoint to form a figure-eight. The figure-eight is not an embedded submanifold, because the midpoint has no locally Euclidean neighborhood in the subspace topology, but the figure-eight is an immersed submanifold diffeomorphic to $\mathbb{R}$.

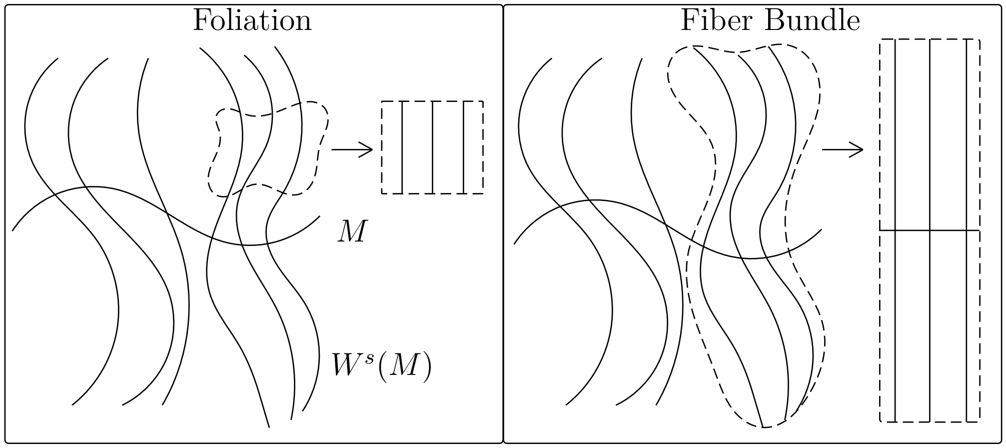


Figure 2.1: The fact that the global stable foliation $W^s(M)$ of a NHIM M is a topological foliation implies that any point in $W^s(M)$ has a neighborhood on which the leaves of the foliation can be straightened via a homeomorphism, depicted on the left. Theorem II.5 shows that $W^s(M)$ is actually a topological disk bundle, which means that $W^s(M)$ admits *local trivializations* whereby unions of entire fibers through a neighborhood of M can be straightened via a homeomorphism, depicted on the right. Furthermore, if $W^s_{\text{loc}}(M)$ is a C^k foliation and $M \in C^k$, then $W^s(M)$ is actually a C^k fiber bundle, which means that these local trivializations can be chosen C^k.

means that M is positively invariant and that the vector field points strictly inward at ∂M. This is the result we actually prove, and indeed, the previously mentioned result follows from this one.

While our fiber bundle result might be expected by dynamicists, we could not find a direct proof in the literature. If the stable foliation happens to be smooth, then we will show that the map sending leaves to their basepoints on M is a submersion with fibers diffeomorphic to disks, and it is shown in [Mei02, Cor. 31] that this automatically implies that the stable manifold $W^s(M)$ is a smooth disk bundle. On the other hand, our proof seems more elementary, directly shows that this bundle is isomorphic to E^s, and handles the general case in which the stable foliation is only continuous.

Next, we investigate global linearizations. A classic result of NHIM theory is that the dynamics in a neighborhood of a NHIM are topologically conjugate to the dynamics linearized at the NHIM [PS70, HPS77, PT77], and there also exist conditions for C^k linearization

[Tak71, Rob71, Sak94, Sel84, Sel83, BK94]. For the special case of a NAIM which is either an equilibrium point or a periodic orbit, [LM13] showed that the linearizing conjugacy can be defined on the entire basin of attraction. We generalize the results of [LM13] in two ways: (i) we show that local linearizability implies global linearizability for arbitrary compact NAIMs[3], and (ii) we prove a global linearizability result for inflowing NAIMs. Since the slow manifolds for slow-fast systems typically have boundary, the latter result is necessary for our goal of deriving a linear normal form for such systems, and we consider this to be our second major theoretical contribution. We state this result roughly below (for the precise statement, see Theorem II.24 in §2.5.2). Recall that the global stable manifold is the basin of attraction in the case of a boundaryless NAIM. For the precise definition of the global stable manifold of an inflowing NAIM, see Equation (2.5) in §2.3.1.

Theorem II.24′. *The dynamics on the global stable manifold of an inflowing NAIM are globally topologically conjugate to the dynamics linearized at the NAIM. If certain additional spectral gap and regularity conditions are assumed, then additionally the dynamics are globally C^k conjugate to the dynamics linearized at the NAIM.*

In order to prove this result, we use a geometric construction in Appendix B which may be of independent interest. Generally speaking, in Appendix B, we show that any compact inflowing NAIM can be embedded into a compact boundaryless NAIM, in such a way that many properties of the original system are preserved, such as asymptotic rates.

After proving these results, we give two applications to geometric singular perturbation theory, under the assumption that the critical manifold is a NAIM (see the references in §2.6 for examples, as well as §2.6.5). Our first application is to show that under this assumption, the Fenichel Normal Form appearing in the literature is valid on the entire union of *global* stable manifolds $\cup_\epsilon W^s(K_\epsilon)$ of the slow manifolds K_ϵ, rather than just on the union of local

[3] As mentioned in a previous footnote, Igor Mezić gives a proof of this boundaryless result in his soon-to-be published textbook [Mez17] (see also Remark II.17).

stable manifolds $\cup_\epsilon W^s_{\text{loc}}(K_\epsilon)$. Our second application is to show that, assuming an additional "nonresonance" condition on the eigenvalues of points on the critical manifold and using our global linearization theorem, we derive a much stronger global normal form which is *linear* in the fast variables. We reiterate that a linearization result for inflowing NAIMs is essential here, since the slow manifolds for singular perturbation problems typically have boundary.

The remainder of the chapter is organized as follows. In §2.3 we give basic definitions, set notation, and give basic constructions to be used in the sequel. In particular, we construct the global stable foliation of a NHIM and show that the local stable foliation is a fiber bundle, and remark that the same constructions work for inflowing NAIMs. In §2.4 we give the proof that, if the local stable foliation and the NHIM are C^k, then the global stable foliation is a C^k fiber bundle isomorphic (as a disk bundle) to E^s. In §2.5, we show that the dynamics on the global stable manifold of an inflowing NAIM are globally conjugate to the linearized dynamics, and other related results. In §2.6, we give applications to geometric singular perturbation theory. In §3.7 we conclude by summarizing what we have and have not done.

2.3 Preliminary constructions

2.3.1 Construction of the global (un)stable foliation of a NHIM

Let Q be an n-dimensional C^∞ Riemannian manifold, let $f\colon Q \to \mathsf{T}Q$ be a $C^{r\geq 1}$ vector field on Q with C^r flow Φ^t and let $M \subset Q$ be a compact r-normally hyperbolic invariant manifold (r-NHIM) for Φ^t. We recall from [HPS77] the definition; specifically we use the most general definition of *eventual relative normal hyperbolicity*. This means that M is a submanifold that is invariant under Φ^t, and there exists a $\mathsf{D}\Phi^t$-invariant continuous splitting into a Whitney sum

$$\mathsf{T}Q|_M = \mathsf{T}M \oplus E^s \oplus E^u \tag{2.1}$$

such that $\mathsf{D}\Phi^t|_{E^s}$ and $\mathsf{D}\Phi^t|_{E^u}$ are exponentially contracting and expanding, respectively. (See Appendix C for the definitions of vector bundles and Whitney sums.) Furthermore, any contraction or expansion of the tangential flow $\mathsf{D}\Phi^t|_{\mathsf{T}M}$ (up to power r) is dominated by the contraction of $\mathsf{D}\Phi^t|_{E^s}$, respectively the expansion of $\mathsf{D}\Phi^t|_{E^u}$. More precisely, there exist $C > 0$ and $a < 0 < b$ such that for all $m \in M$, $t \geq 0$ and $0 \leq i \leq r$ we have

$$\lfloor\!\lfloor \mathsf{D}\Phi^t|_{E^u_m} \rfloor\!\rfloor \geq \frac{e^{bt}}{C} \|\mathsf{D}\Phi^t|_{\mathsf{T}_m M}\|^i \qquad \text{and} \qquad \|\mathsf{D}\Phi^t|_{E^s_m}\| \leq C e^{at} \lfloor\!\lfloor \mathsf{D}\Phi^t|_{\mathsf{T}_m M} \rfloor\!\rfloor^i. \tag{2.2}$$

Here $\lfloor\!\lfloor A \rfloor\!\rfloor \coloneqq \inf\{\|Av\| : \|v\| = 1\}$ denotes the minimum norm of a linear operator A.

Denote by $n = n_m + n_s + n_u$ the ranks of the various bundles and note that $n_m = \dim(M)$. Since the stable and unstable cases are identical under time reversal, we restrict ourselves from now on to the stable case. Tangent to the stable bundle E^s there exists a *local stable manifold* $W^s_{\mathrm{loc}}(M)$, a C^r embedded submanifold[4], with points in $W^s_{\mathrm{loc}}(M)$ asymptotically converging to M in forward time. $W^s_{\mathrm{loc}}(M)$ is invariantly fibered by embedded disks $W^s_{\mathrm{loc}}(m)$ comprising the leaves/fibers of the *local stable foliation*:

$$W^s_{\mathrm{loc}}(M) = \coprod_{m \in M} W^s_{\mathrm{loc}}(m) \tag{2.3}$$

such that $W^s_{\mathrm{loc}}(m)$ intersects M at the unique point m and $\mathsf{T}_m W^s_{\mathrm{loc}}(m) = E^s_m$, see [HPS77, Thm 4.1]. Each of the disks $W^s_{\mathrm{loc}}(m)$ is individually a C^r embedded submanifold, but as a family there is generally only (Hölder) continuous dependence on the basepoint $m \in M$

[4] We will always assume without loss of generality that $W^s_{\mathrm{loc}}(M)$ has no boundary. Otherwise we can simply relabel its manifold interior as $W^s_{\mathrm{loc}}(M)$.

[HPS77, Fen74]. We denote by $P^s_{\mathrm{loc}} : W^s_{\mathrm{loc}}(M) \to M$ the continuous projection map sending each fiber $W^s_{\mathrm{loc}}(m)$ to its corresponding basepoint $m \in M$. Note that the $W^s_{\mathrm{loc}}(m), m \in M$ are only invariant as a foliation — not each $W^s_{\mathrm{loc}}(m)$ individually — since each $m \in M$ is generally not a fixed point of Φ^t. This local invariance of the foliation $W^s_{\mathrm{loc}}(M)$ means that for all $t \geq 0$ and $m \in M$ we have[5]

$$\Phi^t(W^s_{\mathrm{loc}}(m)) \subset W^s_{\mathrm{loc}}(\Phi^t(m)). \tag{2.4}$$

We also have a *global stable manifold* $W^s(M) \supset W^s_{\mathrm{loc}}(M)$ defined by[6]

$$W^s(M) \coloneqq \bigcup_{t \geq 0} \Phi^{-t}\left[(P^s_{\mathrm{loc}})^{-1}(\Phi^t(M))\right]. \tag{2.5}$$

Each of the sets $\Phi^{-t}\left[(P^s_{\mathrm{loc}})^{-1}(\Phi^t(M))\right]$ is an embedded submanifold of Q (diffeomorphic to $W^s_{\mathrm{loc}}(M)$), and thus $W^s(M)$ is a C^r immersed submanifold of Q when given the final topology with respect to the family of inclusions $\Phi^{-t}\left[(P^s_{\mathrm{loc}})^{-1}(\Phi^t(M))\right] \hookrightarrow W^s(M)$. An atlas of charts for $W^s(M)$ consists of the union of atlases for all of the manifolds $\Phi^{-t}\left[(P^s_{\mathrm{loc}})^{-1}(\Phi^t(M))\right]$ — since the flow Φ^t is C^r, it can be checked that this is a C^r atlas.

Let us now construct a global stable foliation as

$$W^s(M) = \coprod_{m \in M} W^s(m), \qquad W^s(m) \coloneqq \bigcup_{t \geq 0} \Phi^{-t}(W^s_{\mathrm{loc}}(\Phi^t(m))). \tag{2.6}$$

[5] When *immediate* relative normal hyperbolicity is assumed (as in [HPS77, Thm 4.1]) then $W^s_{\mathrm{loc}}(M)$ is automatically forward invariant when it has constant diameter. In the case of *eventual* relative normal hyperbolicity, standard proofs construct $W^s_{\mathrm{loc}}(M)$ as the local stable manifold of the map Φ^T for some fixed $T > 0$, so it might not be clear a priori that the inclusion holds for *all* $t \geq 0$, though it is clear that it would hold for t sufficiently large. However, we can always construct a *new* $W^s_{\mathrm{loc}}(M)$ that is forward invariant for all $t \geq 0$, as a sublevel set of a Lyapunov function for M, see [Wil67, Wil69].

[6] This definition works equally well for inflowing NAIMs (see §2.3.3), as opposed to the alternative definition $W^s(M) \coloneqq \bigcup_{t \geq 0} \Phi^{-t}(W^s_{\mathrm{loc}}(M))$.

Note that equation (2.4) implies that the union consists of strictly increasing sets, i.e.,

$$\Phi^{-t}(W^s_{\mathrm{loc}}(\Phi^t(m))) \subset \Phi^{-t'}(W^s_{\mathrm{loc}}(\Phi^{t'}(m))) \quad \text{when } t \leq t'. \tag{2.7}$$

Let us prove that $W^s(M)$ is invariant, that is, for all $t \in \mathbb{R}$ and $m \in M$ we have

$$\Phi^t(W^s(m)) = W^s(\Phi^t(m)). \tag{2.8}$$

This follows from the following sequence of equivalent statements, with $t \in \mathbb{R}$ fixed:

$$\begin{aligned}
& x \in W^s(\Phi^t(m)) \\
\exists \tau_0 \geq 0\colon \forall \tau \geq \tau_0\colon \quad & x \in \Phi^{-\tau}(W^s_{\mathrm{loc}}(\Phi^\tau \circ \Phi^t(m))) \\
\exists \tau_0' \geq 0\colon \forall \tau' \geq \tau_0'\colon \quad & x \in \Phi^t \circ \Phi^{-\tau'}(W^s_{\mathrm{loc}}(\Phi^{\tau'}(m))) \\
& x \in \Phi^t(W^s(m)).
\end{aligned}$$

Note that each global leaf $W^s(m)$ is a C^r embedded submanifold of $W^s(M)$. To see this, note that given any $m \in M$ and $x \in W^s(m)$, by definition of the global foliation there exists $t > 0$ such that $\Phi^t(x) \in W^s_{\mathrm{loc}}(\Phi^t(m))$. Letting U' be a neighborhood of $\Phi^t(x)$ in $W^s_{\mathrm{loc}}(M)$ and considering $U \coloneqq \Phi^{-t}(U') \ni x$, we see that any point $x \in W^s(m)$ has a neighborhood $U \subset W^s(M)$ with $U \cap W^s(m)$ an embedded submanifold of $W^s(M)$ (by invariance of the foliation), so it follows that $W^s(m)$ is embedded in $W^s(M)$. (But since $W^s(M)$ is generally only immersed in Q, any global leaf $W^s(m)$ is generally only immersed in Q.)

We define the global projection $P^s\colon W^s(M) \to M$ to be the map that sends global fibers $W^s(m)$ to their basepoints, just like the local projection P^s_{loc}. Assume now that the local stable foliation is $C^{k \geq 0}$, by which we mean that $P^s_{\mathrm{loc}} \in C^k$. (Recall that $P^s_{\mathrm{loc}} \in C^0$ automatically.) We now show that this implies $P^s \in C^k$.

Let $x \in W^s(M)$ and $t \geq 0$ be such that $x \in \Phi^{-t}(W^s_{\text{loc}}(M))$. This implies that $x' = \Phi^t(x) \in W^s_{\text{loc}}(M)$. Choose a neighborhood $U_{x'}$ of x' in $W^s_{\text{loc}}(M)$. Then $U_x \coloneqq \Phi^{-t}(U_{x'})$ is a neighborhood of x in $W^s(M)$. Now for any $y \in U_x$ we have that $\Phi^t(y) \in W^s_{\text{loc}}(M)$ and by invariance of the local stable foliation it follows that

$$P^s(y) = (\Phi^{-t} \circ P^s_{\text{loc}} \circ \Phi^t)(y). \tag{2.9}$$

Thus it is clear that $P^s \in C^k$ if $P^s_{\text{loc}} \in C^k$ and $k \leq r$ (i.e., $\Phi^t \in C^k$).

We conclude this section by showing that, not only is $P^s \in C^k$ if $P^s_{\text{loc}} \in C^k$, but also that P^s is a submersion if $k \geq 1$.

Proposition II.1. *If P^s_{loc} is C^k with $1 \leq k \leq r$, then $P^s \colon W^s(M) \to M$ is a C^k submersion.*

Proof. We have already shown above that $P^s \in C^k$, so it suffices to show that $\operatorname{rank}(\mathsf{D}P^s_{\text{loc}}|_{\mathsf{T}M}) = \dim(M)$ on all of $W^s(M)$. Since $P^s_{\text{loc}}|_M = \text{id}_M$, it follows that $\operatorname{rank}(\mathsf{D}P^s_{\text{loc}}|_{\mathsf{T}M}) = \dim(M)$. Since being full rank is an open condition, it follows that $\mathsf{D}P^s_{\text{loc}}$ is full rank on some relatively open neighborhood U of M in $W^s_{\text{loc}}(M)$.

Now let $x \in W^s(M)$ be arbitrary. First, by construction of $W^s(M)$ there exists a $T_1 > 0$ such that $\Phi^{T_1}(x) \in W^s_{\text{loc}}(M)$. Next, since every point in $W^s_{\text{loc}}(M)$ asymptotically converges to M, there exists $T_2 > 0$ such that $\Phi^{T_2}(\Phi^{T_1}(x)) \in U$. Defining $T \coloneqq T_1 + T_2 > 0$, we have $\Phi^T(x) \in U$.

Since $\forall t \in \mathbb{R} \colon P^s \circ \Phi^t = \Phi^t \circ P^s$, it follows that $\mathsf{D}\Phi^T_{P^s(x)} \mathsf{D}P^s_x = \mathsf{D}P^s_{\Phi^T(x)} \mathsf{D}\Phi^T_x = \mathsf{D}(P^s_{\text{loc}})_{\Phi^T(x)} \mathsf{D}\Phi^T_x$. The latter composition is formed of two surjective linear maps, and hence $\mathsf{D}\Phi^T_{P^s(x)} \mathsf{D}P^s_x$ is also a surjective linear map. The linear map $\mathsf{D}\Phi^T|_{P^s(x)}$ is invertible since Φ^T is a diffeomorphism, so this implies that $\mathsf{D}P^s_x \colon \mathsf{T}_x W^s(M) \to \mathsf{T}_{P^s(x)} M$ is surjective. ☐

2.3.2 Fiber bundle structure of the local stable foliation

Let $\pi \colon \mathsf{T}Q|_M \to M$ be the natural projection sending $v \in \mathsf{T}_m Q$ to m, and let $\widetilde{E}^s$ be any C^r subbundle of $\mathsf{T}Q|_M$ which C^0 approximates E^s [Wig94, p.72 Prop. 3.2.3]. (Recall that E^s is generally only a continuous subbundle of $TQ|_M$.) As shown in [Fen74, HPS77] there exists a fiber-preserving homeomorphism $\rho_0 \colon U \subset \widetilde{E}^s \to W^s_{\text{loc}}(M)$, where $U \subset \widetilde{E}^s$ is a connected neighborhood of the zero section. Additionally, the restriction of ρ_0 to each fiber $\widetilde{E}^s_m$ is a C^r map. Here we show that if additionally the local stable foliation of $W^s_{\text{loc}}(M)$ is $C^{k\geq 1}$, then ρ_0 can be taken to be a C^k fiber-preserving diffeomorphism.

Fiber bundle concepts from Appendix C (in particular, Definition C.4 and Example C.7) will be used in the proof of Lemma II.2 below. Here and in the rest of the chapter, by a C^k isomorphism of manifolds we mean a homeomorphism if $k = 0$ and a C^k diffeomorphism if $k \geq 1$. A C^k fiber bundle isomorphism is a C^k isomorphism of manifolds which is also fiber-preserving; see Appendix C.

Lemma II.2. *Let M be a 1-NAIM, and assume that $P^s_{\text{loc}} \in C^k$ (hence $W^s_{\text{loc}}(M), M \subset Q$ are necessarily C^k submanifolds). Then $P^s_{\text{loc}} \colon W^s_{\text{loc}}(M) \to M$ is a disk bundle. More specifically, there exists a neighborhood U of the zero section of $\widetilde{E}^s$ and a C^k disk bundle isomorphism $\rho_0 \colon U \to W^s_{\text{loc}}(M)$ covering id_M (identifying M with the zero section of $\widetilde{E}^s$).*

Remark II.3. If M is an r-NAIM for a C^r vector field, then M and $W^s_{\text{loc}}(M)$ are automatically C^r submanifolds of Q (and hence $W^s(M)$ is an immersed C^r submanifold, as we have shown). See [Eld13, Ch. 1] for a discussion of this. We will use Lemma II.2 in proving Theorem II.5 for a 1-NAIM which is a C^r submanifold — a slightly more general situation than an r-NAIM — which explains the slightly weaker hypotheses here.

Proof. As mentioned, for $k = 0$ the result is shown in [Fen74, HPS77] so we may assume $k \geq 1$. The latter case is implicit in the existing proofs of C^k smoothness of local stable fibers, but we make it explicit here for later reference. Consider the extended exponential map[7]

$$\widehat{\exp} = (\pi, \exp)\colon \mathsf{T}Q|_M \to M \times Q \tag{2.10}$$

that remembers the base point $m \in M$. This is a fiber bundle isomorphism between a neighborhood of the zero section of $\mathsf{T}Q|_M$ and a neighborhood of $\operatorname{diag}(M)$ in the trivial bundle $M \times Q$, covering the identity on M, where the zero section of $\mathsf{T}Q|_M$ and $\operatorname{diag}(M)$ are identified with M. Furthermore, we view $W^s_{\text{loc}}(M) \subset Q$ as a C^k submanifold of $M \times Q$ via the embedding (P^s, id_Q), fibered by the images of the leaves $W^s(m)$. It follows that $\widehat{\exp}^{-1}\Big(W^s_{\text{loc}}(M)\Big)$ is a submanifold of $\mathsf{T}Q|_M$ fibered by the leaves $\widehat{\exp}^{-1}(W^s_{\text{loc}}(m))$, and the leaf $\widehat{\exp}^{-1}\Big(W^s_{\text{loc}}(m)\Big) \subset \mathsf{T}_m Q$ is tangent to E^s_m at the zero section since the derivative of $\widehat{\exp}|_{\mathsf{T}Q_m}$ at 0 is the identity for any $m \in M$. Here we are making the usual linear identification $\mathsf{T}_0\mathsf{T}_m Q \cong \mathsf{T}_m Q$.

Let $\tilde{\pi}^s\colon \mathsf{T}Q|_M \to \widetilde{E}^s$ denote orthogonal projection onto $\widetilde{E}^s$. We have that $\tilde{\pi}^s \in C^r$, and when $\widetilde{E}^s$ is sufficiently C^0-close to E^s then $\ker(\tilde{\pi}^s)$ and E^s are transverse. Thus $\mathsf{D}\Big(\tilde{\pi}^s \circ \widehat{\exp}^{-1}|_{W^s_{\text{loc}}(m)}\Big)$ is surjective for each $m \in M$, so by dimension counting this map is a linear bijection between $E^s_m \oplus \mathsf{T}_m M \cong \mathsf{T}_m W^s_{\text{loc}}(M)$ and $\widetilde{E}^s_m \oplus \mathsf{T}_m M \cong \mathsf{T}_m \widetilde{E}^s$ for each $m \in M$.

The global inverse function theorem [GP10, § 1.8 ex. 14] now implies that $\tilde{\pi}^s \circ \widehat{\exp}^{-1}|_{W^s_{\text{loc}}(M)}$ is a C^k diffeomorphism from some neighborhood of $\operatorname{diag}(M)$ onto a neighborhood U of the zero section of $\widetilde{E}^s$. Thus the inverse

$$\rho_0\colon U \to W^s_{\text{loc}}(M)$$

is well-defined and is a fiber-preserving C^k diffeomorphism onto its image. By construction it maps the zero section to M and covers the identity map. □

[7]Recall that we have endowed Q with a Riemannian metric, used in the definition of spectral gap estimates (2.2). Here — and throughout the rest of the chapter — we have in mind the exponential map associated to this Riemannian metric, although for the purposes of this Lemma, any metric will work equally well.

2.3.3 Inflowing and overflowing NAIMs

Suppose now that M is a compact manifold but that M has possibly nonempty boundary, $\partial M \neq \varnothing$. If $\Phi^t(M) \subset M$ for all $t \geq 0$ and the vector field f points strictly inward at ∂M, we call M an *inflowing* invariant manifold. Similarly, if $\Phi^t(M) \subset M$ for all $t \leq 0$ and the vector field f points strictly outward at ∂M, we call M an *overflowing* invariant manifold. If M is inflowing (respectively overflowing) invariant and has a splitting (2.1) satisfying exponential rates (2.2), but with $E^u = \varnothing$, we call M an *inflowing (respectively overflowing) r-normally attracting invariant manifold (r-NAIM).* If $\partial M = \varnothing$ and M is invariant, then M is vacuously both inflowing and overflowing. We refer to such an M simply as an r-NAIM. We sometimes use the term "NAIM" to refer to 1-NAIMs or if we do not wish to emphasize the precise degree of hyperbolicity, and we similarly sometimes use "NHIM".

The main theorem about inflowing NAIMs is that, like boundaryless NHIMs, inflowing NAIMs also have a local stable manifold (with boundary) and a local stable foliation [Fen74, Fen71]. Note that in this case the local stable manifold has boundary, is codimension-0, and its manifold interior is an open neighborhood of the manifold interior of M. Additionally, the interior of the global stable manifold is open in Q and a neighborhood of the manifold interior of the NAIM. Unlike boundaryless NHIMs, however, inflowing NAIMs do not generally persist under perturbations.

The main theorem about overflowing NAIMs is that, like boundaryless NHIMs, overflowing NAIMs persist under perturbations [Fen71]. We will use this fact in §2.6. Unlike boundaryless NHIMs, however, overflowing NAIMs do not generally possess a stable foliation.

Remark II.4. If $P^s_{\text{loc}} \in C^k$ for an inflowing NAIM, then the same proof as for boundaryless NHIMs shows that $P^s \in C^k$ also. Furthermore, Proposition II.1 and Lemma II.2 also hold for inflowing NAIMs. The proof of Lemma II.2 is identical. For the proof of Proposition II.1, one simply pays attention to the facts that (i) since M is positively invariant, points never leave the stable foliation over M when flowing forward in time, and (ii) if $\Phi^t(x) \in W^s(m)$ for $t > 0$, $x \in W^s(M)$, and $m \in M$, then $\Phi^{-t}(m) \in M$. Additionally, the same argument given in §2.3.1 shows that each global fiber $W^s(m)$ is now an embedded submanifold of Q, since the manifold interior of $W^s(M)$ is open in Q and thus trivially embedded. This argument works even for $W^s(m)$ with $m \in \partial M$, since inflowing invariance implies that $\Phi^t(m) \in \operatorname{int} M$ for $t > 0$, with $\operatorname{int} M$ denoting the manifold interior of M.

2.4 The global stable foliation of a NHIM is a fiber bundle

As mentioned in §2.3.1, $P^s_{\text{loc}}\colon W^s_{\text{loc}}(M) \to M$ is in general only (Hölder) continuous. However, in many cases of interest $P^s_{\text{loc}}\colon W^s_{\text{loc}}(M) \to M$ is $C^{k\geq 1}$ (and thus $P^s\colon W^s(M) \to M$ is also C^k as shown in §2.3). In this section we prove that if $P^s_{\text{loc}} \in C^{k\geq 0}$, then $P^s\colon W^s(M) \to M$ is a C^k fiber bundle with fiber $\mathbb{R}^{n_s}$. See Appendix C for the relevant fiber bundle concepts. By reversing time the corresponding result that the global unstable manifold is a fiber bundle follows.

The topology on $W^s(M)$ compatible with its fiber bundle structure is generally finer than the subspace topology induced from Q since $W^s(M)$ is generally only an immersed submanifold of Q, as discussed in §2.3. Consequently, the individual fibers $W^s(m)$ of $W^s(M)$ are generally also only immersed submanifolds of Q, though they are embedded submanifolds of $W^s(M)$ as we have seen in §2.3.

However if M is a NAIM so that $E^u = \varnothing$, then M is asymptotically stable and $W^s(M)$ is an open neighborhood of M, hence trivially an embedded submanifold. More generally, if M has boundary and is an inflowing NAIM, then $W^s(M)$ is an embedded codimension-0 submanifold with boundary[8]. Every boundaryless NHIM M is a NAIM for the dynamics

[8]However, note that the boundary of $W^s(M)$ is only C^k if $P^s_{\text{loc}} \in C^k$, and hence generally not smooth if $P^s_{\text{loc}} \in C^0$ only.

restricted to the invariant manifold $W^s(M)$, and similarly M is a NAIM for the time-reversed dynamics restricted to $W^u(M)$. Hence it suffices to prove that $W^s(M)$ is a fiber bundle over M for the case that M is a NAIM.

To obtain the generality needed for our application in §2.6, we actually prove that $W^s(M)$ is a fiber bundle for M an inflowing NAIM — since a boundaryless NAIM is vacuously inflowing, this implies the other results.

See [Hir94, Ch. 2] for the definition of the Whitney topologies, and also Remark II.6 below.

Theorem II.5. *Let $M \subset Q$ be a compact inflowing 1-NAIM for the flow Φ^t generated by the C^r vector field f on Q, and assume that $M \subset Q$ is a C^r submanifold. Further assume that the local projection $P^s_{\text{loc}}: W^s_{\text{loc}}(M) \to M$ is C^k, with $0 \le k \le r$. Then the global projection $P^s: W^s(M) \to M$ defines a C^k fiber bundle structure on the global stable manifold $W^s(M)$. Furthermore, $W^s(M)$ is C^k isomorphic (as a disk bundle) to any C^k vector bundle over M which approximates E^s in the C^0 Whitney topology.*

In other words, under the hypotheses of Theorem II.5, the global stable foliation $W^s(M)$ is actually a C^k disk bundle.

Remark II.6. Since M is compact, the weak and strong Whitney topologies coincide [Hir94, Ch. 2]. In simpler terms [Wig94, p.72], $\tilde{E}^s$ approximates E^s in the C^0 Whitney topology if there exists a sufficiently small $\epsilon > 0$ such that for every $m \in M$, there exists a neighborhood $U_m \subset M$ of m and local frames $(e_i)_{i=1}^{n_s}$, $(\tilde{e}_i)_{i=1}^{n_s}$ for $E^s, \tilde{E}^s$ such that for all $m' \in U_m$: $\|e_i(m') - e_i(m')\| < \epsilon$.

Remark II.7. Let us reiterate Remark II.3. If M is an r-NHIM for a C^r vector field, then M and $W^s_{\text{loc}}(M)$ are automatically C^r submanifolds of Q. That is, an invariant manifold M being r-normally hyperbolic causes "forced C^r smoothness" of M and of the local and global stable manifolds $W^s_{\text{loc}}(M)$ and $W^s(M)$ [Eld13, Ch. 1]. (Of course this is C^r smoothness of $W^s_{\mathrm{loc}}(M)$ and $W^s(M)$ as submanifolds, not as foliations.) We state Theorem II.5 for a 1-NAIM M which is also assumed to be a C^r submanifold, in order to obtain a slight amount of extra generality.

Remark II.8. The hypotheses required to prove the global linearization Theorems II.18 and II.24 are much stronger than the hypotheses required to prove the fiber bundle Theorem II.5. See Remark II.20 for more details.

Corollary II.9. *Let $M \subset Q$ be a compact inflowing 1-NAIM for the flow Φ^t generated by the C^1 vector field f on Q. Then $P^s\colon W^s(M) \to M$ defines a C^0 fiber bundle structure on $W^s(M)$, isomorphic (as a disk bundle) to E^s.*

Proof. As mentioned earlier, $P^s_{\mathrm{loc}}, E^s, W^s_{\mathrm{loc}}(M) \in C^0$ is automatically satisfied for a compact inflowing 1-NAIM. Hence the result follows from Theorem II.5. □

Corollary II.10. *Let $M \subset Q$ be a compact inflowing 1-NAIM for the flow Φ^t generated by the C^r vector field f on Q, and assume that $M \subset Q$ is a C^r submanifold. Additionally, assume that there exist constants $K > 0$ and $\alpha < 0$ such that for all $m \in M$, $t \geq 0$ and $0 \leq i \leq k < r$ the k-center bunching condition holds:*

$$\|\mathsf{D}\Phi^t|_{T_mM}\|^i \|\mathsf{D}\Phi^t|_{E^s_m}\| \leq K e^{\alpha t} \llbracket \mathsf{D}\Phi^t|_{T_mM} \rrbracket. \tag{2.11}$$

Then $P^s\colon W^s(M) \to M$ defines a C^k fiber bundle structure on $W^s(M)$, isomorphic (as a disk bundle) to E^s.

Proof. It is shown in [Fen77, Thm 5] that the condition (2.11) implies that $P^s_{\mathrm{loc}}, E^s \in C^k$. The result then follows from Theorem II.5. □

Corollary II.11. *Assume now that M is a general compact r-NHIM, rather than a C^r inflowing 1-NAIM as in Theorem II.5, and assume that $P^s_{\mathrm{loc}} \in C^k$. Then $P^s\colon W^s(M) \to M$*

defines a C^k fiber bundle structure on $W^s(M)$, when $W^s(M)$ is endowed with the structure of an immersed submanifold as described in §2.3.1. This bundle is isomorphic (as a disk bundle) to any C^k vector bundle over M which approximates E^s.

Similarly for the unstable manifold $W^u(M)$, if $P^u_{\text{loc}} \in C^k$.

Proof. This follows immediately from Theorem II.5 and the remarks preceding it. □

Remark II.12. We leave it to the reader to formulate corollaries analogous to Corollaries II.9 and II.10 for the case of general compact NHIMs.

We assume that $M \in C^r$ is an inflowing 1-NAIM for the remainder of §2.4, unless stated otherwise.

2.4.1 Overview of the proof of Theorem II.5

(Recall that by a C^k isomorphism of manifolds, we mean a homeomorphism if $k = 0$ and a C^k diffeomorphism if $k \geq 1$. A C^k fiber bundle isomorphism is a C^k isomorphism of manifolds which is also fiber-preserving, and a C^k vector bundle isomorphism is a C^k fiber bundle isomorphism which is linear on the fibers; see Appendix C.)

By Lemma II.2 and Remark II.4, we have a C^k isomorphism of fiber bundles $\rho_0 \colon U \subset \widetilde{E}^s \to W^s_{\text{loc}}(M)$, where $\widetilde{E}^s$ is a vector bundle approximating E^s and $U \subset \widetilde{E}^s$ is open. We will construct a global C^k fiber-preserving isomorphism $\rho \colon \widetilde{E}^s \to W^s(M)$ using the local version $\rho_0 \colon U \to W^s_{\text{loc}}(M)$, according to the following plan. (It might be useful to first read Definition C.4 and Example C.7 from Appendix C.)

First, we define a flow[9] $\Psi^t := \rho_0^*(\Phi^t) = \rho_0^{-1} \circ \Phi^t \circ \rho_0$ on a neighborhood of M contained

[9] For simplicity of presentation, we henceforth ignore the fact that Ψ^t and other "flows" that we subsequently define, such as Θ^t, have possibly smaller domains of definition due to the fact that M is only assumed inflowing invariant. The domains of these "flows" always contain an appropriate neighborhood of $M \times \mathbb{R}_{\geq 0}$, and we only flow backwards in time along one "flow" after flowing forward by an equal time along another appropriate "flow"; as an example, consider equation (2.12). We will still call these objects "flows", and it should be clear what is meant when discussing such objects defined on a bundle over an inflowing invariant manifold.

in U. Adapting a technique of [PS70], we will find a C^k Lyapunov function $V\colon \widetilde{E}^s \to [0,\infty)$ for Ψ^t, such that $V^{-1}(0) = M$, the sublevel set $U_c := V^{-1}(-\infty, c)$ is positively invariant for all $0 < c \leq 1$, and V is strictly decreasing along trajectories starting in such level sets. Furthermore, V will be radially monotone (i.e., $V(\delta y) > V(y)$ if $\delta > 1$), and therefore it will have the nice property that any of its level sets intersect radial rays in each fiber $\widetilde{E}^s_m$ in precisely one point. This enables us to define a family of radial retractions $R_c\colon \widetilde{E}^s \setminus M \to V^{-1}(c)$ onto level sets of V, and we will show that this family is C^k.

We next construct a C^k flow Θ^t on $\widetilde{E}^s \setminus M$ that preserves level sets of V and covers $\Phi^t|_M$. $\Theta^t(y)$ is defined to be $R_{V(y)} \circ \Pi^t(y)$, where the radial retraction family R_c is as defined above, and Π^t is the smooth linear parallel transport covering $\Phi^t|_M$, constructed in Appendix A.

The global C^k isomorphism $\rho\colon \widetilde{E}^s \to W^s(M)$ is now constructed as follows. First, to a point $x \in \widetilde{E}^s$ we assign a time $t(x)$ roughly proportional to the value $V(x) > 0$, but with $t \equiv 0$ on a neighborhood of M. Next, we use the family of retractions R_c to construct a (nonlinear) rescaling diffeomorphism that maps $\widetilde{E}^s$ diffeomorphically onto $U_1 := V^{-1}(-\infty, 1)$, with the image of x denoted $\xi(x) \in U_1$. Finally, we define ρ by first flowing $\xi(x)$ forward by $\Theta_{t(x)}$, applying ρ_0, and then flowing backward in time by applying $\Phi^{-t(x)}$:

$$\rho(x) = \Phi^{-t(x)} \circ \rho_0 \circ \Theta^{t(x)} \circ \xi(x). \tag{2.12}$$

See Figure 2.2. The map ρ is C^k and fiber-preserving by construction. Properness of ρ will follow from asymptotic stability of M, and this will in turn imply surjectivity of ρ. The map ρ will be injective on V level sets since $x \mapsto t(x)$ will be constant on V level sets. Since V is strictly decreasing along trajectories of Ψ^t contained in U_1, it will follow that ρ takes disjoint level sets of V to disjoint subsets of $W^s(M)$, so that ρ will be injective. Therefore ρ is a homeomorphism since it is a continuous and closed bijection, so this will complete the proof if $k = 0$ — if $k \geq 1$, a computation in the proof of Theorem II.5 in §2.4.3 will show

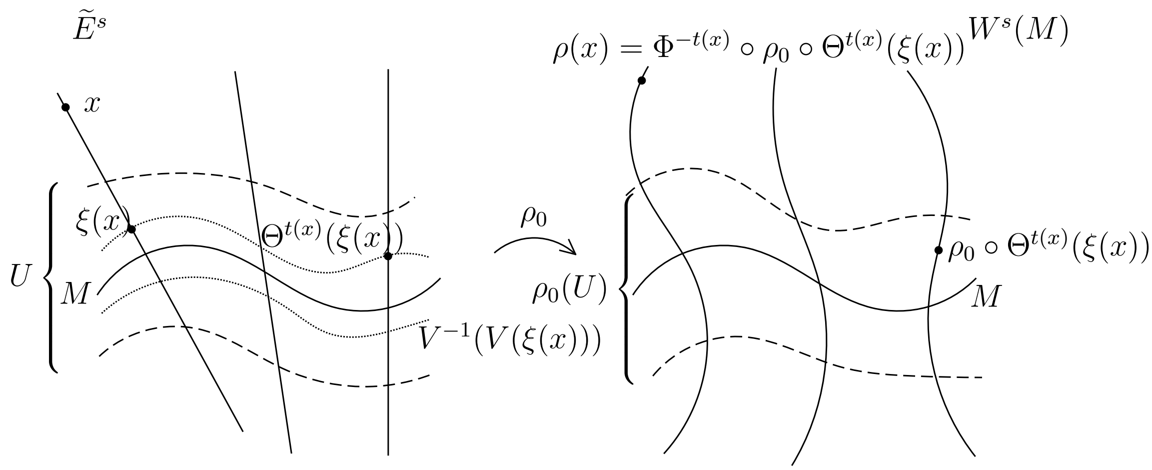


Figure 2.2: An illustration of the proof of Theorem II.5. The neighborhood $U \subset \widetilde{E}^s$ and its image $\rho_0(U)$ are bounded by dashed curves. The level set $V^{-1}(V(\xi(x))) \subset \widetilde{E}^s$ is depicted by the dotted curves.

that $\mathsf{D}\rho$ is an isomorphism everywhere, completing the proof of Theorem II.5.

The purpose of §2.4.2 is to construct the technical devices V, R_c, and Θ^t that will be used in the proof of Theorem II.5. The idea behind the proof of Theorem II.5 is simple, but our constructions are careful in order to avoid the loss of degrees of differentiability of ρ.

2.4.2 Preliminary Results

In order to carry out the proof of Theorem II.5, we need some tools. We will use the following result adapted from [PS70]; for the definition of a fiber metric, see Definition C.12 in Appendix C.

Proposition II.13. *Suppose that M is an inflowing invariant manifold for the C^r flow Φ^t on Q, and let $\pi\colon E \to M$ be a C^r subbundle of $TQ|_M$ equipped with any fiber metric. Let A^t be a continuous linear flow on E such that the time one map is a uniform contraction on fibers:*

$$\exists \alpha < 1\colon \forall m \in M\colon \|A^1|_{E_m}\| \leq \alpha.$$

Let Ψ^t be a $C^{k\geq 0}$ local flow with Ψ defined at least on a set of the form $[0,1]\times E(\epsilon)$ for some $\epsilon > 0$, where $E(\epsilon) \coloneqq \{y \in E : \|y\| \leq \epsilon\}$. Suppose that Ψ^t also covers Φ^t, leaves the zero section of E invariant, and is Lipschitz close to A^t for small t, by which we mean:

$$\forall 0 \leq t \leq 1, m \in M\colon \mathrm{Lip}((\Psi^t - A^t)|_{E_m}) \leq \mu < \min\left(\frac{1}{3}\kappa, 1-\alpha\right), \tag{2.13}$$

for $\kappa \coloneqq \inf\{\lfloor\!\lfloor A^t|_{E_m}\rfloor\!\rfloor : m \in M, 0 \leq t \leq 1\}$. Then there exists a continuous, nonnegative, and proper function $V\colon E \to \mathbb{R}$ such that $V^{-1}(0) = M$ and[10]:

1. *V is radially monotone on E. For any $c > 0$, $V^{-1}(c)$ intersects each radial ray in exactly one point $y \in E$. By a radial ray we mean any set of the form $\{\lambda x : \lambda > 0\}$, where $x \in E$ is nonzero.*

2. *For any c with $0 < c \leq 1$, the sublevel set $V^{-1}(-\infty, c]$ is contained in $E(\epsilon)$ and is positively invariant under Ψ^t.*

3. *V is radially bi-Lipschitz: there are constants $0 < b_1 < b_2 < 0$ such that for any $y \in E \setminus M$ and $\delta \neq 1$, we have the estimate*

$$0 < b_1 \leq \frac{|V(\delta y) - V(y)|}{\|\delta y - y\|} \leq b_2. \tag{2.14}$$

4. *If $\Psi \in C^{k\geq 1}$, then V is C^k on $E \setminus M$, and the derivative of V along any trajectory of Ψ starting in $V^{-1}(0,1)$ is strictly negative.*

Remark II.14. In the proof of Proposition II.13 below, we make use of Rademacher's theorem [Fed68, Thm 3.1.6]. This is to provide a unified proof for both the $C^{k\geq 1}$ and C^0 cases. In the $C^{k\geq 1}$ case we differentiate V in the radial direction in order to obtain the inequalities

[10] To limit excessive parentheses, here and henceforth we abuse notation by writing, e.g., $V^{-1}(a,b)$ instead of $V^{-1}((a,b))$, etc.

(2.14). This is not possible in the C^0 case, but condition (2.13) implies that the function V is locally Lipschitz, hence Rademacher's theorem implies that V is differentiable almost everywhere in the measure-theoretic sense. This is sufficient for our purposes.

Note that Ψ^t is actually radially differentiable in the context of Theorem II.5, even when $k = 0$. However, by using Rademacher's Theorem we simplify the statement of Proposition II.13, while weakening its hypotheses.

Proof. Define the function $g\colon E(\epsilon) \to \mathbb{R}$ by

$$g(y) := \int_0^1 \|\Psi^t(y)\| dt.$$

In the proof of [PS70, Thm 4.1] it is shown that g is continuous, radially monotone, and that for each $0 < \mu' \leq \mu$, $g^{-1}(\mu'\epsilon)$ intersects each radial ray in exactly one point $y \in E(\epsilon)$. It also follows from the proof that corresponding sublevel sets $g^{-1}(-\infty, \mu'\epsilon]$ are positively invariant. It follows from the last inequality in the proof of [PS70, Lem. 4.2] that for any $y \in E(\epsilon)$ and $\delta > 0$:

$$0 < (\kappa - 3\mu) \leq \frac{|g(\delta y) - g(y)|}{\|\delta y - y\|} \leq (\alpha + \mu),$$

where $\delta > 0$ is small enough that this expression is defined. Now let us assume that $k \geq 1$ and $\Psi \in C^k$ — it is clear that g is C^k on the complement of the zero section. We compute

$$\frac{\partial}{\partial t} g \circ \Psi^t(y) = \frac{\partial}{\partial t} \int_0^1 \|\Psi^{t+s}(y)\| ds = \frac{\partial}{\partial t} \int_t^{t+1} \|\Psi^s(y)\| ds = \|\Psi^1 \Psi^t(y)\| - \|\Psi^t(y)\| < 0,$$

with the last term being negative since A^1 is an α-contraction and Ψ^1 is μ-Lipschitz close to A^1, with $\mu + \alpha < 1$, hence Ψ^1 decreases the norm of $\Psi^t(y)$.

Replacing g by $g/(\mu\epsilon)$, we may assume that g satisfies 1, 2, and 4, and also

$$0 < b_1 \leq \frac{|g(\delta y) - g(y)|}{\|\delta y - y\|} \leq \beta \tag{2.15}$$

if we define $b_1 \coloneqq (\kappa - 3\mu)/(\mu\epsilon)$ and $\beta \coloneqq (\alpha + \mu)/(\mu\epsilon)$.

Now let $0 < \epsilon' < \epsilon$ be such that $g^{-1}(-\infty, 1] \subset E(\epsilon') \subset E(\epsilon)$. We are going to extend g to a C^k function $V : E \to \mathbb{R}$ such that $V|_{E(\epsilon')} = g|_{E(\epsilon')}$, with V satisfying 1, 2, 3, and 4. Let $\chi \colon [0, \infty) \to [0, \infty)$ be a C^∞ nonnegative, increasing function satisfying $\chi \equiv 0$ on $[0, \epsilon']$ and $\chi \equiv 1$ on $[\epsilon, \infty)$, and define $\psi \colon E \to \mathbb{R}$ via $\psi(y) \coloneqq \chi(\|y\|)$. We now define $V \colon E \to \mathbb{R}$ via

$$V \coloneqq (1 - \psi)g + \psi\beta\|\cdot\|, \tag{2.16}$$

with the understanding that $V(y) = \beta\|y\|$ for $\|y\| > \epsilon$. Clearly V is continuous. By the definition of ψ, we see that V is C^k on $E \setminus M$ and $V|_{E(\epsilon')} = g|_{E(\epsilon')}$, so that when we replace ϵ by ϵ' then 2 and 4 are automatically satisfied. Clearly 3 implies 1, so it suffices to show that V satisfies 3. To do this, fix any $y \in E \setminus M$. By (2.15), the function $\delta \mapsto g(\delta y)$ is locally Lipschitz, and the same is true of the other functions in Equation (2.16) defining V. Since V is a sum of products of such functions, V is also locally Lipschitz. Hence even if $k = 0$, by Rademacher's theorem $\delta \mapsto V(\delta y)$ and $\delta \mapsto g(\delta y)$ are differentiable except at a set of Lebesgue measure zero. The following statements must be interpreted to hold almost everywhere in the Lebesgue measure sense. We obtain

$$\frac{\partial}{\partial \delta} V(\delta y) = [(1 - \psi)\frac{\partial}{\partial \delta} g(\delta y) + \psi\beta\|y\|] + (\beta\|\delta y\| - g)\frac{\partial}{\partial \delta}\psi(\delta y),$$

where here and henceforth g and ψ are implicitly evaluated at δy. From this, we obtain the inequalities

$$[(1 - \psi)\frac{\partial}{\partial \delta} g(\delta y) + \psi\beta\|y\|] \leq \frac{\partial}{\partial \delta} V(\delta y) \leq [(1 - \psi)\frac{\partial}{\partial \delta} g(\delta y) + \psi\beta\|y\|] + \beta\|\delta y\|\frac{\partial}{\partial \delta}\psi(\delta y).$$

The leftmost inequality was obtained using $\frac{\partial}{\partial \delta}\psi(\delta y) \geq 0$ and the fact that Equation (2.15) implies that $\beta\|\delta y\| \geq g(\delta y)$, and the rightmost inequality was obtained since $g(\delta y), \frac{\partial}{\partial \delta}\psi(\delta y) \geq 0$.

Now Equation (2.15) implies that $b_1\|y\| \leq \frac{\partial}{\partial\delta}g(\delta y) \leq \beta\|y\|$ for $\delta y \in \operatorname{supp}(1-\psi)$, and $\beta \geq b_1$, so it follows that $b_1\|y\| \leq [(1-\psi)\frac{\partial}{\partial\delta}g(\delta y) + \psi\beta\|y\|] \leq \beta\|y\|$. Consequently, we have

$$b_1\|y\| \leq \frac{\partial}{\partial\delta}V(\delta y) \leq \beta\|y\| + \beta\|\delta y\|\frac{\partial}{\partial\delta}\psi(\delta y) = \beta[1 + \delta\frac{\partial}{\partial\delta}\psi(\delta y)]\|y\|. \tag{2.17}$$

The derivative term can be rewritten into a radial derivative

$$\delta\,\frac{\partial}{\partial\rho}\psi(\rho y)\bigg|_{\rho=\delta} = \frac{\partial}{\partial r}\psi(r\delta y)\bigg|_{r=1} =: \psi'(\delta y),$$

which is zero for $\delta y \notin E(\epsilon)$, and bounded inside the precompact set $E(\epsilon)$. Defining $b_2 :=$ $\beta[1 + \sup_{x\in E(\epsilon)}\psi'(x)] < \infty$, we see that the right hand side of (2.17) is bounded by $b_2\|y\|$.

The function $\delta \mapsto V(\delta y)$ is absolutely continuous since it is locally Lipschitz, so the fundamental theorem of Lebesgue integral calculus[11] implies that for any $\delta > 1$,

$$V(\delta y) - V(y) = \int_1^\delta \frac{\partial}{\partial s}V(sy)\,ds \geq b_1\|y\|(\delta - 1) = b_1\|\delta y - y\|,$$

and a similar argument shows that $V(\delta y) - V(y) \leq b_2\|\delta y - y\|$, with b_2 defined as before. This completes the proof. □

By Proposition II.13, for each $c > 0$ we may define a retraction $R_c\colon E \setminus M \to V^{-1}(c)$ by sliding along radial rays. We then define $R\colon (E \setminus M) \times (0,\infty) \to E$ by $R(\,\cdot\,, c) := R_c$.

Lemma II.15. *Let all notation be as in Proposition II.13 and let $R\colon (E\setminus M)\times(0,\infty) \to E$ be as defined above. Then $R \in C^k$.*

Proof. If $k \geq 1$, then $V \in C^{k\geq 1}$ and Equation (2.14) together with the mean value theorem imply that the derivative of V in the radial direction is nonzero. We may therefore apply the

[11]This technicality is needed only for the case that the differentiability degree $k = 0$. If $k \geq 1$, the mean value theorem or the elementary fundamental theorem of calculus will suffice.

implicit function theorem to the function $F(\delta, x, c) \coloneqq V(\delta x) - c$, defined on $(0, \infty) \times (E \setminus M) \times (0, \infty)$, to obtain a C^k $\mathbb{R}$-valued function $\delta(x, c)$ such that $V(\delta(x, c)x) = c$. It follows that $R(x, c) = \delta(x, c)x$, and therefore $R \in C^k$.

If $k = 0$, we will make use of a different argument which is effectively a "Lipschitz implicit function theorem". The argument is sketched as follows. We will define an auxiliary C^0 map $T \coloneqq (0, \infty) \times (E \setminus M) \times (0, \infty) \to \mathbb{R}$ such that $T_{x,c} \coloneqq T(\,\cdot\,, x, c)$ has a unique fixed point given by $\delta(x, c)$, and additionally such that $T_{x,c}$ is a contraction mapping. The domain of each $T_{x,c}$ is not a complete metric space, but the existence of the fixed point of each $T_{x,c}$ will follow from Proposition II.13 point 1, and these fixed points $R_c(x)$ are unique since $T_{x,c}$ is a contraction. The theorem then follows from the general fact that the fixed points $R_c(x)$ of a continuous family $T_{x,c}$ of contractions depends continuously on the parameters (x, c).

We now proceed with the proof. Define a continuous function T by

$$T_{x,c}(\delta) \equiv T(\delta, x, c) \coloneqq \delta - \frac{1}{b_2} \frac{V(\delta x) - c}{\|x\|}$$

on $(0, \infty) \times (E \setminus M) \times (0, \infty)$, where b_2 is as in Proposition II.13. We already know from Proposition II.13 that for each x and c, $T_{x,c}$ has a unique fixed point $\delta(x, c)$. $T_{x,c}$ is a contraction uniformly in x and c since

$$\begin{aligned} T_{x,c}(\delta_1) - T_{x,c}(\delta_2) &= \delta_1 - \delta_2 - \frac{1}{b_2} \frac{V(\delta_1 x) - V(\delta_2 x)}{\|x\|} \\ &= \left(1 - \frac{1}{b_2} \frac{V(\delta_1 x) - V(\delta_2 x)}{(\delta_1 - \delta_2)\|x\|}\right)(\delta_1 - \delta_2), \end{aligned}$$

so that by Equation (2.14) we have

$$|T_{x,c}(\delta_1) - T_{x,c}(\delta_2)| \leq k|\delta_1 - \delta_2|,$$

where $k := \left(1 - \frac{b_1}{b_2}\right) < 1$. It follows that the fixed point $\delta(x, c)$ depends continuously on (x, c), since it is a general fact that the fixed points of a (uniform) family of contractions $T_{x,c}$ depend continuously on the parameters (x, c). Since $V(\delta(x, c)x) = c$, it follows that $R(x, c) = \delta(x, c)x$, and therefore $R \in C^0$. This completes the proof. □

Lemma II.16 (Nonlinear parallel transport)**.** *Let all notation be as in Proposition II.13. Then there exists a C^k flow Θ^t on E such that Θ covers the base flow and preserves level sets of V:*

$$\forall t \colon V \circ \Theta^t = V.$$

Proof. Let Π^t be any C^r linear parallel transport covering Φ^t as in Lemma A.1 (see Appendix A). We define Θ^t for $t > 0$ by flowing x forward via the linear flow Π^t and then projecting onto the $V(x)$ level set of V:

$$\Theta^t(x) := R_{V(x)} \circ \Pi^t(x).$$

It follows from Lemma II.15 that $\Theta \in C^k$. Since for each t the linear flow Π^t maps radial rays into radial rays, it follows that Θ^t is injective for fixed $t \geq 0$ and also that Θ indeed satisfies the group property.

By Lemma A.1 in Appendix A, Π^t is a C^r linear isomorphism for each $t > 0$, and R is C^k by Lemma II.15. Using the fact that Π^t preserves radial rays, it follows that the map $x \mapsto R_{V(x)} \circ \Pi^{-t}(x)$ defined on $\Pi^t(E)$ is a C^k inverse for Θ^t, so Θ^t is a C^k isomorphism onto its image.

□

2.4.3 The proof of Theorem II.5

Now we start the proof that $W^s(M)$ is a fiber bundle isomorphic to $\widetilde{E}^s$ over the inflowing NAIM M. (For the reader new to fiber bundles, see Appendix C and in particular Example C.7).

Proof of Theorem II.5. Let $\rho_0 \colon U \subset \widetilde{E}^s \to W^s_{\text{loc}}(M)$ be the C^k fiber-preserving isomorphism constructed using Lemma II.2 and Remark II.4. With Proposition II.13 in mind, we define a C^k local flow Ψ^t on U and a global $C^{\max\{k-1,0\}}$ linear flow A^t on $\widetilde{E}^s$ as follows:

$$\Psi^t := \rho_0^*(\Phi^t) = \rho_0^{-1} \circ \Phi^t \circ \rho_0, \qquad \forall m \in M \colon A^t|_{\widetilde{E}^s_m} := [\mathsf{D}(\rho_0|_{\widetilde{E}^s_{\Phi^t(m)}})]^{-1} \circ \mathsf{D}\Phi^t \circ \mathsf{D}(\rho_0|_{\widetilde{E}^s_m}), \tag{2.18}$$

for all $t > 0$. Here we are viewing $\mathsf{D}(\rho_0|_{\widetilde{E}^s_m})$ as a map $\widetilde{E}^s_m \to E^s_m$ via the canonical linear identification $\mathsf{T}_0 \widetilde{E}^s_m \cong \widetilde{E}^s_m$. Note that by compactness of M, the linear flow A^t is eventually uniformly contracting relative to the fiber metric (see Def. C.12 in Appendix C) on $\widetilde{E}^s$ induced by the Riemannian metric on TQ: i.e., there exists $t_0 > 0$ and $0 \le \alpha < 1$ such that

$$\forall m \in M \colon \|A^{t_0}|_{E_m}\| \le \alpha.$$

Furthermore, even if $k = 0$, the restrictions $\rho_0|_{\widetilde{E}^s_m}$ of ρ_0 to individual linear fibers of $\widetilde{E}^s$ are smooth (see, e.g., [Fen74, HPS77, Thm 1, Thm 4.1]). It follows that $\Psi^t|_{\widetilde{E}^s_m}$ is smooth, and

$$\forall m \in M \colon \mathsf{D}(\Psi^t|_{\widetilde{E}^s_m})_0 = A^t|_{\widetilde{E}^s_m}. \tag{2.19}$$

This is because the restrictions $\rho_0|_{\widetilde{E}^s_m}$ of ρ_0 to individual linear fibers of $\widetilde{E}^s$ are smooth. By NHIM theory (see [PS70, p. 191, (2.4)]), we also have that the map $(m, y) \mapsto \mathsf{D}(\rho_0|_{\widetilde{E}^s_m})_y$ is uniformly continuous at the zero section in the sense that $\mathsf{D}(\rho_0|_{\widetilde{E}^s_{m'}})_y$ tends uniformly to $\mathsf{D}(\rho_0|_{\widetilde{E}^s_m})_0$ as $m' \to m$ and $\|y\| \to 0$.

It therefore follows in either case ($k > 0$ or $k = 0$), possibly after a rescaling of time, that Ψ^t and A^t satisfy the hypotheses of Proposition II.13 on some uniform neighborhood of the zero section — the Lipschitz condition hypothesis in Proposition II.13 follows from the preceding sentence, see also [PS70, p. 191, (2.4)(b')]. Hence we obtain a radially monotone function $V\colon \widetilde{E}^s \to \mathbb{R}$ as in Proposition II.13, and the corresponding C^k family of radial retractions $R_c\colon \widetilde{E}^s \setminus M \to V^{-1}(c)$ with $0 < c < \infty$ as in Lemma II.15. As in Lemma II.16, we also obtain a C^k flow Θ^t defined on $E \setminus M$, covering Φ^t, and preserving level sets of V. For the sake of notation, for any $0 < c < \infty$ we henceforth let U_c denote the sublevel set $V^{-1}(-\infty, c)$.

We next define the following smooth functions. Let $\chi \in C^\infty\big([0,\infty);[0,1)\big)$ be a global diffeomorphism such that $\chi(\delta) = \delta$ for $\delta \le \frac{1}{2}$ and $\chi'(\delta) \in (0,1)$ for $\delta > \frac{1}{2}$. Secondly, define $\tau(\delta) := \delta - \chi(\delta)$. Hence we have $\tau \in C^\infty\big([0,\infty);[0,\infty)\big)$ with $\tau(\delta) = 0$ for $\delta \le \frac{1}{2}$ and $\tau'(\delta) > 0$ for $\delta > \frac{1}{2}$. Thus τ restricted to $(\frac{1}{2}, \infty)$ is a diffeomorphism onto $(0,\infty)$.

Finally, we construct the global fiber bundle isomorphism $\rho\colon \widetilde{E}^s \to W^s(M)$ as follows. Let $x \in \widetilde{E}^s_m$ at the base point $m \in M$. Define the rescaled $\xi(x) := R_{\chi(V(x))}(x)$ if $\|x\| \neq 0$ and $\xi(x) = x$ otherwise. Note that $\xi(x) \in U_1 \subset \widetilde{E}^s$ is C^k dependent on x by Lemma II.15, and ξ is a C^k isomorphism since its C^k inverse is given by $y \mapsto R_{\chi^{-1}(V(y))}(y)$. Secondly, define $t(x) := \tau(V(x))$ if $\|x\| \neq 0$ and $t(x) = 0$ otherwise, and note that t is C^k dependent on x since $\tau(\delta) = 0$ for $\delta \le \frac{1}{2}$. Now define

$$\rho\colon \widetilde{E}^s \to W^s(M), \qquad \rho(x) = \Phi^{-t(x)} \circ \rho_0 \circ \Theta^{t(x)} \circ \xi(x). \tag{2.20}$$

By construction it is clear that ρ is a C^k fiber-preserving map covering the identity on M.

We now show that ρ is injective. First, note that ρ restricted to any level set of V is injective since the function $t \mapsto t(x)$ is constant on such level sets by construction, and for any fixed $t_0 \ge 0$ the map $x \mapsto \Phi^{-t_0} \circ \rho_0 \circ \Theta^{t_0} \circ \xi(x)$ is a C^k isomorphism. Hence it suffices to show that $\rho(V^{-1}(a)) \cap \rho(V^{-1}(b)) = \varnothing$ for any $a \neq b$, $a, b > 0$. Let $t_1 = t(V^{-1}(a))$, $t_2 = t(V^{-1}(b))$, and assume without loss of generality that $b > a$ and hence $t_2 > t_1$. The following are equivalent statements:

$$
\begin{aligned}
& \rho(V^{-1}(a)) \cap \rho(V^{-1}(b)) = \varnothing \\
\iff \quad & \Phi^{-t_1} \circ \rho_0 \circ \Theta^{t_1} \circ \xi\Big(V^{-1}(a)\Big) \cap \Phi^{-t_2} \circ \rho_0 \circ \Theta^{t_2} \circ \xi\Big(V^{-1}(b)\Big) = \varnothing \\
\iff \quad & \Phi^{t_2 - t_1} \circ \rho_0 \circ \Theta^{t_1} \circ \xi\Big(V^{-1}(a)\Big) \cap \rho_0 \circ \Theta^{t_2} \circ \xi\Big(V^{-1}(b)\Big) = \varnothing \\
\iff \quad & \Psi^{t_2 - t_1} \circ \Theta^{t_1} \circ \xi\Big(V^{-1}(a)\Big) \cap \Theta^{t_2} \circ \xi\Big(V^{-1}(b)\Big) = \varnothing \\
\iff \quad & \Psi^{t_2 - t_1} \circ \xi\Big(V^{-1}(a)\Big) \cap \xi(V^{-1}(b)) = \varnothing \\
\iff \quad & \Psi^{t_2 - t_1}\Big(V^{-1}(\chi(a))\Big) \cap V^{-1}(\chi(b)) = \varnothing,
\end{aligned}
$$

where we used in the last line that by construction of ξ, $\xi(V^{-1}(a)) = V^{-1}(\chi(a))$ and $\xi(V^{-1}(b)) = V^{-1}(\chi(b))$. Since $a < b$ we have $0 < \chi(a) < \chi(b) < 1$, and since for any $t \geq 0$ we have that $V^{-1}(-\infty, a]$ is Ψ^t-invariant by Proposition II.13, it follows that indeed $\Psi^{t_2 - t_1}\Big(V^{-1}(\chi(a))\Big) \cap V^{-1}(\chi(b)) = \varnothing$. Hence ρ is injective.

We continue with surjectivity of ρ. Letting $(y_n)_{n \in \mathbb{N}}$ be any sequence in $\widetilde{E}^s$ with $\|y_n\| \to \infty$, it follows that $t(y_n) \to \infty$ and $\chi(y_n) \to 1$. For any $\delta > 0$, let U_δ denote the δ-sublevel set of V, consistent with our notation U_1. Let $K \subset W^s(M)$ be any compact set. By compactness and asymptotic stability of M, there exists $t_0 > 0$ such that $\forall t \geq t_0 \colon \Phi^t(K) \subset U_{1/2}$. It follows that for all sufficiently large $n \in \mathbb{N}$:

$$\rho(y_n) \in W^s(M) \cap \bigcup_{t \geq t_0} \Phi^{-t}(\rho_0(U_1 \setminus U_{1/2})) \subset W^s(M) \setminus K.$$

Hence ρ takes diverging sequences to diverging sequences and is therefore a proper map, so ρ is also a closed map. We have already shown that the continuous map ρ is injective. Using these facts, we establish surjectivity of ρ as follows. Since ρ maps the manifold interior $\operatorname{int} \widetilde{E}^s$ of $\widetilde{E}^s$ into the manifold interior $\operatorname{int} W^s(M)$ of $W^s(M)$, it follows by invariance of domain that $\rho|_{\operatorname{int} \widetilde{E}^s} \colon \operatorname{int} \widetilde{E}^s \to \operatorname{int} W^s(M)$ is an open map, and since we also know that ρ is a closed map, it follows by connectivity that $\operatorname{int} W^s(M) = \rho(\operatorname{int} \widetilde{E}^s)$. Next, since $\rho(\partial \widetilde{E}^s) \subset \partial W^s(M)$ and since $\partial \widetilde{E}^s$ and $\partial W^s(M)$ are topological manifolds, we may invoke invariance of domain again and similarly conclude that $\rho(\partial \widetilde{E}^s) = \partial W^s(M)$. This completes the proof of surjectivity of ρ.

To summarize, we have shown that ρ is a bijective, continuous, and closed map. Therefore, ρ is a homeomorphism. This completes the proof if $k = 0$.

Assuming now that $k \geq 1$, it suffices to show that ρ is a local diffeomorphism. Since ρ agrees with the diffeomorphism ρ_0 on $U_{1/2}$, it suffices to consider $x \in \widetilde{E}^s \setminus M$. Let $y :=$ $\rho_0 \circ \Theta^{t(x)} \circ \xi(x)$, $\xi' := \frac{\partial}{\partial \delta}\xi(\delta x)|_{\delta=1}$, and $\kappa := \frac{\partial}{\partial \delta} t(\delta x)|_{\delta=1}$. A computation using $\frac{\partial}{\partial t}\Phi^t = \mathsf{D}\Phi^t \circ f$ shows that

$$\frac{\partial}{\partial \delta}\rho(\delta x)|_{\delta=1} = \mathsf{D}\Phi^{-t(x)}\left[-\kappa f(y) + \mathsf{D}\rho_0\left(\kappa g(\xi(x)) + \mathsf{D}\Theta^{t(x)}\xi'\right)\right], \tag{2.21}$$

where $g\colon \widetilde{E}^s \to \mathsf{T}\widetilde{E}^s$ is the vector field generating Θ — note that g is tangent to V level sets. The vector in brackets points outward to $\Phi^{t(x)} \circ \rho(V^{-1}(a))$, where $a = V(x)$. To see this, first note that Proposition II.13 and the inequality $\tau' > 0$ imply that $\kappa > 0$, and therefore $-\kappa f(y)$ points outward to $\Phi^{t(x)} \circ \rho(V^{-1}(a))$. Similar reasoning also shows that ξ' is outward pointing at V level sets. Since Θ^t is a flow, it follows that $\mathsf{D}\Theta^{t(x)}$ can be smoothly deformed to the identity through isomorphisms (in other words, an "isotopy"), which implies that $\mathsf{D}\Theta^{t(x)}\xi'$ is also outward pointing at V level sets. Since ρ_0 is a diffeomorphism which maps the zero section of $\tilde{E}^s$ to M, $\mathsf{D}\rho_0$ maps outward pointing vectors at V level sets to outward pointing vectors at $V \circ \rho_0^{-1}$ level sets. Taken together, these facts show that the quantity in brackets indeed points outward to $\Phi^{t(x)} \circ \rho(V^{-1}(a))$. Now since Φ^t is a flow, $\mathsf{D}\Phi^{-t(x)}$ can also be smoothly deformed to the identity through isomorphisms, and therefore the same reasoning above in the case of $\mathsf{D}\Theta^{t(x)}$ establishes that $\frac{\partial}{\partial \delta}\rho(\delta x)|_{\delta=1}$ is outward pointing to $\rho(V^{-1}(a))$ at $\rho(x)$. On the other hand, $\mathsf{D}\rho$ takes a basis for $\mathsf{T}_x V^{-1}(a)$ to a basis for $\mathsf{T}_{\rho(x)}\rho(V^{-1}(a))$, so $\mathsf{D}\rho$

is an isomorphism. This completes the proof. □

2.5 Global linearization

A classic result in the theory of normally hyperbolic invariant manifolds is that the dynamics are "linearizable" on some (a priori small) neighborhood of the NHIM [PS70, HPS77, PT77], which is to say that there is some neighborhood U of $M \subset Q$ and a fiber-preserving homeomorphism $\varphi\colon U \to \varphi(U) \subset E^s \oplus E^u$ onto a neighborhood of the zero section such that

$$\varphi \circ \Phi^t|_U = \mathsf{D}\Phi^t|_{E^s \oplus E^u} \circ \varphi, \tag{2.22}$$

for all $t \in \mathbb{R}$ such that both sides of the expression are defined. This is a vast generalization of the Hartman–Grobman Theorem.

In [LM13], this local result is extended to a global result for the special cases of exponentially stable equilibria and periodic orbits. More precisely, it is shown that the domain of the linearization can actually be taken to be the entire basin of attraction for these attractors. As conjectured in the conclusion of [LM13], this globalization result should generalize to hold for arbitrary (boundaryless) NAIMs. In this section, we establish this generalization; our methods are similar to theirs.

We would like to apply this linearization result in the context of slow-fast systems to derive linear normal forms on a neighborhood of a slow manifold, improving upon the Fenichel Normal Form to be discussed in §2.6. However, the relevant slow manifolds are often compact manifolds with boundary. To the best of our knowledge, neither the results mentioned above nor the existing (local) linearization results in the literature directly apply in this case [PS70, Rob71, HPS77, PT77, Sel84, Sel83, Sak94, BK94]. Thus, the content of this section can be divided as follows.

1. In §2.5.1, we prove that the dynamics restricted to the basin of attraction of a compact *boundaryless* NAIM are globally linearizable (and smoothly linearizable, assuming some additional hypotheses). This is the content of Theorem II.18 and its corollaries.

2. In §2.5.2, we turn to the main goal of §2.5, which is to prove that the dynamics restricted to the global stable manifold of a compact *inflowing* NAIM are globally linearizable (and smoothly linearizable, assuming some additional hypotheses). This is the content of Theorem II.24 and its corollaries. We prove this local result in the course of proving the stronger global result. To achieve this, we use a topological construction developed in Appendix B, which might be of independent interest.

Remark II.17. After we had proved Theorem II.18 (global linearization for the boundaryless case), we learned that Igor Mezić independently obtained this theorem before us. A very readable proof appears in his soon-to-be published textbook on Koopman operator theory [Mez17]. His proof technique is the same as ours. However, his result applies only to boundaryless NAIMs, and therefore we need our Theorem II.24 (global linearization for inflowing NAIMs) for our goal of deriving a linear normal form for a class of slow-fast systems, which we do in §2.6.

2.5.1 Global linearization for boundaryless NAIMs

In the following results, recall that by a C^k isomorphism, we mean a homeomorphism if $k = 0$ and a C^k diffeomorphism if $k \geq 1$. Theorem II.18 will be used as a stepping stone to prove a global linearization result for inflowing NAIMs in §2.5.2 below, which we apply to slow-fast systems in §2.6.

Theorem II.18. *Let $M \subset Q$ be a compact (boundaryless) 1-NAIM for the C^r flow Φ^t on Q. Assume that $E^s \in C^k$, with $0 \leq k \leq r-1$, and that Φ^t is locally C^k conjugate to the linear* *flow* $\mathsf{D}\Phi^t|_{E^s}$ *on some neighborhood of* $M \subset W^s(M)$. *Then* Φ^t *is globally* C^k *conjugate to* $\mathsf{D}\Phi^t|_{E^s}$*, which is to say that there exists a* C^k *fiber-preserving isomorphism* $\varphi\colon W^s(M) \to E^s$ *such that*

$$\forall t \in \mathbb{R}\colon \varphi \circ \Phi^t = \mathsf{D}\Phi^t|_{E^s} \circ \varphi. \tag{2.23}$$

Additionally, φ *agrees with the local conjugacy on its domain.*

Remark II.19. The hypotheses of Theorem II.18 *assume* the existence of a local linearizing C^k conjugacy. Theorem II.18 shows that any local linearizing conjugacy may be extended to a global linearizing conjugacy having the same regularity.

Remark II.20. The relationship between Theorems II.5, II.18, and II.24 (see §2.5.2 below) are as follows. The hypotheses of Theorem II.18 and II.24 are much stronger than those required for Theorem II.5, and in particular the hypotheses of Theorem II.5 are not sufficient to prove the conclusion of Theorems II.18 and II.24. However, the hypotheses of Theorem II.18 and II.24 suffice to prove the conclusion of Theorem II.5 in the cases that $\partial M = \varnothing$ and M is inflowing with $\partial M \neq \varnothing$, respectively, since the conjugacy $\varphi\colon W^s(M) \to E^s$ is in particular a C^k fiber-preserving isomorphism.

Remark II.21. As pointed out in [LM13], the flow is automatically locally C^1 linearizable near an exponentially stable equilibrium or periodic orbit. See the references therein. It is shown in [PS70, HPS77, PT77] that the flow is always locally C^0 linearizable near a NHIM. There are also various results in the literature giving conditions ensuring that Φ^t is locally C^k linearizable near a general invariant manifold, such as [Sak94, Tak71, Rob71, Sel84, Sel83]. See also [BK94, Chap. VI] for similar results, as well as historical remarks. In particular, we obtain the following easy corollary.

Corollary II.22. *Let* M*,* Q*,* P^s*, and* Φ^t *be as in Theorem II.18. Assume that* Φ^t *is a* C^1 *flow and that* M *is a* 1*-NAIM. Then* $\Phi^t|_{W^s(M)}$ *is globally topologically conjugate to* $\mathsf{D}\Phi^t|_{E^s}$*.*

Proof. As remarked already, it is shown in [PS70, HPS77, PT77] that Φ^t is locally topologically conjugate to $\mathsf{D}\Phi^t|_{E^s}$ near M. Thus Theorem II.18 yields the existence of a global topological conjugacy between Φ^t and $\mathsf{D}\Phi^t|_{E^s}$. □

Remark II.23. This furnishes a proof alternative to the one given in §2.4.3 that for a 1-NAIM M, $P^s \colon W^s(M) \to M$ is always a topological fiber bundle isomorphic to E^s over M.

Proof of Theorem II.18. By assumption, there exists a neighborhood U of M and a C^k fiber-preserving isomorphism $\varphi_{\text{loc}} \colon U \to \varphi_{\text{loc}}(U) \subset E^s$ such that for all $t > 0$:

$$\varphi_{\text{loc}} \circ \Phi^t|_U = \mathsf{D}\Phi^t|_{E^s} \circ \varphi_{\text{loc}}. \tag{2.24}$$

We now extend this local conjugacy to a global one.

Let V be a strict C^∞ Lyapunov function for the flow Φ^t [Wil67, Wil69], and let f be the vector field generating that flow. V is nonnegative, $V^{-1}(0) = M$, V is proper, and the Lie derivative $L_f V$ of V along trajectories not contained in M is strictly negative. Since V is proper, there is $c > 0$ such that $V^{-1}(c) \subset U$, for example, take $c < \inf_{x \in W^s(M) \setminus U} V(x)$.

Since $L_f V < 0$ on $V^{-1}(c)$, it follows that the vector field f intersects $V^{-1}(c)$ transversally. The properties of V imply that for all $x \in W^s(M) \setminus M$ there exists a unique "impact time" $\tau(x) \in \mathbb{R}$ such that $\Phi^{\tau(x)}(x) \in V^{-1}(c)$. Using transversality of f to $V^{-1}(c)$ and the implicit function theorem applied to $(x, \tau) \mapsto V(\Phi^\tau(x))$, we see that $\tau \colon W^s(M) \setminus M \to \mathbb{R}$ is C^r.

Now define a map $\varphi \colon W^s(M) \to E^s$ by

$$\varphi(x) \coloneqq \begin{cases} \mathsf{D}\Phi^{-\tau(x)} \circ \varphi_{\text{loc}} \circ \Phi^{\tau(x)}(x), & x \in W^s(M) \setminus M \\ \varphi_{\text{loc}}, & x \in U. \end{cases} \tag{2.25}$$

See Figure 2.3. Note that φ is well-defined because Equation (2.24) implies that the two functions in (2.25) agree on $U \setminus M$, and hence $\varphi \in C^k$ since clearly both maps in (2.25) are.

Note also that φ maps fibers $W^s(m)$ into fibers E^s_m by invariance of the stable foliation and stable vector bundle under the nonlinear and linear flows, respectively. It is easy to check directly that φ conjugates the flows as in Equation (2.23) — we now show that φ is a C^k isomorphism.

We first show that $\varphi\colon W^s(M) \to E^s$ is injective. Define the C^k function $V'\colon \varphi_{\text{loc}}(U) \to \mathbb{R}$ by $V' \coloneqq V \circ (\varphi_{\text{loc}})^{-1}$. We have that $\forall v = \varphi_{\text{loc}}(x) \in \varphi_{\text{loc}}(U)$ and all $t > 0$:

$$V' \circ \mathsf{D}\Phi^t(v) = V' \circ \mathsf{D}\Phi^t \circ \varphi_{\text{loc}}(x) = V' \circ \varphi_{\text{loc}} \circ \Phi^t(x) = V \circ \Phi^t(x). \tag{2.26}$$

It follows that V' is strictly decreasing along trajectory segments of $\mathsf{D}\Phi^t|_{E^s}$ contained in $\varphi_{\text{loc}}(U)$, so that any trajectory of $\mathsf{D}\Phi^t|_{E^s}$ starting in $\varphi_{\text{loc}}(U \setminus M)$ intersects the c level set $\Sigma \coloneqq (V')^{-1}(c)$ of V' in precisely one point. Now suppose that $\varphi(x) = \varphi(y)$ with $x \neq y$. Then we have $\mathsf{D}\Phi^{-\tau(x)}(v) = \mathsf{D}\Phi^{-\tau(y)}(w)$, where $v \coloneqq \varphi_{\text{loc}} \circ \Phi^{\tau(x)}(x) \in \Sigma$ and $w \coloneqq \varphi_{\text{loc}} \circ \Phi^{\tau(y)}(y) \in \Sigma$. It follows that $v = \mathsf{D}\Phi^{\tau(x)-\tau(y)}(w)$ which, by the previous comments, implies that $\tau(x) = \tau(y)$ and that $v = w$. By injectivity of φ_{loc} we therefore have $\Phi^{\tau(x)}(x) = \Phi^{\tau(x)}(y)$. Since $\Phi^{\tau(x)}$ is injective, $x = y$. Hence φ is injective.

We next show that $\varphi\colon W^s(M) \to E^s$ is surjective. Note that φ_{loc} maps M one-to-one and onto the zero section of E^s (this must be the case since homeomorphisms preserve ω-limit sets). Now consider any $v \in E^s \setminus M$, identifying M with the zero section as usual. Since M is a NAIM, the zero section of E^s is asymptotically stable for the linear flow $\mathsf{D}\Phi^t|_{E^s}$. This fact and continuity imply that there is $t_0 > 0$ such that $\mathsf{D}\Phi^{t_0}(v) \in \Sigma$. Let $x' \in U$ be the unique point with $\varphi_{\text{loc}}(x') = \mathsf{D}\Phi^{t_0}(v)$. Setting $x = \Phi^{-t_0}(x')$, we see that $\tau(x) = t_0$ and that

$$\varphi(x) = \mathsf{D}\Phi^{-\tau(x)} \circ \varphi_{\text{loc}} \circ \Phi^{\tau(x)}(x) = \mathsf{D}\Phi^{-t_0} \circ \varphi_{\text{loc}} \circ \Phi^{t_0}(x) = \mathsf{D}\Phi^{-t_0} \circ \varphi_{\text{loc}}(x') = \mathsf{D}\Phi^{-t_0} \circ \mathsf{D}\Phi^{t_0}(v) = v.$$

To complete the proof, it suffices to prove that $\varphi^{-1} \in C^k$. Since $\varphi^{-1}|_{\varphi(U)} = (\varphi_{\text{loc}})^{-1}$, φ^{-1}

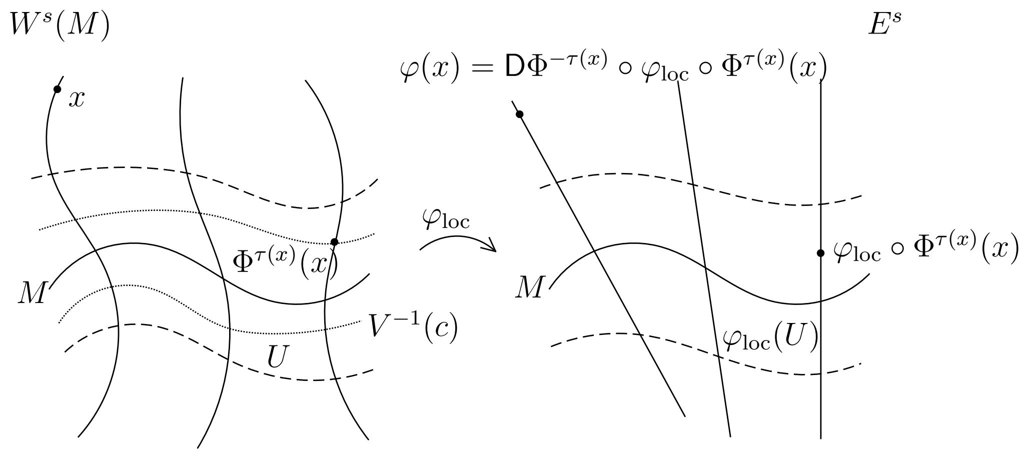


Figure 2.3: An illustration of the proof of Theorem II.18. The neighborhood $U \subset W^s(M)$ and its image $\varphi_{\text{loc}}(U)$ are bounded by dashed curves. The level set $V^{-1}(c) \subset W^s(M)$ is depicted by the dotted curves.

is C^k on U. Now let $v \in E^s \setminus \varphi(U)$. By asymptotic stability of the zero section for $\mathsf{D}\Phi^t|_{E^s}$, there exists $t_0 > 0$ such that $\mathsf{D}\Phi^{t_0}(v) \in U$. Since $\varphi^{-1} = \Phi^{-t_0} \circ \varphi^{-1} \circ \mathsf{D}\Phi^{t_0}|_{E^s}$ by (2.23), it follows that

$$\varphi^{-1}|_{\mathsf{D}\Phi^{-t_0}(\varphi(U))} = \Phi^{-t_0} \circ (\varphi_{\text{loc}})^{-1} \circ \mathsf{D}\Phi^{t_0}|_{\mathsf{D}\Phi^{-t_0}(\varphi(U))}$$

is a composition of C^k maps, so that φ^{-1} is C^k on a neighborhood of v. This completes the proof. □

2.5.2 Global linearization for inflowing NAIMs

We next proceed to our main goal for §2.5, which is to prove a global linearization theorem for inflowing NAIMs. The key tool we use is Proposition B.1 in Appendix B, which shows that many results about boundaryless NAIMs can be transferred to inflowing NAIMs. We reiterate that this result is necessary for our derivation of a linear normal form (see Theorem II.28) for slow-fast systems, since the slow manifolds for these systems are typically manifolds with boundary.

Theorem II.24. *Let $M \subset Q$ be a compact inflowing r-NAIM for the flow Φ^t generated by the C^r vector field f on Q, where $r \geq 3$. Assume further that there exist constants $0 < \delta < -\alpha < -\beta$ and $K \geq 1$ such that $-\alpha > r\delta$, $-\beta < -2\alpha - (r-1)\delta$, and such that for all $t \geq 0$*

$$\begin{aligned} K^{-1}e^{-\delta t} \leq \lfloor\!\lfloor \mathsf{D}\Phi^t|_{TM} \rfloor\!\rfloor \leq \|\mathsf{D}\Phi^t|_{TM}\| \leq Ke^{\delta t}, \\ K^{-1}e^{-\delta t} \leq \lfloor\!\lfloor (\mathsf{D}\Phi^t|_{TM})^{-1} \rfloor\!\rfloor \leq \|(\mathsf{D}\Phi^t|_{TM})^{-1}\| \leq Ke^{\delta t}, \\ K^{-1}e^{\beta t} \leq \lfloor\!\lfloor \mathsf{D}\Phi^t|_{E^s} \rfloor\!\rfloor \leq \|\mathsf{D}\Phi^t|_{E^s}\| \leq Ke^{\alpha t} \end{aligned} \tag{2.27}$$

hold uniformly on TM and E^s. Then $E^s \in C^{r-1}$ and $\Phi^t|_{W^s(M)}$ is globally C^{r-1} conjugate to $\mathsf{D}\Phi^t|_{E^s}$.

Proof. By Proposition B.1 in Appendix B, there exists a C^∞ manifold $\widehat{Q}$, an open neighborhood $U \supset W^s(M)$, a C^∞ embedding $\iota\colon U \to \widehat{Q}$, a C^r vector field $\hat{f}$ on $\widehat{Q}$ generating a C^r flow $\widehat{\Phi}^t$, and a C^r compact and boundaryless r-NAIM $\widehat{N} \subset \widehat{Q}$ for $\widehat{\Phi}^t$ with the following properties.

1. $\iota_*(f|_{W^s(M)}) = \hat{f}|_{\iota(W^s(M))}$.

2. $\forall m \in M : \iota(W^s(m)) = \widehat{W}^s(\iota(m))$, where $W^s(M)$ and $\widehat{W}^s(\widehat{N})$ are the global stable foliations of M for f and $\widehat{N}$ for $\hat{f}$, respectively.

3. There exist constants δ', α', β' arbitrarily close to δ, α, β such that (2.27) holds uniformly on $T\widehat{N}$ and $\widehat{E}^s$, after replacing δ, α, β, M, E^s, and Φ^t by δ', α', β', $\widehat{N}$, $\widehat{E}^s$, and $\widehat{\Phi}^t$, respectively. Here, $\widehat{E}^s$ is the stable vector bundle of $\widehat{N}$ for $\widehat{\Phi}^t$.

In [Sak94, p. 335, Thm B] it is shown[12] that item 3 implies that $\widehat{W}^s(\widehat{N}), \widehat{E}^s \in C^{r-1}$ and that $\widehat{\Phi}^t$ is locally C^{r-1} conjugate to $\mathsf{D}\widehat{\Phi}^t|_{\widehat{E}^s}$ near $\widehat{N}$. By Theorem II.18, it follows

[12]Comparing (2.27) to [Sak94, Eq. (2.7)], note that we express the conditions on $\mathsf{D}\Phi^t|_{TM}$ only for $t \geq 0$, but also on the inverse to prevent issues with $\Phi^t(m)$ leaving M. We do not have an equivalent for Sakamoto's estimate for Z since we have no unstable bundle. Furthermore, the lack of minimum norms for the lower bounds seems to be a minor oversight in Sakamoto's conditions. Next, Sakamoto's result is actually only stated for a NAIM with trivial normal bundle, but this is easily extended to the general case by locally writing the dynamics on the normal bundle and then extending them to the total space of the direct sum

that $\widehat{\Phi}^t|_{\widehat{W}^s(\widehat{N})}$ is globally C^{r-1} conjugate to $\mathsf{D}\widehat{\Phi}^t|_{\widehat{E}^s}$. Let $\widehat{\varphi}\colon \widehat{W}^s(\widehat{N}) \to \widehat{E}^s$ be such a conjugacy. Items 1 and 2 imply that restriction of $\widehat{\varphi}$ yields a well-defined global C^{r-1} conjugacy $\widehat{\varphi}|_{\widehat{W}^s(\iota(M))}\colon \widehat{W}^s(\iota(M)) \to \widehat{E}^s|_{\iota(M)}$ between $\widehat{\Phi}^t|_{\widehat{W}^s(\iota(M))}$ and $\mathsf{D}\widehat{\Phi}^t|_{\widehat{E}^s|_{\iota(M)}}$. Hence $\varphi := \iota|^{-1}_{\widehat{W}^s(\iota(M))} \circ \widehat{\varphi} \circ \iota$ is a global C^{r-1} conjugacy from $\Phi^t|_{W^s(M)}$ to $\mathsf{D}\Phi^t|_{E^s}$. □

Corollary II.25. *Let $M \subset Q$ be a compact inflowing 1-NAIM for the C^r flow Φ^t generated by the C^r vector field f on Q. Then $\Phi^t|_{W^s(M)}$ is globally topologically conjugate to $\mathsf{D}\Phi^t|_{E^s}$.*

Proof. The proof is identical to that of Theorem II.24, but with [PS70, Thm 2] used instead of [Sak94, p. 335, Thm B] to provide a local linearizing C^0 conjugacy. □

2.6 Applications to Geometric Singular Perturbation Theory

We give two applications of Theorem II.5 and Theorem II.24 to slow-fast systems in the context of geometric singular perturbation theory (GSP). Our applications assume the special case in which the slow manifold is attracting. Both applications are improvements of the so-called Fenichel Normal Form, discussed below, and are contained in Theorems II.26 and II.28 below.

The Fenichel Normal Form [JK94, Jon95, Kap99, JT09] is the form that the equations of motion take near the slow manifold of a slow-fast system, when written in local coordinates which are adapted to the slow manifold and its stable and unstable foliations. One application of this normal form was to derive the estimates used to prove the so-called Exchange Lemma and its extensions, which are useful tools for establishing the existence of heteroclinic and homoclinic orbits in slow-fast systems; see, e.g., [JK94, Jon95, JKK96, Bru96, KJ01,

with an *inverse bundle*; this useful trick is briefly mentioned in a different context on [Sak94, pp. 333-334], but see also [PSW97, Sec. 3] for more details. Finally, it is actually claimed in [Sak94, p. 333] that $\widehat{E}^s \in C^r$, but to the best of our knowledge this seems to be a minor oversight — the references in [Sak94, p. 333] provided to support this statement either claim C^{r-1} smoothness only [Sak90, Lem. 3.2 (ii)], or omit the details of higher degrees of smoothness in their proof [Yi93a, Thm 3.1], [Yi93b, Thm 3.1]. For a proof that $\widehat{E}^s \in C^{r-1}$, see [Fen71, Thm 7] or [Sak90, App. B].

Liu06, Sch08, JT09] and the references therein. In the special case that the slow manifold is attracting, another application of the Fenichel Normal Form is to dimensionality reduction: in this normal form, the dynamics of the transformed slow variable are decoupled from the transient dynamics of the transformed fast variable, and therefore the transformed slow dynamics serves as a reduction of the full dynamics in a clear way. Stated differently, in the coordinates placing the system in Fenichel Normal Form, the map P^s sending stable fibers to their basepoints is simply an orthogonal projection; the coordinate change "straightens out" the stable fibers $W^s(m)$, for m in the slow manifold. We remark that the recent paper [JT09] is a useful source of historical information on the Fenichel Normal Form.

As mentioned above, for our applications to GSP we will assume the special case in which the critical manifold is a NAIM. This special case arises naturally in many concrete applications, such as in understanding nonholonomic dynamics as a limit of friction forces [Eld16], in biolocomotion [EJ16], in the context of chemical reactions and combustion [LG94], in various problems in control theory [KOS76], and many more [Kue15, Ch. 20]. For general background on GSP, one may consult, e.g., the seminal paper [Fen79], the expository articles [Kap99, Jon95], or the recent book [Kue15].

Our two applications are as follows.

1. Using Theorem II.5 and assuming that the slow manifold is a NAIM, we show that the Fenichel Normal Form is valid on the union of *global* stable manifolds $\cup_\epsilon W^s(K_\epsilon)$ of slow manifolds K_ϵ, rather than just on the union of local stable manifolds $\cup_\epsilon W^s_{\text{loc}}(K_\epsilon)$. This is the content of Theorem II.26.

2. Using Theorem II.24 and assuming that the slow manifold is a NAIM, we show that under additional spectral assumptions on the critical manifold, there exists a stronger normal form which is *linear* in the fast variables. This normal form can be viewed as a stronger version of the Fenichel Normal Form. Additionally, this linear normal form

is also valid on the union of *global* stable manifolds $\cup_\epsilon W^s(K_\epsilon)$ of slow manifolds K_ϵ. This is the content of Theorem II.28.

The remainder of this section is as follows. We first introduce the context for Theorems II.26 and II.28 by describing the GSP setup in §2.6.1. Next, §2.6.2 contains the global extension of the Fenichel Normal Form as an application of Theorem II.5. Following this, §2.6.3 contains the derivation of the linear normal form, as well as its global extension, as an application of Theorem II.24. Next, in in §2.6.4 we discuss our results and relate them to the so-called method of straightening out fibers (SOF method) recently appearing in the literature [KBS14]. Finally, in §2.6.5 we illustrate our results in an example involving a classical mechanical system.

2.6.1 Setup and classic results

Consider a singularly perturbed system of the form

$$\begin{aligned} x' &= f(x, y, \epsilon) \\ \epsilon y' &= g(x, y, \epsilon), \end{aligned} \tag{2.28}$$

where $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$ are functions of "slow time" τ, ϵ is a small parameter, and[13] $f, g \in C^{r \geq 2}$. For all $\epsilon \neq 0$, this system is equivalent via a time-rescaling $t = \tau/\epsilon$ to the regularized system

$$\begin{aligned} \dot{x} &= \epsilon f(x, y, \epsilon) \\ \dot{y} &= g(x, y, \epsilon). \end{aligned} \tag{2.29}$$

[13] Note that we have adopted Fenichel's convention of letting x denote the "slow" variable here, as a matter of personal style.

We let a "prime" denote a derivative with respect to τ, and a "dot" denote a derivative with respect to the "fast time" t.

Now suppose that $K_0 \subset \operatorname{int} \widehat{K}_0 \subset \widehat{K}_0$ are compact manifolds with boundary contained in $S \coloneqq \{(x, y) : g(x, y, 0) = 0\}$, with $\operatorname{int} \widehat{K}_0$ denoting the manifold interior of $\widehat{K}_0$. Noting that S consists of critical points of the ($\epsilon = 0$) system, let us assume that the eigenvalues of $\mathsf{D}_2 g(x, y, 0)$ have strictly negative real part on $\widehat{K}_0$. In particular, this implies that $\widehat{K}_0$ can be locally written as a graph $\widehat{K}_0 \coloneqq \{(x, F(x))\}$ over some domain $B \subset \mathbb{R}^{n_x}$.

By making local modifications to the vector field defined by (2.29) in arbitrarily small neighborhoods of ∂K_0 and $\partial \widehat{K}_0$, we may henceforth assume without loss of generality that the vector field is inward pointing at ∂K_0 and outward pointing at[14] $\partial \widehat{K}_0$.

By our assumption on the eigenvalues of $\mathsf{D}_2 g$, we have that $K_0 \times \mathbb{R}$ and $\widehat{K}_0 \times \mathbb{R}$ are noncompact NAIMs for the dynamics

$$\begin{aligned} \dot{x} &= 0 \\ \dot{y} &= g(x, y, 0) \\ \dot{\tilde{\epsilon}} &= 0, \end{aligned} \tag{2.30}$$

since the equations for $\dot{x}$ and $\dot{y}$ in (2.30) are independent of $\tilde{\epsilon}$. Here, $\tilde{\epsilon} \in \mathbb{R}$ is a new parameter, and its relation to ϵ will be determined subsequently. We compactify these NAIMs by replacing $\mathbb{R}$ with its one-point compactification S^1, and we thereby henceforth consider (2.30) to be defined on $\mathbb{R}^{n_x + n_y} \times S^1$. For this new domain of definition, $K_0 \times S^1$ and $\widehat{K}_0 \times S^1$ are compact inflowing and overflowing NAIMs, respectively.

Next, following [Eld13, p. 142], we use a scaling parameter $\kappa > 0$ to slowly "turn on" the $\tilde{\epsilon}$ dependence. Let $\chi \colon \mathbb{R} \to [0, 1]$ be a C^∞ nonnegative bump function such that $\chi \equiv 1$ on $[-1, 1]$ and $\chi \equiv 0$ on $\mathbb{R} \setminus (-2, 2)$, and — anticipating a parameter substitution $\epsilon = \kappa \tilde{\epsilon}$ —

[14]Similar constructions are carried out in greater detail in [Jos00, § 2].

consider the vector field defined by

$$
\begin{aligned}
\dot{x} &= \chi(\tilde{\epsilon})\kappa\tilde{\epsilon} f(x, y, \chi(\tilde{\epsilon})\kappa\tilde{\epsilon}) \\
\dot{y} &= g(x, y, \chi(\tilde{\epsilon})\kappa\tilde{\epsilon}) \\
\dot{\tilde{\epsilon}} &= 0.
\end{aligned}
\tag{2.31}
$$

One can verify that this vector field can be made arbitrarily C^r-close to (2.30) by taking $0 < \kappa \ll 1$ sufficiently small. It follows from Fenichel's theorem on persistence of overflowing NAIMs [Fen71, Thm 1] that there exists a $\kappa > 0$ such that $\widehat{K}_0 \times S^1$ persists to a C^r-nearby overflowing r-NAIM for (2.31), and that $K_0 \times S^1$ persists to an inflowing NAIM inside it, since $K_0 \times S^1 \subset \widehat{K}_0 \times S^1$ and inflowing invariance is an open condition. Because K_0 consisted entirely of critical points for (2.30), by a theorem of Fenichel the local stable foliation of the inflowing NAIM is C^{r-1} [Fen77, Thm 5].

We now make the change of variables $\epsilon = \kappa\tilde{\epsilon}$ and see that (2.31) is equivalent to

$$
\begin{aligned}
\dot{x} &= \chi(\epsilon/\kappa)\epsilon f(x, y, \chi(\epsilon/\kappa)\epsilon) \\
\dot{y} &= g(x, y, \chi(\epsilon/\kappa)\epsilon) \\
\dot{\epsilon} &= 0,
\end{aligned}
\tag{2.32}
$$

so it follows that (2.32) has compact r-NAIMs M and $\widehat{M}$ which are respectively inflowing and overflowing, and with M contained in the manifold interior of $\widehat{M}$. Since M and $\widehat{M}$ are the images of the NAIMs for (2.31) through a diffeomorphism, the local stable foliation $W^s_{\text{loc}}(M)$ of M for (2.32) is also C^{r-1} in all variables x, y, ϵ.

2.6.2 Globalizing the Fenichel Normal Form

Continuing the analysis of §2.6.1, we may apply Theorem II.5 to deduce that the leaves of the global stable foliation of M for (2.32) fit together to form a C^{r-1} disk bundle

$P^s\colon W^s(M) \to M$ isomorphic (as a disk bundle) to E^s. By the definition of χ we see that for $\epsilon \in [-\kappa, \kappa]$, (2.32) reduces to the system

$$
\begin{aligned}
\dot{x} &= \epsilon f(x, y, \epsilon), \\
\dot{y} &= g(x, y, \epsilon), \\
\dot{\epsilon} &= 0.
\end{aligned}
\tag{2.33}
$$

As in [Jon95], let us make the commonly made assumption[15] that $\widehat{K}_0$ is the graph of a map $B \subset \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$, where B is a closed ball in $\mathbb{R}^{n_x}$ — as we have already noted, by the implicit function theorem this can always be achieved by shrinking $\widehat{K}_0$ if necessary. Thus, if κ is sufficiently small, we can write M as the graph of a C^r map $y = F(x, \epsilon)$ defined on a suitable open subset of $\mathbb{R}^{n_x} \times S^1$. Making the coordinate change $(x, y, \epsilon) \mapsto (x, y - F(x, \epsilon), \epsilon)$, we may assume that M is contained in $\mathbb{R}^{n_x} \times \{0\} \times S^1$. Since we assumed that $\widehat{K}_0$ is contractible it follows that M deformation retracts onto S^1, and hence the bundle $P^s\colon W^s(M) \to M$ must be trivializable over any subset of the form $M_0 \coloneqq M \cap (\mathbb{R}^{n_x + n_y} \times (-\epsilon_0, \epsilon_0))$, for any sufficiently small $\epsilon_0 > 0$. It follows that there exists a C^{r-1} fiber-preserving diffeomorphism $W^s(M_0) \cong M_0 \times \mathbb{R}^{n_y}$ of the form $(x, y, \epsilon) \mapsto (\tilde{x}, \tilde{y}, \epsilon) \coloneqq (P^s(x, y, \epsilon), \phi(x, y, \epsilon), \epsilon)$, with $\phi(x, 0, \epsilon) \equiv 0$. Making this final coordinate change, it follows that when restricted to $W^s(M_0)$, the system (2.33) takes the form:

$$
\begin{aligned}
\dot{\tilde{x}} &= \epsilon h(\tilde{x}, \epsilon), \\
\dot{\tilde{y}} &= \Lambda(\tilde{x}, \tilde{y}, \epsilon)\tilde{y}, \\
\dot{\epsilon} &= 0,
\end{aligned}
\tag{2.34}
$$

[15] A more general situation where M cannot be written as a graph can be handled using a tubular neighborhood modeled on the normal bundle of M.

where $(\tilde{x}, \tilde{y}, \epsilon) \mapsto \Lambda(\tilde{x}, \tilde{y}, \epsilon)$ is a C^{r-3} family of $n_x \times n_x$ matrices and $(\tilde{x}, \epsilon) \mapsto h(\tilde{x}, \epsilon)$ is[16] C^{r-1}. The $\dot{\tilde{x}}$ equation depends only on $\tilde{x}$ and ϵ because we are using an invariant fiber bundle trivialization for coordinates on $W^s(M_0)$, and $\tilde{x}$ and $\tilde{\epsilon}$ are coordinates for M_0. By our choice of coordinates, $\dot{\tilde{x}}$ is zero when $\epsilon = 0$ because $\dot{x} = 0$ when $\epsilon = 0$ — this fact and Hadamard's Lemma implies that $\dot{\tilde{x}}$ is of the form $\epsilon h(\tilde{x}, \epsilon)$. Hadamard's Lemma similarly implies that $\dot{\tilde{y}}$ is of the form $\Lambda(\tilde{x}, \tilde{y}, \epsilon)\tilde{y}$, because after our coordinate changes M_0 corresponds to the set of points in $W^s(M_0)$ with $\tilde{y} = 0$, and also M_0 is positively invariant, so it must be the case that $\dot{\tilde{y}} = 0$ when $\tilde{y} = 0$. Suppressing the $\dot{\epsilon} = 0$ equation, we have proven the following result, which we record here as a theorem.

Theorem II.26. *Assume that $\widehat{K_0}$ can be written as the graph of a C^r map $B \subset \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$, with B a closed ball in $\mathbb{R}^{n_x}$. Then there exists $\kappa > 0$ such that for any $\epsilon \in [-\kappa, \kappa]$, there is a C^{r-1} fiber-preserving diffeomorphism $\varphi_\epsilon \colon W^s(K_\epsilon) \to K_\epsilon \times \mathbb{R}^{n_y}$ such that in the coordinates $\tilde{x}, \tilde{y} = \varphi_\epsilon(x, y)$, the system (2.28) takes the form*

$$\begin{aligned} \dot{\tilde{x}} &= \epsilon h(\tilde{x}, \epsilon), \\ \dot{\tilde{y}} &= \Lambda(\tilde{x}, \tilde{y}, \epsilon)\tilde{y}. \end{aligned} \tag{2.35}$$

In the new coordinates, K_ϵ corresponds to $\{(\tilde{x}, \tilde{y}) | \tilde{y} = 0\}$. The diffeomorphism ρ_ϵ is C^{r-1} in ϵ. Also, $h \in C^{r-1}$, $\Lambda \in C^{r-3}$, and the function $(\tilde{x}, \epsilon) \mapsto \epsilon h(\tilde{x}, \epsilon)$ is C^r.

Restricting attention now to only positive values of $\epsilon > 0$, in the original slow time-scale (2.35) is equivalent to

$$\begin{aligned} \tilde{x}' &= h(\tilde{x}, \epsilon), \\ \epsilon \tilde{y}' &= \Lambda(\tilde{x}, \tilde{y}, \epsilon)\tilde{y}. \end{aligned} \tag{2.36}$$

[16]However, the maps $(\tilde{x}, \epsilon) \mapsto \epsilon h(\tilde{x}, \epsilon)$ and $(\tilde{x}, \tilde{y}, \epsilon) \mapsto \Lambda(\tilde{x}, \tilde{y}, \epsilon)\tilde{y}$ are C^r and C^{r-2}, respectively. The first map is C^r because $\epsilon h(x, \epsilon) \equiv f(x, F(x, \epsilon), \epsilon)$, and the right hand side is C^r in x and ϵ.

Remark II.27. Because of our assumption that the critical manifold of (2.29) was a NAIM, the normal form which we were able to derive and state in Theorem II.26 appears considerably simpler than the *Fenichel Normal Form* — c.f. [Jon95, p. 82], [Kap99, pp. 109-111], [JT09, p. 973] or [Kue15, pp. 72–73], although our normal form actually directly follows from the general Fenichel Normal Form. Our contribution is that, using Theorem II.5, we have shown that this normal form is valid on a neighborhood which consists of the entire union of *global* stable manifolds $\cup_\epsilon W^s(K_\epsilon)$, as opposed to being valid merely on the union of local stable manifolds $\cup_\epsilon W^s_{\text{loc}}(K_\epsilon)$.

2.6.3 Smooth global linearization: A stronger GSP normal form

In this section we continue to assume that the critical manifold is a NAIM for (2.29), but we make the following additional "nonresonance" assumption on the eigenvalues of the critical points. Let $r_{\min}(x,y) \leq r_{\max}(x,y) < 0$ denote the minimum and maximum real parts of eigenvalues of $D_2 g(x,y)$, where $(x,y) \in \widehat{K}_0$, and $\widehat{K}_0$ is defined following (2.29). We assume that there exist negative real constants α, β such that $2\alpha < \beta < \alpha < 0$ and

$$\forall (x,y) \in \widehat{K}_0 : \beta < r_{\min}(x,y) \leq r_{\max}(x,y) < \alpha. \tag{2.37}$$

The payoff for this assumption is that we can obtain a C^{r-1} normal form which is *linear* in $\tilde{y}$, improving upon the Fenichel Normal Form (2.35) significantly. This normal form is also global in the sense that it holds on the entire union of global stable manifolds $\cup_\epsilon W^s(K_\epsilon)$. This is the content of the following result.

Theorem II.28. *Assume that the vector field defined by* (2.28) *is* $C^{r\geq 3}$*, and assume that the condition* (2.37) *holds for the regularized system* (2.29)*, and that* $\widehat{K}_0$ *can be written as the graph of a* C^r *map* $B \subset \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$*, with* B *a closed ball in* $\mathbb{R}^{n_x}$*. Then there exists* $\kappa > 0$ *such that for any* $\epsilon \in [-\kappa, \kappa]$*, there is a* C^{r-1} *fiber-preserving diffeomorphism*

$\varphi_\epsilon\colon W^s(K_\epsilon) \to K_\epsilon \times \mathbb{R}^{n_y}$ such that in the coordinates $\tilde{x}, \tilde{y} = \varphi_\epsilon(x, y)$, the system (2.28) takes the form

$$\begin{aligned} \dot{\tilde{x}} &= \epsilon h(\tilde{x}, \epsilon), \\ \dot{\tilde{y}} &= A(\tilde{x}, \epsilon)\tilde{y}. \end{aligned} \tag{2.38}$$

In the new coordinates, K_ϵ corresponds to $\{(\tilde{x}, \tilde{y}) | \tilde{y} = 0\}$. The diffeomorphism φ_ϵ is C^{r-1} in ϵ. Also, $h \in C^{r-1}$, $A \in C^{r-1}$, and the function $(\tilde{x}, \epsilon) \mapsto \epsilon h(\tilde{x}, \epsilon)$ is C^r.

Restricting attention now to only positive values of $\epsilon > 0$, in the original slow time-scale (2.38) is equivalent to

$$\begin{aligned} \tilde{x}' &= h(\tilde{x}, \epsilon), \\ \epsilon\tilde{y}' &= A(\tilde{x}, \epsilon)\tilde{y}. \end{aligned} \tag{2.39}$$

If the condition (2.37) does not hold, then there exists a homeomorphism φ_ϵ such that the same result holds, but φ_ϵ is generally not differentiable in that case.

Proof. Consider the compact inflowing NAIM M for the system (2.32) defined on $\mathbb{R}^{n_x+n_y} \times S^1$. As described in §2.6.2, our assumption that $\widehat{K_0}$ is a graph implies that if $\kappa > 0$ is sufficiently small, then M is the graph of a C^r map $(x, \epsilon) \mapsto y$. Hence we may assume without loss of generality that $M \subset \mathbb{R}^{n_x} \times \{0\} \times S^1$.

Let E^s be the stable vector bundle of M and let Φ^t be the flow of the system (2.32) on $\mathbb{R}^{n_x+n_y} \times S^1$. By continuity and compactness, it can be shown that assumption (2.37) implies that if $\kappa > 0$ is sufficiently small, then there exist constants $\delta > 0$ and $K \geq 1$ such that $-\alpha > r\delta$, and such that for all $t \geq 0$

$$\begin{aligned} K^{-1}e^{-\delta t} &\leq [\![\mathsf{D}\Phi^t|_{TM}]\!] \leq \|\mathsf{D}\Phi^t|_{TM}\| \leq Ke^{\delta t}, \\ K^{-1}e^{-\delta t} &\leq [\![(\mathsf{D}\Phi^t|_{TM})^{-1}]\!] \leq \|(\mathsf{D}\Phi^t|_{TM})^{-1}\| \leq Ke^{\delta t}, \\ K^{-1}e^{\beta t} &\leq [\![\mathsf{D}\Phi^t|_{E^s}]\!] \leq \|\mathsf{D}\Phi^t|_{E^s}\| \leq Ke^{\alpha t} \end{aligned} \tag{2.40}$$

uniformly on TM and E^s. By Theorem II.24, there exists a global C^{r-1} fiber-preserving diffeomorphism $\varphi\colon W^s(M) \to E^s$ which conjugates $\Phi^t|_{W^s(M)}$ to $\mathsf{D}\Phi^t|_{E^s}$ and maps M diffeomorphically onto the zero section of E^s.

Now in any local trivialization of E^s, the vector field generating the flow $\mathsf{D}\Phi^t|_{E^s}$ is of the form (2.38) augmented with $\dot{\epsilon} = 0$ (where coordinates for the zero section are given by $\tilde{x}$ and coordinates for the fibers given by $\tilde{y}$). It follows that if we define $\varphi_\epsilon(\,\cdot\,,\,\cdot\,) \coloneqq \varphi(\,\cdot\,,\,\cdot\,,\epsilon)$, then it suffices to show that $W^s(M)$ is trivializable over the subset $M \cap (\mathbb{R}^{n_x+n_y} \times [-\kappa,\kappa])$. But $M \cap (\mathbb{R}^{n_x+n_y} \times [-\kappa,\kappa])$ is contractible since it is diffeomorphic to $K_0 \times [-\kappa,\kappa]$, so $W^s(M)$ is indeed trivializable over M_κ.

The statement about (2.39) follows easily by replacing t with the rescaled slow time $\tau = \epsilon t$.

Finally, to justify the last statement for the case that (2.37) does not hold, we simply apply Corollary II.25 instead of Theorem II.24. This completes the proof. $\square$

Remark II.29. Assume that $n_y = \dim(y) = 1$, so that the fast variable is one-dimensional and the slow manifold is codimension-1. Then the eigenvalue condition (2.37) can always be made to hold by taking $\widehat{K_0}$ sufficiently small.

Remark II.30. We see from the proof that, since K_ϵ is a manifold with boundary, our linearization result Theorem II.24 for inflowing invariant manifolds is crucial. This is because, to the best of our knowledge, all of the linearization results in the literature assume a boundaryless invariant manifold [PS70, Tak71, Rob71, HPS77, PT77, Sel84, Sel83, Sak94, BK94].

2.6.4 Discussion

We have proven Theorems II.26 and II.28, both of which are statements about normal forms for slow-fast systems in the framework of geometric singular perturbation theory (GSP). These results assume that the slow manifold is attracting.

Let us first discuss some literature regarding the Fenichel Normal Form for attracting slow manifolds, which is the subject of Theorem II.26. Because of the practical benefits afforded by dimensionality reduction, there has been interest in actually *computing* the coordinate change placing the system in Fenichel Normal Form for the attracting slow manifold case. Recently, the so-called method of straightening out fibers (SOF method) has been developed to iteratively approximate the Taylor polynomials of this coordinate change[17] [KBS14]; similar techniques for systems near equilibria were previously developed in [Rob89, Rob00], and we also mention that it was shown in [ZKK04] that the Computational Singular Perturbation (CSP) method initially developed in [LG89, Lam93] iteratively approximates the first-order Taylor polynomial of this coordinate change.

Theorem II.26 does not yield a new normal form; it shows that, in the attracting slow manifold case, the domain of the coordinate change placing the system in Fenichel Normal Form actually extends to the entire global stable manifold of the slow manifold. This result seems to be of primarily theoretical interest. For example, the state-of-the-art SOF method only provides a means for computing Taylor polynomials centered at the slow manifold. Since these Taylor polynomials are only guaranteed to accurately approximate the coordinate near the slow manifold, they are unlikely to approximate the global coordinate change. Hence the global coordinate change, guaranteed to exist by Theorem II.26, might not be explicitly computable except in special cases.

On the other hand, Theorem II.28 does yield a new normal form, and also shows that

[17] The results of [KBS14] actually apply in more general situations, such as the case of a normally elliptic slow manifold. See [KBS14] for more details.

the domain of the associated coordinate change extends to the entire global stable manifold. In order for the coordinate change to be differentiable, some additional spectral conditions (2.37) need to be satisfied, although these are automatically satisfied on a small enough domain of the slow manifold in the codimension-1 case (see Remark II.29). The payoff is that this normal form is *linear* in the fast variables. Furthermore, by combining the SOF method of [KBS14] with additional normal form computations [GH83, Rob89, Rob00] for the fast variable, it seems to us that it should be possible in principle to compute the Taylor polynomials of this coordinate change in a systematic way. We hope to explore this in future work. Of course, computing this coordinate system *globally* suffers the same difficulties mentioned in the previous paragraph. Finally, we observe that our normal form is quite similar in form to the dynamics produced by "high-gain" nonlinear control schemes — suggesting that linearly controlled fast variables are an inherent feature of a broad class of systems, rather than a convenient requirement imposed by control theorists.

2.6.5 Example

In this section, we consider an example of a forced pendulum with damping. This example was chosen so that the natural state space is not Euclidean. This will allow us to illustrate Theorems II.5, II.18, and II.24 by directly applying these theorems to obtain stronger results than those obtainable via Theorems II.26 and II.28, which we formulated for dynamics on a Euclidean space.

We allow the damping coefficient of the pendulum to be a function of the pendulum angle, and consider an applied torque which depends on the pendulum angle and time. We assume that the applied torque is periodic in time, and for simplicity we assume that the period is 2π. Specifically, we consider the equations of motion

$$\epsilon\theta'' + \frac{\epsilon g}{l}\sin\theta + c(\theta)\theta' = \tau(\theta, t), \tag{2.41}$$

where ϵ is the pendulum mass which we assume to be small[18], l is the pendulum length, g is the acceleration due to gravity, c is the angle-dependent damping coefficient, and $\tau(\theta, t)$ is the applied torque — not to be confused with the slow time variable that is also denoted by τ with some abuse of notation. We are assuming that $\forall \theta, t : \tau(\theta, t + 2\pi) = \tau(\theta, t)$. We define the angular velocity $\omega \coloneqq \theta'$. The periodicity of τ allows us to introduce a circular coordinate α and write (2.41) in the following extended state space form:

$$\begin{aligned} \theta' &= \omega \\ \alpha' &= 1 \\ \epsilon\omega' &= -\frac{\epsilon g}{l}\sin\theta - c(\theta)\omega + \tau(\theta, \alpha). \end{aligned} \tag{2.42}$$

We consider (θ, α) to be angle coordinates on the two-torus $T^2 \coloneqq S^1 \times S^1$, so that the state space is $T^2 \times \mathbb{R}$. As in §2.6.1, for $\epsilon \neq 0$ this "slow time" system is equivalent via a time-rescaling $t = \tau/\epsilon$ to the "fast time" system

$$\begin{aligned} \dot{\theta} &= \epsilon\omega \\ \dot{\alpha} &= \epsilon \\ \dot{\omega} &= -\frac{\epsilon g}{l}\sin\theta - c(\theta)\omega + \tau(\theta, \alpha). \end{aligned} \tag{2.43}$$

To relate this to our earlier notation from §2.6.1, here (θ, α) is playing the role of x and ω is playing the role of y. For $\epsilon = 0$, the set of critical points of (2.43) are given by

[18] Strictly speaking, in a physical context we should define ϵ to be a dimensionless quantity in order to refer to it as "small" in an absolute sense. However, this will cause no problem whatsoever for applying and illustrating our results, and we therefore do not bother with this.

$S \coloneqq \{(\theta, \alpha, \omega) \colon c(\theta)\omega = \tau(\theta, \alpha)\}$.

Let us first consider the special case of a constant positive damping coefficient $c(\theta) \equiv c_0 > 0$. Then S is the graph of the map $F_0(\theta, \alpha) \coloneqq \frac{1}{c_0}\tau(\theta, \alpha)$. We henceforth assume that $\tau \in C^r$, with[19] $3 \le r < \infty$. It follows that S is a C^r manifold diffeomorphic to the torus T^2. Furthermore, the eigenvalues of all critical points in the critical manifold S are readily checked to be $(0, 0, -c_0)$, with the zero eigenvalues corresponding to the tangent spaces of S and $-c_0$ corresponding to $\operatorname{span}\{(0, 0, 1)\}$. Therefore, S is an r-NAIM for (2.43) when $\epsilon = 0$. Since $\partial S = \varnothing$, there exists $\epsilon_0 > 0$ such that for all $0 \le \epsilon \le \epsilon_0$, there is a unique persistent NAIM S_ϵ close to S, with $S_0 = S$. As in §2.6.2, S_ϵ is the graph of a C^r map $\omega = F(\theta, \alpha, \epsilon)$ with $F_0 = F(\,\cdot\,, \,\cdot\,, 0)$.

Using a technique from [SW99], we next prove the following proposition.

Proposition II.31. *For all sufficiently small $\epsilon > 0$, S_ϵ is globally asymptotically stable. In other words, for all sufficiently small $\epsilon > 0$, we have $W^s(S_\epsilon) = T^2 \times \mathbb{R}$.*

Proof. We already know that S_ϵ is locally asymptotically stable for $\epsilon > 0$ sufficiently small, so it suffices to show that S_ϵ is globally attracting for $\epsilon > 0$ sufficiently small. We fix any $\epsilon_0 > 0$ and define

$$\eta \coloneqq \frac{1}{c_0}\left(\frac{\epsilon_0 g}{l} + \max_{(\theta,\alpha)\in T^2} |\tau(\theta, \alpha)| + 1\right).$$

Note that for all $0 \le \epsilon \le \epsilon_0$, the compact subset

$$D_\eta \coloneqq \{(\theta, \alpha, \omega) \colon |\omega| < \eta\}$$

of $T^2 \times \mathbb{R}$ is positively invariant, and every point in $(T^2 \times \mathbb{R}) \setminus D_\eta$ will flow into D_η in some finite time; indeed, $\dot{\omega} < -1$ on $(T^2 \times \mathbb{R}_{\ge 0}) \setminus D_\eta$, $\dot{\omega} > 1$ on $(T^2 \times \mathbb{R}_{\le 0}) \setminus D_\eta$, and

[19]Even if $r = \infty$, we can only derive results for a finite smoothness degree. This is because persistent NHIMs generally have only a finite degree of smoothness, even if the dynamics are C^∞ and the spectral gap is infinite [Eld13, Remark 1.12].

the vector field points inward at ∂D_η. Therefore it suffices to show such that S_ϵ attracts all states in D_η for sufficiently small $\epsilon > 0$. Next, by the same reasoning as in §2.6.1, we know that the compact set $\bigcup_{0\leq\epsilon\leq\epsilon_0} S_\epsilon$ is locally asymptotically stable for the augmented dynamics (adding $\dot{\epsilon} = 0$) on $T^2 \times \mathbb{R} \times \mathbb{R}$. Hence there exists $\delta > 0$ such that for all $\epsilon > 0$ sufficiently small, the basin of attraction of S_ϵ contains the set N_δ of points $(\theta, \alpha, \omega) \in T^2 \times \mathbb{R}$ satisfying $|\omega - F(\theta, \alpha, \epsilon)| < \delta$. In order to obtain a contradiction, suppose that there exist arbitrarily small values of $\epsilon > 0$ such that S_ϵ does not attract all states in D_η, and let Φ^t_ϵ denote the flow of (2.43). Then there exist sequences $(\epsilon_n)_{n\in\mathbb{N}}$ and $(\theta_n, \alpha_n, \omega_n)_{n\in\mathbb{N}} \subset D_\eta$ such that $\epsilon_n \to 0$ and $\forall t > 0, n > 0 : \Phi^t(\theta_n, \alpha_n, \omega_n) \notin N_\delta$. Since D_η is compact, by passing to a subsequence we may assume that $(\theta_n, \alpha_n, \omega_n) \to (\theta_0, \alpha_0, \omega_0) \in D_\eta$. Since S_0 is globally asymptotically stable for Φ^t_0, for all sufficiently large $t > 0$, $\Phi^t_0(\theta_0, \alpha_0, \omega_0) \in N_{\delta/2}$. By continuity of the map $(t, \epsilon, \theta, \alpha, \omega) \mapsto \Phi^t_\epsilon(\theta_\epsilon, \alpha_\epsilon, \omega_\epsilon)$, it follows that for all sufficiently large $t, n > 0$, $\Phi^t_{\epsilon_n}(\theta_n, \alpha_n, \omega_n) \in N_\delta$. This is a contradiction, showing that for all sufficiently small $\epsilon > 0$, $W^s(S_\epsilon) = T^2 \times \mathbb{R}$ for the dynamics (2.43).

□

Because the eigenvalues of the critical manifold are $(0, 0, -c_0)$, after taking ϵ_1 smaller if necessary we see that the r-center bunching conditions (2.11) are satisfied. Therefore, Proposition II.31 and Corollary II.10 of Theorem II.5 show that there exists a a C^{r-1} diffeomorphism $\varphi_\epsilon : T^2 \times \mathbb{R} \to T^2 \times \mathbb{R}$ mapping S_ϵ onto $T^2 \times \{0\}$ and mapping stable fibers of S_ϵ onto sets of the form[20] $(\theta, \alpha) \times \mathbb{R}$. Using the coordinates $\tilde{\theta}, \tilde{\alpha}, \tilde{\omega} = \varphi_\epsilon(\theta, \alpha, \omega)$ and changing

[20] Here, and during the rest of this example, we are using the fact that the normal bundle — and hence also the stable bundle E^s — of the slow manifold is trivial.

back to the original time scale, (2.42) takes the form

$$\begin{aligned}\tilde{\theta}' &= F(\tilde{\theta}, \tilde{\alpha}, \epsilon) \\ \tilde{\alpha}' &= 1 \\ \epsilon\tilde{\omega}' &= \Lambda(\tilde{\theta}, \tilde{\alpha}, \tilde{\omega}, \epsilon)\tilde{\omega},\end{aligned} \tag{2.44}$$

for some function Λ. The same reasoning as in §2.6.2 can be used to show that φ_ϵ is jointly C^{r-1} in all variables including ϵ. This result should be compared with Theorem II.26, which was formulated for dynamics on a Euclidean space. We see that Theorem II.5 yields a global coordinate system on all of $T^2 \times \mathbb{R}$ placing (2.42) in the form (2.44). In contrast, without Theorem II.5 and using only the available results in the literature, we would have only been able to obtain such a coordinate system on a precompact neighborhood of S_ϵ.

Alternatively, because the eigenvalues of the critical manifold are $(0, 0, -c_0)$, after taking ϵ_1 smaller if necessary we see that the stronger spectral conditions of Theorem II.24 are also satisfied (c.f. (2.37)). Hence Theorem II.24 implies that there exists a global C^{r-1} diffeomorphism $\psi_\epsilon \colon T^2 \times \mathbb{R} \to T^2 \times \mathbb{R}$ mapping S_ϵ onto $T^2 \times \{0\}$ and mapping stable fibers of S_ϵ onto sets of the form $(\theta, \alpha) \times \mathbb{R}$. Using the coordinates $\tilde{\theta}, \tilde{\alpha}, \tilde{\omega} = \psi_\epsilon(\theta, \alpha, \omega)$ and changing back to the original time scale, (2.42) takes the form

$$\begin{aligned}\tilde{\theta}' &= F(\tilde{\theta}, \tilde{\alpha}, \epsilon) \\ \tilde{\alpha}' &= 1 \\ \epsilon\tilde{\omega}' &= A(\tilde{\theta}, \tilde{\alpha}, \epsilon)\tilde{\omega},\end{aligned} \tag{2.45}$$

for some function A. The same reasoning as in §2.6.2 can be used to show that φ_ϵ is jointly C^{r-1} in all variables including ϵ. This result should be compared with Theorem II.28, which was formulated for dynamics on a Euclidean space. We used Theorem II.24 to derive (2.45), but since $\partial S_\epsilon = \varnothing$ this result can also be obtained by combining Theorem II.18 with the local smooth linearization results of [Sak94].

Still considering (2.42), we will now consider specific choices of a non-constant damping function $c(\theta)$ and applied torque $\tau(\theta, \alpha)$ which will be chosen so that Theorem II.18 does not apply, but so that Theorem II.24 does apply to yield a linear normal form. For the sake of concreteness, let $c(\theta) \coloneqq \cos(\theta) + 1$ and $\tau(\theta, \alpha) \coloneqq -\sin(\theta) + (1/2)\cos(\alpha)$. Then $c(\pi) = 0$, so it follows that the critical set $S \coloneqq \{(\theta, \alpha, \omega) \colon c(\theta)\omega = \tau(\theta, \alpha)\}$ is not normally hyperbolic for the fast time system (2.43) everywhere. However, e.g. $c(\theta) > 1$ for $|\theta| < \pi/2$, so it follows in particular that the subset $K_0 \coloneqq \{(\theta, \alpha, \omega) \in S \colon |\theta| \leq \pi/4\}$ is r-normally attracting. Furthermore, K_0 is inflowing for the slow time system (2.42) restricted to S when $\epsilon = 0$, because K_0 is the graph of $F(\theta, \alpha, 0)$ with

$$F(\theta, \alpha, 0) \coloneqq \frac{\tau(\theta, \alpha)}{c(\theta)} = \frac{-\sin(\theta) + (1/2)\cos(\alpha)}{\cos(\theta) + 1}$$

with $|\theta| \leq \pi/4$. Therefore, the projection of the slow time dynamics restricted to K_0 are given by

$$\theta' = \frac{-\sin(\theta) + (1/2)\cos(\alpha)}{\cos(\theta) + 1}$$

$$\alpha' = 1$$

and clearly the vector field points inward at the boundary of $\{(\theta, \alpha) \colon |\theta| \leq \pi/4\}$. We can modify the flow locally near the boundary of any larger set $\widehat{K}_0 \supset K_0$ to render $\widehat{K}_0$ overflowing, and therefore there exists $\epsilon_0 > 0$ such that for all $0 \leq \epsilon \leq \epsilon_0$, $\widehat{K}_0$ (and hence also K_0) persists to a nearby r-NAIM for the fast time system (2.43). Since inward pointing of a vector field is an open condition, after possibly shrinking ϵ_0 it follows that K_ϵ is also inflowing for all $0 \leq \epsilon \leq \epsilon_0$. Additionally, after possibly shrinking ϵ_0, we see that the hypotheses of Theorem II.24 are satisfied for K_ϵ for all $0 \leq \epsilon \leq \epsilon_0$ (check that (2.37) is satisfied on K_0 by using

$\alpha = -\sqrt{2}/2 - 1 \approx 1.7$ and $\beta = -2$, and use the fact that the hypotheses of Theorem II.24 are open conditions). Hence Theorem II.24 implies that there exists a C^{r-1} diffeomorphism $\psi_\epsilon \colon W^s(K_\epsilon) \to K_\epsilon \times \mathbb{R}$ mapping K_ϵ onto $K_\epsilon \times \{0\}$ and mapping stable fibers of K_ϵ onto sets of the form $\{(\theta, \alpha)\} \times \mathbb{R}$. Using the coordinates $\tilde{\theta}, \tilde{\alpha}, \tilde{\omega} = \psi_\epsilon(\theta, \alpha, \omega)$ and changing back to the original time scale, (2.43) takes the form

$$\begin{aligned} \tilde{\theta}' &= F(\tilde{\theta}, \tilde{\alpha}, \epsilon) \\ \tilde{\alpha}' &= 1 \\ \epsilon\tilde{\omega}' &= A(\tilde{\theta}, \tilde{\alpha}, \epsilon)\tilde{\omega}, \end{aligned} \tag{2.46}$$

for suitable functions A and F. The same reasoning as in §2.6.2 can be used to show that φ_ϵ is jointly C^{r-1} in all variables including ϵ. This result should be compared with Theorem II.28, which was formulated for dynamics on a Euclidean space. Here we had to use Theorem II.24 to derive (2.46), because Theorem II.18 does not apply since $\partial K_\epsilon \neq \varnothing$. Without Theorem II.24 and using only the explicitly available results in the literature, we would not have been able to obtain even a local version of this coordinate system.

Finally, we note that the Taylor polynomials of the coordinate change for the normal form (2.44) can in principle be obtained using the SOF method, although as mentioned in §2.6.4 this does not help to compute the coordinates *globally.* We do not pursue this here. As mentioned in §2.6.4, we believe it should be possible in principle to additionally compute the Taylor polynomials of the coordinate changes for the normal forms (2.45) and (2.46), which we hope to explore in future work.

2.7 Discussion

Stated technically, we have proven some results for NHIMs which are of two types: (i) global versions of well-known local results, and (ii) linearization results for inflowing NAIMs.

We restricted our attention to flows.

We first showed that the global stable foliation of an inflowing NAIM is a fiber bundle, with fibers coinciding with the leaves of the global stable foliation, and that this fiber bundle is as smooth as the local stable foliation. From that result, we deduced the corresponding result for the global (un)stable foliation of a general NHIM, though one needs to be careful in interpreting this statement as the global (un)stable manifold is generally only an immersed submanifold of Q.

We next considered global linearizations, and showed that the linearization result of [PS70, HPS77] for boundaryless NHIMs applies also to inflowing NAIMs. Furthermore, this linearization extends to the entire global stable manifold — inflowing NAIMs are globally linearizable, or topologically conjugate to the flow linearized at the NAIM. If some additional spectral gap conditions are assumed, then the global linearizing conjugacy can be taken to be C^k. This extends the results of [LM13] to the case of arbitrary inflowing NAIMs (although see Remark II.17). A key tool in our proof was the geometric construction of Appendix B, which allowed us to reduce to the boundaryless case.

We then used our theoretical results to give two applications to slow-fast systems with attracting slow manifolds, in the context of geometric singular perturbation theory (GSP). First, using our fiber bundle theorem we extended the domain of the Fenichel Normal Form [JK94, Jon95, Kap99]. Second, under an additional spectral gap assumption, we derived a global smooth *linear* normal form for GSP problems. If the slow manifold is codimension-1, this assumption can always be made to hold (after possibly shrinking the slow manifold; see Remark II.29). For this application it was essential that we proved a linearization theorem for inflowing NAIMs, since the slow manifolds appearing in slow-fast systems typically have boundary. We then illustrated these results on an example of a mechanical system. We noted that it might be interesting to combine the method of straightening out fibers (SOF method) of [KBS14] with additional normal form computations [GH83, Rob89, Rob00] for the fast variable, in order to develop a systematic technique for computing the Taylor polynomials of the coordinate change for the linear normal form. We hope to explore this idea in future work.

Less formally, what we have shown is that the local structure next to an inflowing NAIM extends globally, in terms of structure (as a disk bundle), in its degree of smoothness, and in the fact that the dynamics are often conjugate to their linearization. In fact, the linearization is so robust that it can be extended consistently to yield a system linear in its fast variables throughout all sufficiently small perturbations of a singularly perturbed system.

We have considered only compact NHIMs and compact inflowing NAIMs in stating our results. From our experience, we expect that extending these results to noncompact manifolds should be possible, but possibly quite technical. However, our results for compact inflowing NAIMs allow our work to be applied to (for example) positively invariant compact subsets of the phase space of a mechanical system.

CHAPTER III

Estimating Phase from Observed Trajectories Using the Temporal 1-Form

3.1 Acknowledgements

This chapter is based on joint work with Simon Wilshin, Clayton Scott, John Guckenheimer, and Shai Revzen. In particular, Wilshin deserves the credit for the figures in this chapter, and for writing the code used to generate these figures. Kvalheim would like to thank Peter J. Thomas for useful comments on stochastic differential equations, which ultimately pointed him in the right direction regarding Theorems III.7 and III.10. We are indebted to Daniel Guralnik and G. Bard Ermentrout for valuable insights. This work was supported by ARO Young Investigator Award #61770 and W911NF-14-1-0573 to S. Revzen. The guineafowl kinematics were gathered under the supervision of M. Daley with the support of a Human Frontier Science Program grant (RGY0062/2010) and ethical approval was obtained from Royal Veterinary College Ethics and Welfare Committee under the protocol title, "Kinematics and kinetics in birds running over uneven terrain".

3.2 Introduction

In this chapter we shift our attention to oscillators, a special case of the dynamical systems studied in Chapter II. Our interest in oscillators arose from their utility as models of animal locomotion, but oscillators appear in virtually every physical science: in biological models at all scales, from the coupling neuronal firing [HI97] to coupled oscillations of predator and prey populations [May72]; in chemistry [EP98]; in physics [GW02]; in electrical [VdP34], civil [Duf18], and mechanical [Sto50] engineering; etc. For oscillations which are robustly stable under perturbations, their underlying mathematical structure shares many properties across all these cases. One of the most significant of these properties is the existence of an *asymptotic phase* which encodes the long term outcome of any (recoverable) perturbation. We have discovered that by representing phase as a *Temporal 1-Form*, we could bring to bear interpolation tools from machine learning to allow asymptotic phase to be computed in a data-driven way — directly from collections of arbitrarily short time series of measurements — and without a need to know the governing equations. This would allow investigators to perform *phase reduction* directly from experimental measurements, and thereby construct models of oscillatory systems and how they couple to each other on a broad swath of real world systems. In this chapter we introduce the Temporal 1-Form and an algorithm to compute it. We test our algorithm using simulations with known asymptotic phase, as well as animal locomotion data.

3.3 Mathematical background

In the sciences we often model the time evolution of a system using a smooth flow Φ^t on a smooth n-dimensional manifold $\mathbf{X}$ (frequently $\mathbb{R}^n$) such that $\Phi : \mathbf{X} \times \mathbb{R} \to \mathbf{X}$, $\Phi^0(x) = x$, and $\Phi^t(\Phi^s(x)) = \Phi^{t+s}(x)$. We define a vector field $f(x) := \frac{\partial}{\partial t}\Phi^t(x)|_{t=0}$ and assume that it is C^2, so that the *ordinary differential equation* (ODE) $\dot{x} = f(x)$ generates the C^2 flow

Φ^t; familiar equations like Newton's laws, chemical reaction equations, or the dynamics of electrical circuits are usually stated in terms of these ODEs, but are often easier to analyze mathematically by viewing the flow as the primary object of study. The curve $t \mapsto \Phi^t(x)$ obtained by fixing x and letting t vary is the *trajectory* or *orbit* of the flow with initial condition x, sometimes denoted $x(t)$. A *periodic orbit* γ of period $T > 0$ is a trajectory that satisfies $\gamma(t+T) = \gamma(t)$ with $\gamma(t) \neq \gamma(0)$ for all $0 < t < T$. We focus on dynamical systems possessing only one nontrivial periodic trajectory γ and denote by $\Gamma \subset \mathbf{X}$ the image of γ.

We further assume that this periodic trajectory is exponentially stable, i.e., the distance of any trajectory in the stability basin to the limit cycle satisfies $d(x(t), \Gamma) < Cd(x(0), \Gamma)e^{-\lambda t}$ for some $\lambda, C > 0$, where d is the distance induced by some Riemannian metric. This property is generic in the sense that a stable periodic orbit Γ persists under all C^1-small perturbations if and only if it is exponentially stable, which is equivalent to Γ being attracting and normally hyperbolic (see Chapter II or [HPS77]). For brevity, we will henceforth simply refer to such exponentially stable periodic orbits as *limit cycles*. We refer to the stability basin of a limit cycle and the dynamics within it as an *oscillator*.

The asymptotic behavior of any oscillator is described fully by its *asymptotic phase*, which is the map $P^s \colon W^s(\Gamma) \to \Gamma$ of Chapter II; see also [Hal80, p. 221, Theorem 2.1]. P^s is a retraction ($P^s|_\Gamma = \mathrm{id}|_\Gamma$) and a semiconjugacy[1] ($\forall t \in \mathbb{R} : P^s \circ \Phi^t = \Phi^t \circ P^s$). The stability basin $W^s(\Gamma)$ of the oscillator is partitioned into manifolds of points $W^s(x) = (P^s)^{-1}(x), x \in \Gamma$ having the same asymptotic phase; these manifolds are called *isochrons* (see Chapter II or [Guc75, Win80]). $W^s(x)$ can be characterized as the set of initial conditions whose trajectories coalesce with the trajectory starting at x. We note that the geometry of isochrons can be surprisingly complicated, especially when the vector field f has "multiple time scales" [OM10, MRMM14, LKO14, LKO15].

[1] It follows that $P^s \in C^2$ since $\forall x \in W^s(\Gamma) : P^s(x) = o(\tau(x))$, where $\tau : W^s(\Gamma) \to \mathbb{R}$ is defined by $P^s \circ \Phi^{-\tau(x)}(x) = o(0)$. By the implicit function theorem, τ — and hence P^s — are C^2.

The main problem motivating this chapter is the estimation of asymptotic phase of an oscillator from empirically measured trajectories, without recourse to the underlying equations of motion. Thus in the practical situations of interest our algorithm will necessarily be subject to noise in some way, since measurements are always subject to uncertainty. We are interested both in situations where the noise comes from measurement errors, and in situations where the vector field is subject to stochastic and/or deterministic perturbations. In this chapter we show how to estimate asymptotic phase when the dynamics are observed empirically, e.g. as an ensemble of noisy measurements from system trajectories.

While there has been work to define generalized notions of asymptotic phase for stochastic oscillators [SP10, SP13, TL14], in this chapter we restrict ourselves to estimation of the classical asymptotic phase of a deterministic oscillator using data from a possibly perturbed version of the underlying deterministic system. An interesting direction for future work is the possibility of modifying our algorithm to compute one or both of these notions of stochastic phase.

3.3.1 Classical phase and the Temporal 1-Form

Classically, phase is taken to be angle-valued or *phasor*-valued, i.e. S^1-valued. This kind of phase can be constructed by first taking a C^2 map $\varphi|_\Gamma : \Gamma \to S^1 \subset \mathbb{C}$ with the property that $\varphi|_\Gamma(\gamma(t)) = \exp(2\pi i \frac{t}{T})\varphi|_\Gamma(\gamma(0))$, and then extending to $W^s(\Gamma)$ by defining $\varphi(x) := \varphi|_\Gamma \circ P^s(x)$. Such a phase is determined up to the choice of $\gamma(0) \in \Gamma$, and its level sets are the isochrons.

We now define the Temporal 1-Form and show that it appears naturally as a consequence of the existence of the asymptotic phase map P^s. The vector field f is nowhere zero on Γ, and Γ is one-dimensional. Thus there exists a unique C^2 differential 1-form ω on Γ satisfying $\langle \omega, f \rangle = 1$ identically on Γ (in Euclidean space, viewing f as a column vector, this would be $\omega := f^\mathsf{T}/\|f\|^2$). The *Temporal 1-Form* $\phi \colon W^s(\Gamma) \to \mathsf{T}^* W^s(\Gamma)$ is defined to be the pullback

$\phi := (P^s)^*\omega$ of ω via P^s. Hence the Temporal 1-Form has the following two properties: (1) $\langle \phi, f \rangle = 1$ everywhere; (2) it is *closed*, i.e. the exterior derivative $\mathbf{d}\phi = 0$.

The following theorem shows that the Temporal 1-Form is uniquely defined by these two properties. We give a proof that does not use hyperbolicity of the limit cycle.

Theorem III.1. *Let a dynamical system have an asymptotically stable limit cycle Γ. If a C^1 Temporal 1-Form exists on any positively invariant tubular neighborhood $U \subset W^s(\Gamma)$ of Γ, it is unique.*

Remark III.2. For example, Theorem II.5 of Chapter II implies in particular that $W^s(\Gamma)$ is a tubular neighborhood, and hence we may take $U = W^s(\Gamma)$ in Lemma III.1.

Proof. Let ϕ_1 and ϕ_2 be any two C^1 Temporal 1-Forms.

Since U is a tubular neighborhood, this implies that the limit cycle is a deformation retraction of U[2]. The limit cycle is diffeomorphic to S^1. The first de Rham cohomology of S^1 is isomorphic to $\mathbb{R}$, and de Rham cohomology is a homotopy invariant ([Lee13] Chapter 17). It follows that there exist C^1 functions $V_1, V_2 : U \to \mathbb{R}$, real numbers $c_1, c_2 \in \mathbb{R}$, and a closed 1-form θ on U such that[3]

$$\phi_1 = c_1\theta + \mathbf{d}V_1$$

$$\phi_2 = c_2\theta + \mathbf{d}V_2.$$

Let T be the period of the limit cycle. Since the integrals of the exact 1-forms $\mathbf{d}V_1, \mathbf{d}V_2$ over

[2] In fact, a strong deformation retract. The limit cycle is a strong deformation retract of a tubular neighborhood [Lee13, Ch. 6] containing it. Using the flow, one may construct a strong deformation retract of $W^s(\Gamma)$ onto this tubular neighborhood. Concatenating the two homotopies gives the desired strong deformation retraction.

[3] Strictly speaking, this would be immediate from the standard theory of de Rham cohomology only if $\mathbf{X}$, ϕ_1 and ϕ_2 were C^∞. However, it can be shown that on a C^∞ manifold, any C^r closed form differs from some C^∞ closed form by a C^r exact form ([dR84] p. 68 Theorem 12). Additionally, every C^{r+1} manifold has a unique compatible C^∞ structure ([Hir94] p. 51 Theorem 9). These facts, together with topological invariance of de Rham cohomology, yield the equalities.

closed loops must vanish, we have

$$c_1 \int_\Gamma \theta = \int_\Gamma \phi_1 = T = \int_\Gamma \phi_2 = c_2 \int_\Gamma \theta.$$

Thus the integral of θ over Γ must be nonzero, and we obtain $c_1 = c_2$. It remains only to show that $\mathbf{d}V_1 = \mathbf{d}V_2$.

From the above, it follows that $\phi_1 - \phi_2 = \mathbf{d}(V_1 - V_2)$, and so $V := V_1 - V_2$ satisfies the differential equation $\langle \mathbf{d}V, f(x)\rangle = 0$, and thus every solution trajectory is confined to a level surface of V. In particular $\Gamma \subset V^{-1}(\kappa)$ for some κ. However, all $y \in U$ are contained in trajectories that approach Γ arbitrarily closely. Since V is continuous and U is positively invariant, the value of V on any such trajectory must also be κ. We conclude that $V = \kappa$ on all of U, and thus $\mathbf{d}V_1 = \mathbf{d}V_2$, showing that $\phi_1 = \phi_2$. □

We conclude this section by noting that a choice of basepoint $\gamma(0) \in \Gamma$ uniquely determines the associated phasor-valued phase $\varphi\colon W^s(\Gamma) \to S^1 \subset \mathbb{C}$ via line integration:

$$\varphi(x) = \exp\left(\frac{2\pi i}{T} \int_\sigma \phi\right), \tag{3.1}$$

where σ is any smooth path joining $\gamma(0)$ to x. That this integral is independent of the choice of σ follows from the fact that the limit cycle is a generator for the fundamental group of $W^s(\Gamma)$, the fact that the line integral of ϕ around the limit cycle is the period T, and the fact that line integrals of closed 1-forms are invariant under path homotopy.

3.3.2 The Temporal 1-Form measures the phase response to perturbations

Suppose $\dot{x} = f(x)$ defines a (deterministic) oscillator and let ϕ be its Temporal 1-Form. Let us assume that during a finite time interval $[t_1, t_2]$, the dynamics are perturbed, and are given instead by a model of the form

$$\dot{x} = f(x) + \eta(t, x). \tag{3.2}$$

Here, we do not necessarily assume that the model in (3.2) represents an ODE; rather, we loosely mean that (3.2) is any type of "equation" such that (i) for $\eta(t, x) \equiv 0$, "solutions" of (3.2) coincide with trajectories of the unperturbed oscillator, and (ii) a well-defined notion of line integrals of ϕ along "solutions" of (3.2) exists. For example, $\eta : [t_1, t_2] \times \mathbf{X} \to \mathsf{T}\mathbf{X}$ could simply be a C^1 vector field. Alternatively, η could represent a stochastic process, e.g., *white noise*; in this case, we may interpret (3.2) as a *stochastic differential equation*, along the solutions of which line integrals of differential forms can be defined rigorously [IM79].

For any given perturbation, at times long after t_2 the dynamics of the system coalesce with some trajectory γ taking values in the limit cycle Γ. The phase, or value of $\gamma(0)$ – a distinct trajectory γ exists for every choice of $\gamma(0) \in \Gamma$ – is the only remaining record of the perturbation. This map from perturbation to phase change is the *phase response*, and it plays a special role in oscillator dynamics and determines the long term effects of oscillator coupling [Win80].

A particularly attractive feature of the Temporal 1-Form is the ease with which it allows us to compute the phase response to a perturbation as in (3.2). The line integral of ϕ over the "trajectory" is $\int_{x([t_1,t_2])} \phi = \int_{t_1}^{t_2} \langle \phi, f(x(t)) \rangle \, dt + \int_{t_1}^{t_2} \langle \phi, \eta(t, x(t)) \rangle \, dt$ and thus is $\Delta\phi + t_2 - t_1$ for some value $\Delta\phi$. Since $t_2 - t_1$ is known, the phase response $\Delta\phi$ to the perturbation $\eta(t, x)$ is as easy to compute as the integral itself. By using a computational approach such as we propose in §3.5, an investigator may approximate ϕ for a system from observed trajectories. He or she may then use this to estimate the influence of perturbations on phase and derive oscillator coupling models, despite having no explicit model of the vector field f.

3.4 The Temporal 1-Form estimated from uncertain systems

One of our primary motivations for consideration of the Temporal 1-Form was the hope that the Temporal 1-Form could be computed from real-world data which may be modeled as coming from noisy systems of the form (3.2), and which may additionally be subject to measurement noise. This raises the question: to what extent can the Temporal 1-Form be computed from such real-world data? In this section, we prove some results addressing this question.

First, in §3.4.1 we prove two general theorems relating the true Temporal 1-Form of an underlying deterministic oscillator to an estimate computed by minimizing the specific cost function used in our algorithm. Theorem III.4 relates the quantitative performance of an estimate (from noisy data) of the Temporal 1-Form to that of the true Temporal 1-Form. Theorem III.6 is a qualitative result addressing to what extent the isochrons of the estimated Temporal 1-Form approach the isochrons of the true Temporal 1-Form. The formulation in §3.4.1 is sufficiently general that our results apply for systems subject to system noise of the form (3.2), or to systems subject to measurement noise, or both; see Remark III.3. Next, in §3.4.2 we elaborate in more detail on the applicability of our results to the situation that (3.2) represents a stochastic differential equation (SDE), since SDE data is one of the primary tools we use for testing our algorithm. We consider both Itô and Stratonovich SDEs.

Throughout the remainder of this section, we assume for simplicity that the state space $\mathbf{X}$ is an open subset of $\mathbb{R}^n$.

3.4.1 A general performance bound for both measurement and system noise

For the remainder of §3.4, we will assume that data takes values in a compact tubular neighborhood with smooth boundary $K \subset W^s(\Gamma)$ which contains[4] Γ.

[4] If it does not, we may simply discard the data not in K.

The remainder of this section is outlined as follows. §3.4.1.1 contains the definition of a norm making the space of C^1 1-forms on K into a Banach space, and the definition of a compact subspace $\mathcal{F}$ of such 1-forms. §3.4.1.2 contains our assumptions on the data and noise. §3.4.1.3 contains the main results described above (Theorems III.4 and III.6), which rely on the preliminaries introduced in §3.4.1.1 and 3.4.1.2.

Finally, we remark that the motivation for defining the compact subspace $\mathcal{F}$ introduced in §3.4.1.1 is two-fold. First, this enables the second simpler statement of Theorem III.4. Second — and most importantly — the introduction of $\mathcal{F}$ enables us to prove Theorem III.6 on the convergence of the estimated isochrons to the true isochrons.

3.4.1.1 Definitions and a compact space of 1-forms $\mathcal{F}$

In what follows we let $\|\cdot\|$ be the Euclidean norm, which we use to identify vector fields and 1-forms. By abuse of notation we also denote by $\|\cdot\|$ the norm on covectors induced by this identification.

By a C^m function ω on K, we mean a continuous map ω which is C^m on the interior of K, such that all partial derivatives of order less than or equal to K extend continuously to the boundary of K. Given a d-multilinear map $A\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$, define

$$\|A\| \coloneqq \sup_{\|v_1\|=1,\ldots,\|v_d\|=1} A(v_1, \ldots, v_n).$$

This norm makes the space $\mathcal{L}^d(\mathbb{R}^n, \mathbb{R}^m)$ of all such d-linear maps into a Banach space. Given a continuous map $A\colon K \to \mathcal{L}^d(\mathbb{R}^n, \mathbb{R}^m)$, we define the sup norm $\|A\|_\infty \coloneqq \sup_{x\in K}\|A(x)\|$. Recall that the d-th derivative $\mathsf{D}^d\omega$ of a C^d map ω is naturally identified with an element of $\mathcal{L}^d(\mathbb{R}^n, \mathbb{R}^m)$. Given a C^m map $\omega\colon K \to \mathbb{R}^m$, we define the C^m norm $\|\omega\|_m$ via

$$\|\omega\|_m \coloneqq \max\{\|\omega\|_\infty, \|\mathsf{D}\omega\|_\infty, \ldots, \|\mathsf{D}^m\omega\|_\infty\}.$$

Note that under this definition, we have $\|\omega\|_0 = \|\omega\|_\infty$. We denote by $\Omega^1_m(K)$ the Banach space of all C^m 1-forms on K, topologized with the norm $\|\cdot\|_m$.

We define the following subset $\mathcal{F} \subset \Omega^1_1(K)$ which will serve as the domain of cost functions defined subsequently, over which we perform our optimization.

$$\mathcal{F} := \{\omega \in \Omega^1_1(K) | \forall k : \|\omega\|_1 \leq M_0, \mathrm{Lip}(\mathsf{D}\omega) \leq M_0, \mathbf{d}\omega = 0\}, \tag{3.3}$$

where M_0 is a positive number sufficiently large that $\phi \in \mathcal{F}$. Here $\mathrm{Lip}(\mathsf{D}\omega)$ is defined to be the Lipschitz constant

$$\mathrm{Lip}(\mathsf{D}\omega) \coloneqq \sup_{\substack{x,y \in K \\ x \neq y}} \frac{\|\mathsf{D}\omega(x) - \mathsf{D}\omega(y)\|}{\|x - y\|}$$

By the Arzelà-Ascoli theorem, $\mathcal{F}$ is a compact subset of $\Omega^1_1(K)$ with the topology induced by the C^1 norm, and $\mathcal{F}$ is convex by linearity of the exterior derivative and the triangle inequality. Note that $\|\omega_n - \omega\|_1 \to 0$ implies that all first order partial derivatives of ω_n uniformly approach those of ω, so that $\mathbf{d}\omega = 0$ is indeed a "closed condition".

3.4.1.2 Assumptions about Data and Noise

We assume that we have a finite collection of pairs $(x_i, \dot{x}_i)_{i=1}^N$ with the x_i taking values in $K \subset W^s(\Gamma) \subset \mathbb{R}^n$, and with $\dot{x}_i$ of the form $\dot{x}_i = f(x_i) + \eta_i$. We consider the $x_i \in K$ and $\eta_i \in \mathbb{R}^n$ to be random variables, and we assume that there are constants $\delta, \sigma^2 \geq 0$ such that the means $\mathbb{E}(\eta_i)$ and variances $\mathrm{var}(\eta_i)$ satisfy the bounds

$$\forall i \colon \|\mathbb{E}(\eta_i)\| < \delta \qquad \forall i \colon \mathrm{var}(\eta_i) < \sigma^2. \tag{3.4}$$

Remark III.3. Our assumptions are sufficiently general that, under the mild assumption (3.4), the η_i could arise from measurement noise, or system noise (i.e., of the form (3.2)), or both. In §3.4.2, we will argue in detail that this formulation applies to data from Itô SDEs. It also applies to data coming from Stratonovich SDEs, but there is a slight twist: to get a sharper result, one should replace f by a modified vector field $\tilde{f}$ arising from Itô's formula. See §3.4.2 for details.

Let $\omega \in \Omega_0^1(K)$ be any continous 1-form. The assumptions (3.4), together with the approximate weak law of large numbers (Corollary D.3) imply the following. Here $\mathbb{P}$ denotes the probability measure and $\delta, \sigma^2 \geq 0$ are the constants from above.

$$\forall \epsilon > 0\colon \lim_{N\to\infty} \mathbb{P}\left(\|\omega\|_\infty^{-2} N^{-1} \sum_{i=1}^{N} \langle \omega(x_i), \eta_i \rangle^2 < \epsilon + \sigma^2 + \delta \right) = 1. \tag{3.5}$$

This can be interpreted as a kind of "lim sup in probability" statement. By considering the $\mathbb{P}(\,\cdot\,)$ expressions to be a function of $(n, N) \in \mathbb{N}^2$ and constructing an appropriate "diagonal sequence," we can find a function $\zeta\colon \mathbb{N} \to \mathbb{R}_{\geq 0}$ such that $\zeta(N) \to 0$ as $N \to \infty$ and such that

$$\lim_{N\to\infty} \mathbb{P}\left(\|\omega\|_\infty^{-2} N^{-1} \sum_{i=1}^{N} \langle \omega(x_i), \eta_i \rangle^2 < \zeta(N) + \sigma^2 + \delta \right) = 1. \tag{3.6}$$

Existence of such a function $\zeta(N)$ with a fast decay rate implies that the decay rate of the $\mathbb{P}(\,\cdot\,)$ expression with respect to N slows only slightly as ϵ decreases. The function ζ thereby provides a bound on the convergence rate of the weak law of large numbers for the noise. We will use ζ to inform an estimate of the convergence rate to a performance bound for estimating the Temporal 1-Form from noisy data.

3.4.1.3 Performance of the estimated Temporal 1-Form

Denote by ϕ the unique true Temporal 1-Form on $W^s(\Gamma)$ corresponding to the vector field f on $\mathbf{X}$. We define the "true" cost function $J_N : \Omega_1^1(K) \to \mathbb{R}$ which has ϕ as its (possibly nonunique) global minimizer with minimum $J_N(\phi) = 0$:

$$J_N(\omega) := \frac{1}{N} \sum_{i=1}^{N} \left(\langle \omega(x_i), f(x_i) \rangle - 1 \right)^2 . \tag{3.7}$$

We define the following "noisy" data-driven cost function $\widehat{J_N} : \Omega_1^1(K) \to \mathbb{R}$:

$$\widehat{J_N}(\omega) := \frac{1}{N} \sum_{i=1}^{N} \left(\langle \omega(x_i), f(x_i) + \eta_i \rangle - 1 \right)^2 . \tag{3.8}$$

Note that for any given realization of the (x_i, η_i), $\widehat{J_N} : \mathcal{F} \to \mathbb{R}$ is easily seen to be continuous. We denote by $\widehat{\phi}_N \in \Omega_1^1(K)$ *any* 1-form with the property that

$$\widehat{J_N}(\widehat{\phi}_N) \leq \widehat{J_N}(\phi) \tag{3.9}$$

In a nutshell, our algorithm to compute the Temporal 1-Form works by attempting to minimize $\widehat{J_N}$, although here we have not proven that any minimizers exist (although minimizers of $\widehat{J_N}|_{\mathcal{F}}$ do exist by compactness of $\mathcal{F}$ and continuity of $\widehat{J_N}$). We have the following result giving a bound on the performance of our algorithm.

Theorem III.4. *Assume $\phi \in \Omega_1^1(K)$ and that the preceding assumptions on noise hold. Then*

$$\lim_{N \to \infty} \mathbb{P} \left(J_N(\widehat{\phi}_N) \leq \left(\zeta(N) + \sigma^2 + \delta \right) \left(\|\widehat{\phi}_N\|_\infty + \left[2\|\widehat{\phi}_N\|_\infty^2 + \|\phi\|_\infty^2 \right]^{\frac{1}{2}} \right)^2 \right) = 1.$$

Further assume that $\phi, \widehat{\phi}_N \in \mathcal{F}$. Then

$$\lim_{N \to \infty} \mathbb{P} \left(J_N(\widehat{\phi}_N) \leq \left(\zeta(N) + \sigma^2 + \delta \right) \left(4 + 2\sqrt{3} \right) M_0^2 \right) = 1,$$

where $M_0 > 0$ is as defined in (3.3).

Remark III.5. Since $\zeta(N) \to 0$, Theorem III.4 states that $J(\widehat{\phi}_N)$ will asymptotically be smaller than or equal to

$$\left(\sigma^2 + \delta\right)\left(4 + 2\sqrt{3}\right) M_0^2$$

with high probability. A similar statement holds for Theorem III.4.

Proof of Theorem III.4. Expanding the summand in the definition of $\widehat{J_N}$ and rearranging, we see that for any $\omega \in \mathcal{F}$:

$$\begin{aligned} J_N(\omega) &= \widehat{J_N}(\omega) - \frac{2}{N}\sum_{i=1}^{N}\left(\langle \omega(x_i), f(x_i)\rangle - 1\right)\langle \omega(x_i), \eta_i\rangle \\ &\quad - \frac{1}{N}\sum_{i=1}^{N}\langle \omega(x_i), \eta_i\rangle^2 \qquad (3.10) \\ &\leq \widehat{J_N}(\omega) + 2\left[J_N(\omega)\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{i=1}^{N}\langle \omega(x_i), \eta_i\rangle^2\right]^{\frac{1}{2}} \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}\langle \omega(x_i), \eta_i\rangle^2, \qquad (3.11) \end{aligned}$$

with the inequality arising from an application of the Cauchy-Schwarz inequality to the middle term. It follows immediately from (3.10) that

$$\widehat{J_N}(\phi) = \frac{1}{N}\sum_{i=1}^{N}\langle \phi(x_i), \eta_i\rangle^2 \tag{3.12}$$

since the true Temporal 1-Form ϕ satisfies $(\langle \phi, f\rangle - 1) \equiv 0$. Next, Equations (3.9) and (3.11) imply that

$$\begin{aligned} J_N(\widehat{\phi}_N) \leq \widehat{J_N}(\phi) + \frac{1}{N}\sum_{i=1}^{N}\left\langle \widehat{\phi}_N(x_i), \eta_i\right\rangle^2 + \dots \qquad (3.13) \\ \dots + 2\left[J_N(\widehat{\phi}_N)\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{i=1}^{N}\left\langle \widehat{\phi}_N(x_i), \eta_i\right\rangle^2\right]^{\frac{1}{2}}. \end{aligned}$$

Define

$$\xi(N) \coloneqq \zeta(N) + \sigma^2 + \delta.$$

Substituting (3.12) into (3.13) and using (3.6) yields an inequality quadratic in $\sqrt{J_N(\widehat{\phi}_N)/\xi(N)}$ which holds in probability tending to 1:

$$\frac{J_N(\widehat{\phi}_N)}{\xi(N)} - 2\|\widehat{\phi}_N\|_\infty \left[\frac{J_N(\widehat{\phi}_N)}{\xi(N)}\right]^{\frac{1}{2}} - \left(\|\phi\|_\infty^2 + \|\widehat{\phi}_N\|_\infty^2\right) \leq 0. \tag{3.14}$$

Solving the quadratic inequality (3.14) yields

$$\left[\frac{J_N(\widehat{\phi}_N)}{\xi(N)\|\widehat{\phi}_N\|_\infty^2}\right]^{\frac{1}{2}} \leq 1 + \left[2 + \frac{\|\phi\|_\infty^2}{\|\widehat{\phi}_N\|_\infty^2}\right]^{\frac{1}{2}}, \tag{3.15}$$

and substituting $\xi(N) = \zeta(N) + \sigma^2 + \delta$ yields our result. □

From $J_N(\phi) = 0$ we observe that $J_N(\widehat{\phi}_N) = |J_N(\widehat{\phi}_N) - J_N(\phi)|$, i.e. the estimate provided by Theorem III.4 bounds how well our algorithm does at finding a minimizer of the true goal function. For a fixed dynamical system and fixed σ^2 and δ, the performance of the estimator comes within $\mathcal{O}(\sigma^2 + \delta)$ of the true 1-form asymptotically with increasing sample size, and does so at the same rate as $\zeta(\cdot)$. This seems to be essentially the best one could hope for under the given assumptions.

The appearance of $\|\phi\|_\infty$ in the bound indicates that a "wilder" Temporal 1-Form is harder to estimate. The appearance of $\|\widehat{\phi}_N\|_\infty$ in the right hand side of the bound shows that this is a bound on the relative, rather than absolute, error. However, if we assume that $\widehat{\phi}_N \in \mathcal{F}$ then it is not strictly necessary to interpret the bound as a relative bound, since $\widehat{\phi}_N \in \mathcal{F}$ has bounded sup norm $\|\widehat{\phi}_N\|_\infty \leq M_0$ in this case. We are aware of no other phase estimation method with comparable guarantees on estimation accuracy and convergence rate.

Theorem III.4 is a result on the quantitative performance of the Temporal 1-Form estimated from noisy data. We now consider the case that all of the x_i are independent and identically distributed random variables given by a continuous and nonvanishing probability density on K. For example, this probability density could represent the stationary distribution of a Fokker-Planck equation. We also assume that $\widehat{\phi}_N \in \mathcal{F}$. Under these assumptions and the previous assumptions on noise, the following Theorem III.6 shows that in the limit of large data, the isochrons corresponding to the estimated Temporal 1-Form restricted to K can be made arbitrarily close to the true isochrons in the C^1 topology by taking the noise variance and mean sufficiently small, with high probability.

Theorem III.6. *Assume $\phi, \widehat{\phi}_N \in \mathcal{F}$ and the hypotheses of Theorem III.4. Additionally assume that K is positively invariant for the underlying deterministic dynamical system generated by the vector field f, and also assume that the x_i are independent and identically distributed random variables given by a continuous probability density on K which is nowhere vanishing on* $\operatorname{int}(K)$. *Then for every $\epsilon > 0$, there exists $k_0 > 0$ such that if the noise variance and mean bounds σ^2, δ satisfy $0 \leq (\sigma^2 + \delta) < k_0$, then*

$$\lim_{N\to\infty} \mathbb{P}\left(\|\widehat{\phi}_N - \phi\|_\infty > \epsilon\right) = 0. \tag{3.16}$$

Proof. Let $\rho\colon K \to \mathbb{R}_{>0}$ be the continuous probability density function of the x_i. We define a new cost function $J\colon \mathcal{F} \to \mathbb{R}$ to be the expected value

$$J(\omega) \coloneqq \int_K \left(\langle \omega(x), f(x)\rangle - 1\right)^2 \rho(x)\, dx. \tag{3.17}$$

By Theorem III.1, ϕ is the unique minimizer of J since ρ is continuous and nonvanishing on $\operatorname{int}(K)$: for any other 1-form, the integrand in (3.17) would be positive on an open set, and $J(\phi) = 0$. Since $K \times \mathcal{F}$ is compact and $(x, \omega) \mapsto (\langle \omega(x), f(x) \rangle - 1)^2$ is continuous, there exists a constant $M_1 > 0$ uniformly bounding $(\langle \omega(x), f(x) \rangle - 1)^4$, and therefore

$$\forall i \colon \forall \omega \in \mathcal{F} \colon \operatorname{var}\left[(\langle \omega(x_i), f(x_i) \rangle - 1)^2\right] < M_1.$$

Hence by the explicit convergence rate in Chebyshev's weak law of large numbers (Corollary D.2), we obtain the uniform bound

$$\forall \omega \in \mathcal{F} \colon \forall \epsilon, N > 0 \colon \mathbb{P}(|J_N(\omega) - J(\omega)| > \epsilon) < \frac{M_1}{\epsilon^2 N} \tag{3.18}$$

and therefore

$$\forall \epsilon > 0 \colon \lim_{N \to \infty} \mathbb{P}(|J_N(\widehat{\phi}_N) - J(\widehat{\phi}_N)| > \epsilon) = 0. \tag{3.19}$$

By Theorem III.4, there exists an appropriate constant $K > 0$ such that for every choice of noise mean bound $\delta \geq 0$ and variance bound $\sigma^2 \geq 0$, for every $\kappa > 0$ we have

$$\lim_{N \to \infty} \mathbb{P}\left(J_N(\widehat{\phi}_N) > K(\sigma^2 + \delta) + \kappa\right) = 0. \tag{3.20}$$

Let $B_\epsilon \subset \mathcal{F}$ be the open ball of radius ϵ centered at ϕ. As mentioned, we know that ϕ is the unique minimizer of J and that $J(\phi) = 0$. Since $\mathcal{F}$ and therefore $\mathcal{F} \setminus B_\epsilon$ is compact, J attains a minimum value $\alpha > 0$ on $\mathcal{F} \setminus B_\epsilon$. Hence if $\sigma^2 + \delta$ is sufficiently small that $2K(\sigma^2 + \delta) < \alpha$, then we may choose $\kappa > 0$ sufficiently small so that $2K(\sigma^2 + \delta) + 2\kappa < \alpha$, and therefore

$\forall N \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}\left(\|\widehat{\phi}_N - \phi\|_\infty > \epsilon\right) &\le \mathbb{P}\left(J(\widehat{\phi}_N) > \alpha\right) \\
&\le \mathbb{P}\left((J(\widehat{\phi}_N) > 2K(\sigma^2+\delta) + 2\kappa\right) \\
&\le \mathbb{P}\left(J_N(\widehat{\phi}_N) + |J_N(\widehat{\phi}_N) - J(\widehat{\phi}_N)| > 2K(\sigma^2+\delta) + 2\kappa\right) \\
&\le \mathbb{P}\left(J_N(\widehat{\phi}_N) > K(\sigma^2+\delta) + \kappa\right) \\
&+ \mathbb{P}\left(|J_N(\widehat{\phi}_N) - J(\widehat{\phi}_N)| > K(\sigma^2+\delta) + \kappa\right).
\end{aligned} \tag{3.21}$$

Taking limits as $N \to \infty$ and using (3.19) and (3.20) completes the proof. □

3.4.2 The Temporal 1-Form estimated from stochastic differential equation data

In this section, we consider the collection of data from solutions of a stochastic differential equation (SDE). We consider both Itô and Stratonovich SDEs. We show that the data collected from an Itô SDE satisfies the hypotheses of Theorem III.4, and that the data also satisfies the hypotheses of Theorem III.6 if we assume that the data is drawn from an appropriate stationary distribution of the SDE. We then draw similar conclusions for Stratonovich SDEs, but show that the vector field f should be modified with an Itô correction term in the statements of Theorems III.4 and III.6.

3.4.2.1 Itô stochastic differential equations

Let us suppose that we observe an oscillator perturbed by system noise of the form (3.2), and furthermore that the perturbed system is described precisely by the Itô stochastic differential equation (SDE) [Arn74, Fri75, Gar04, Eva13, Øks13]

$$dX_t = f(X_t)\,dt + \lambda B(t, X_t)\,dW_t, \tag{3.22}$$

where λ is a real number, $t \mapsto B(t,x)$ is nonanticipating with respect to W_t for each fixed x, $f \in C^2$, and such that $f\colon \mathbb{R}^n \to \mathbb{R}^n$ and $B\colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy the following hypotheses of the standard SDE existence and uniqueness theorem [Fri75, Ch. 5, Thm 1.1]:

$$\|f(x) - f(y)\| + \|B(t,x) - B(t,y)\| \leq L_1\|x - y\| \tag{3.23}$$

$$\|f(x)\| + \|B(t,x)\| \leq L_2(1 + \|x\|), \tag{3.24}$$

for some constants $L_1, L_2 > 0$. Here $\|B\| = \mathrm{trace}(B^T B)$.

Let $t_{\max} > 0$ and suppose that we measure (possibly distinct) trajectories of the system at $2N \in \mathbb{N}$ times of the form $t_i \in [0, t_{\max}]$ and $t_i + (\Delta t)_i$, where $(\Delta t)_i > 0$. Assume that at time t_i we measure the state of the system to be $x_i \in K$ (K is the compact neighborhood defined in §3.4.1; if $x_i \notin K$, then discard that measurement), and that at time $t_i + (\Delta t)_i$ we measure the state of the system to be $x_i + (\Delta x)_i$. For each $1 \leq i \leq N$, define $\eta_i \coloneqq \frac{(\Delta x)_i}{(\Delta t)_i} - f(x_i)$.

We have the following result establishing the applicability of Theorems III.4 and III.6 to the data pairs $(x_i, f(x_i) + \eta_i)$.

Theorem III.7. *The data $(x_i, f(x_i)+\eta_i)$ satisfies the hypotheses of Theorem III.4. Furthermore, for every $\delta > 0$ and $\lambda_{\max} > 0$ there exists $(\Delta_t)_{\max} > 0$ such that for all $0 < (\Delta t)_i < (\Delta_t)_{\max}$ and $|\lambda| < \lambda_{\max}$:*

$$\|\mathbb{E}[\eta_i]\| \leq \delta. \tag{3.25}$$

Next, for each fixed $(\Delta t)_{\min} > 0$ such that $(\Delta t)_{\min} < (\Delta t)_i$, and for each fixed $\sigma^2 > 0$, there exists $\lambda_{\max} > 0$ such that for all $|\lambda| \leq \lambda_{\max}$,

$$\mathrm{var}(\eta_i) \leq \sigma^2. \tag{3.26}$$

Finally, additionally assume that K is positively invariant for the deterministic dynamical system with vector field f, and that the Fokker-Planck equation associated to (3.22) has a stationary distribution given by a continuous probability density which is nowhere vanishing on K. If the x_i are drawn from this stationary distribution, then the data $(x_i, f(x_i) + \eta_i)$ satisfy the hypotheses of Theorem III.6.

To prove Theorem III.7 we will use Corollary III.9 of the following Proposition. Proposition III.8 is a special case of the result on uniform mean-square continuity with respect to parameters from [Eva13, Sec. 5.3]. For a proof of a slightly more general result, see [Fri75, Ch. 5, Thm 5.2].

Proposition III.8 (Continuous dependence of SDE solutions on parameters)**.** *For each $k \in \mathbb{N}$, let $f^k\colon [t_0, t_{\max}] \times \mathbb{R}^n \to \mathbb{R}^n$ and $B^k\colon [t_0, t_{\max}] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfy*

$$\|f^k(t,x) - f^k(t,y)\| + \|B^k(t,x) - B^k(t,y)\| \le L_1\|x-y\|$$
$$\forall k \in \mathbb{N}\colon \|f^k(t,x)\| + \|B^k(t,x)\| \le L_2(1+\|x\|),$$

for some constants $L_1, L_2 > 0$ independent of $k \in \mathbb{N}$, $x, y \in \mathbb{R}^n$, and $t \in [t_0, t_{\max}]$. Let X_t^k denote the unique solution of the Itô initial value problem with deterministic initial condition x_0^k:

$$dX_t^k = f^k(t,x)\,dt + B^k(t,x)\,dW_t$$
$$X_0^k = x_0^k.$$

Assume that $x_0^k \to x_0 \in \mathbb{R}^n$ and that for each $M \ge 0$,

$$\lim_{k\to\infty} \max_{\substack{t_0\le t\le t_{\max} \\ \|x\|\le M}} \Big(\|f^k(t,x) - f(t,x)\| + \|B^k(t,x) - B(t,x)\|\Big) = 0.$$

Then

$$\lim_{k\to\infty} \mathbb{E}\left[\max_{t_0\le t\le t_{\max}} \|X_t^k - X_t\|^2\right] = 0, \tag{3.27}$$

where X_t is the unique solution of

$$dX_t = f(t,x)\,dt + B(t,x)\,dW_t$$
$$X_0 = x_0.$$

Corollary III.9. *Assume the hypotheses of Proposition III.8. Then*

$$\lim_{k\to\infty} \max_{t_0\le t\le t_{\max}} \left\|\mathbb{E}\left[X_t^k\right] - \mathbb{E}\left[X_t\right]\right\| = 0$$
$$\lim_{k\to\infty} \max_{t_0\le t\le t_{\max}} |\mathrm{var}(X_t^k) - \mathrm{var}(X_t)| = 0.$$

Proof of Corollary III.9. Since in general the supremum of the expected values of a family of random variables is less than or equal to the expected value of the supremum, we have

$$\begin{aligned}\max_{t_0\le t\le t_{\max}} \mathbb{E}\left[\|X_t^k - X_t\|\right]^2 &\le \max_{t_0\le t\le t_{\max}} \mathbb{E}\left[\|X_t^k - X_t\|^2\right] \\ &\le \mathbb{E}\left[\max_{t_0\le t\le t_{\max}} \|X_t^k - X_t\|^2\right],\end{aligned} \tag{3.28}$$

with the first inequality following from Cauchy-Schwarz. The first inequality together with Proposition III.8 implies convergence of the means

$$\mathbb{E}\left[X_t^k\right] \to \mathbb{E}\left[X_t\right] \tag{3.29}$$

uniformly on $[t_0, t_{\max}]$, because

$$\left\|\mathbb{E}\left[X_t^k\right] - \mathbb{E}\left[X_t\right]\right\| = \left\|\mathbb{E}\left[X_t^k - X_t\right]\right\| \le \mathbb{E}\left[\|X_t^k - X_t\|\right].$$

Next, the second inequality of (3.28), Proposition III.8, and the L^2 triangle inequality imply that

$$\left| \left[\mathbb{E}\left[\|X_t^k\|^2 \right] \right]^{\frac{1}{2}} - \left[\mathbb{E}\left[\|X_t\|^2 \right] \right]^{\frac{1}{2}} \right| \leq \left[\mathbb{E}\left[\|X_t^k - X_t\|^2 \right] \right]^{\frac{1}{2}} \to 0$$

uniformly on $[t_0, t_{\max}]$. Hence $\mathbb{E}\left[\|X_t^k\|^2 \right] \to \mathbb{E}\left[\|X_t\|^2 \right]$ uniformly on $[t_0, t_{\max}]$. Together with (3.29), this implies that

$$\text{var}(X_t^k) \to \text{var}(X_t) \tag{3.30}$$

uniformly on $[t_0, t_{\max}]$.

□

Proof of Theorem III.7. Throughout the proof, we use the notation η_i^λ to highlight the dependence of η_i on λ. We let $X_t^{\lambda,i}$ denote the solution of (3.22), defined for $t \geq t_i$, with initial condition $X_{t_i}^{\lambda,i} = x_i$.

We first establish (3.25). We have

$$\eta_i^\lambda = -f(x_i) + \frac{1}{(\Delta t)_i} \int_{t_i}^{t_i + (\Delta t)_i} f(X_s^{\lambda,i})\, ds + \frac{1}{(\Delta t)_i} \int_0^t B(t, (X_s^{\lambda,i}))\, dW_s. \tag{3.31}$$

Since B is nonanticipating, Itô calculus rules imply that the expected value of the second integral vanishes [Gar04, Sec. 4.2.6.e], and therefore

$$\mathbb{E}[\eta_i^\lambda] = R(x_i, \lambda, t_i, (\Delta t)_i) \coloneqq -f(x_i) + \frac{1}{(\Delta t)_i} \int_{t_i}^{t_i + (\Delta t)_i} \mathbb{E}\left[f(X_s^{\lambda,i}) \right] ds. \tag{3.32}$$

We estimate

$$\|R(x_i, t_i, \lambda, (\Delta t)_i)\| \leq \frac{1}{(\Delta t)_i} \int_{t_i}^{t_i + (\Delta t)_i} \|\mathbb{E}\left[f(X_s^{\lambda,i}) \right] - f(x_i)\|\, ds. \tag{3.33}$$

The sample path $s \mapsto X_s^{\lambda,i}$ is continuous with probability one [Gar04], and therefore by the fundamental theorem of calculus we see for each fixed (t_i, x_i, λ) that the right hand side of (3.33) tends to 0 as $(\Delta t)_i \to 0$ with probability one. Additionally, nonnegativity of the integrand implies that this convergence is monotone. Finally, Proposition III.8 and the Lipschitz condition (3.23) imply that the right hand side of (3.33) is a continuous function of (t_i, x_i, λ) for each $(\Delta t)_i > 0$. Hence Dini's theorem [Rud76, Thm 7.13] applies to show that $R \to 0$ uniformly on the compact set $K \times [0, \lambda_{\max}] \times [t_i, t_{\max}]$ as $(\Delta t)_i \to 0$. Therefore, $\delta > 0$ in (3.25) can be made as small as desired by taking $(\Delta t)_{\max}$ sufficiently small.

Next, we establish (3.26). Since $X_t^{0,i}$ is the deterministic solution of an ODE, we have $\mathrm{var}(X_t^{0,i}) \equiv 0$. Since

$$\mathrm{var}(\eta_i^{\lambda}) = \frac{1}{((\Delta t)_i)^2}\mathrm{var}(X_t^{\lambda,i}),$$

uniform convergence of variances (Corollary III.9) implies that for each value of $(\Delta t)_{\min} > 0$, $\mathrm{var}(\eta_i)$ can be made arbitrarily small by taking $|\lambda|$ sufficiently small.

To complete the proof, note that the final statement merely adds the additional hypotheses needed for Theorem III.6. □

3.4.2.2 Stratonovich stochastic differential equations

Let us again suppose that we observe an oscillator perturbed by system noise which is described by a stochastic differential equation, but now we consider an autonomous Stratonovich (rather than Itô) SDE

$$dX_t = f(X_t)\,dt + \lambda B(X_t) \circ dW_t. \tag{3.34}$$

Here $\circ$ is not composition, but merely a somewhat standard choice of notation indicating that this is a Stratonovich equation. We assume that f and B satisfy all of the same hypotheses as in §3.4.2.1, and we assume that we have data x_i, $x_i + (\Delta x)_i$ collected at times and t_i, $t_i + (\Delta t)_i$ exactly as in §3.4.2.1.

We cannot simply repeat the proof of Theorem III.7, because to obtain Equation (3.32) we used the fact that the expected value of the Itô integral of a nonanticipating function is zero. This is not true of the expected value of a Stratonovich integral [Eva13, Sec. 6.5.7]. However, we can circumvent this issue by converting (3.34) into an equivalent Itô SDE:

$$dX_t = \left(f(X_t) + \frac{\lambda^2}{2}c(X_t)\right) dt + \lambda B(t, X_t)dW_t, \tag{3.35}$$

where the i-th component $c^i(x)$ is given by

$$c^i(t,x) = \sum_{k=1}^{m}\sum_{j=1}^{n}\left(\frac{\partial}{\partial x^j}B^i_k\right)B^j_k. \tag{3.36}$$

I.e., X_t is a solution to the Stratonovich SDE (3.34) if and only if X_t is a solution to the Itô SDE (3.35) [Eva13, Sec. 6.5.6].

We define a new autonomous vector field (this is why we assumed that we started with an autonomous Stratonovich SDE; otherwise, c would also depend on t)

$$\tilde{f}(x) \coloneqq f(x) + \frac{\lambda^2}{2}c(x) \tag{3.37}$$

and new noise terms

$$\tilde{\eta}_i \coloneqq \frac{(\Delta x)_i}{(\Delta t)_i} - \tilde{f}(x_i). \tag{3.38}$$

From Theorem III.7, we immediately obtain the following.

Theorem III.10. *The data $(x_i, \tilde{f}(x_i) + \tilde{\eta}_i)$ satisfies the hypotheses of Theorem III.4. Furthermore, for every $\delta > 0$ and $\lambda_{\max} > 0$ there exists $(\Delta_t)_{\max} > 0$ such that for all $0 < (\Delta t)_i < (\Delta_t)_{\max}$ and $|\lambda| < \lambda_{\max}$:*

$$\|\mathbb{E}[\tilde{\eta}_i]\| \leq \delta. \tag{3.39}$$

Next, for each fixed $(\Delta t)_{\min} > 0$ such that $(\Delta t)_{\min} < (\Delta t)_i$, and for each fixed $\sigma^2 > 0$, there exists $\lambda_{\max} > 0$ such that for all $|\lambda| \leq \lambda_{\max}$,

$$\mathrm{var}(\tilde{\eta}_i) \leq \sigma^2. \tag{3.40}$$

Finally, additionally assume that K is positively invariant for the deterministic dynamical system with vector field $\tilde{f}$, and that the Fokker-Planck equation associated to (3.35) has a stationary distribution given by a continuous probability density which is nowhere vanishing on K. If the x_i are drawn from this stationary distribution, then the data $(x_i, \tilde{f}(x_i) + \tilde{\eta}_i)$ satisfy the hypotheses of Theorem III.6.

Remark III.11. In principle, we could still analyze the performance of our algorithm on data from (3.34) using f and η_i in Theorem III.4 rather than $\tilde{f}$ and $\tilde{\eta}_i$. However, it would then no longer be possible to make δ in (3.39) arbitrarily small by merely taking $(\Delta t)_{\max}$ small, although in this case δ can still be made small by taking both $(\Delta t)_{\max}$ and $\lambda_{\max}$ sufficiently small.

3.5 Algorithm

For systems in which the dimension is low, the noise is small or the data is plentiful, data pairs $(x_k, \dot{x}_k)$, $k = 1 \ldots N$ may be used to estimate the vector field f. Numerous investigators proposed such approaches, primarily in the 1990s. Classical papers include Kostelich and Yorke [KY90] who fit arbitrary continuous dynamics to trajectories, and the rich literature on time-delay embeddings, derivatives and principal components [SWT06, GFCE92]. When dynamical equations are known precisely, methods based on forward integration [Win80, MM12] or on automatic differentiation and continuation [OM10, HdlL12, LKO14] can be used to compute the isochrons or equivalently asymptotic phase.

With fewer data points and higher levels of noise, one can no longer expect the off-cycle dynamics to be entirely amenable to accurate estimation, and only a few of the slower decaying modes might be observable off of the limit cycle [RG12]. At the extreme, only phase itself might be detectable, through methods such as those proposed in [RG08], or through various phase-locking approaches [PS08, SA10, SPNI05].

Several investigators have pursued the construction of linearized (Floquet) models for dynamics near the limit cycle. An approach of Data Driven Floquet Analysis (DDFA) was proposed in [RGF11, Rev09] which consisted of the estimation of phase followed by the construction of affine models conditioned on phase. Tytell [Tyt13] reported a DDFA approach based on harmonic balance. Wang and Srinivasan [WS12] proposed to construct a Floquet model using "factored Poincaré maps", and then derive a phase estimate from this model. Ankarali *et. al* [AC14] applied "harmonic transfer functions" and achieved good results on a hybrid spring-mass hopper model. Regardless of how they are obtained, linear models of the dynamics around the limit cycle can at best only represent the hyperplanes tangent to the isochrons at their intersection with the limit cycle. By contrast, the method we propose here can construct nonlinear isochron approximations wherever in state space flow data convergent to the limit cycle is available.

Finally, we remark that methods based on forward integration [Win80, MM12] can in principle be applied directly to data without first attempting to estimate the vector field f. However, such methods necessarily require collections of time series with lengths exceeding several periods of oscillation. By contrast, *the method we propose here can use collections of time series having arbitrarily small lengths.* This is arguably the most valuable property of our method. We additionally have explicit theoretical guarantees on the performance of our method applied to noisy data (§3.4); we know of no comparable guarantees for methods based on forward integration.

3.5.1 Approximation by topologically motivated basis functions

We construct a series approximation to the Temporal 1-Form using the basis[5] functions described below.

A naive approach would be to attempt using localized vector-valued basis functions, such as the exterior derivatives of radial basis functions. However, the fact that the line integral of the Temporal 1-Form along the limit cycle is non-zero implies that it is not exact, i.e., it is not the exterior derivative of any scalar valued function. Because linear combinations of exact forms are exact, the Temporal 1-Form cannot be expressed as a sum of exact 1-forms. The key insight enabling our algorithm is that although the closed 1-form we seek is not exact, the first de Rham cohomology group of the stability basin $W^s(\Gamma)$ or any other tubular neighborhood of the limit cycle Γ is isomorphic to $\mathbb{R}$, meaning that all closed 1-forms fall into equivalence classes, each class corresponding to a real number. Given an exemplar 1-form in some non-zero class, its scalar multiples generate the entire de Rham cohomology, and any two closed 1-forms in the same cohomology class differ by an exact form[6]. Therefore, if we provide one basis function with non-zero cohomology and correctly identify its multiplier, the residual can be approximated by an $\mathbb{R}$-linear combination of exact basis functions.

We approximate the Temporal 1-Form ϕ using a two-step process described below. First we construct a closed form $\mathbf{d}\theta$, which[7] we refer to as the θ form, which lies in the same cohomology class as ϕ. The θ term captures the nontrivial cohomology of the Temporal 1-Form. As such it cannot be exact, and must have a nonzero line integral around the limit cycle. Second, we approximate the residual exact form $(\phi - \mathbf{d}\theta)$ by a series using basis

[5] We use the term "basis" informally, as it is used e.g. in the notion of "radial basis functions" in machine learning, to mean a collection of functions whose finite linear combinations are dense in the function space of interest.

[6] Note that any C^1 closed form on a C^∞ manifold differs from some C^∞ closed form by a C^1 exact form [dR84, pp. 61-70], so our statements apply to C^1 forms.

[7] Despite the unfortunate notation, $\mathbf{d}\theta$ is closed but not exact.

functions of Taylor-Fourier type.

Remark III.12. The specific algorithm described below is just one possible method for computing the Temporal 1-Form. It is certainly possible to compute a different θ form having the same cohomology as ϕ via other means, and it is certainly possible to fit the residual exact form using a different set of basis functions. Theoretically such algorithms are all capable of computing the Temporal 1-Form, but in practice different choices may impact performance for numerical reasons. Indeed, we are currently testing a variant of the algorithm described below, to appear in a future publication.

3.5.1.1 Fitting the θ form

In typical real-world applications, data takes the form of a sequence of n-tuples of real numbers (i.e., a time series). In these cases, the state space $\mathbf{X}$ is an open subset of $\mathbb{R}^n$, and so in this and all following sections we assume that $\mathbf{X}$ is an open subset of $\mathbb{R}^n$. We will take full advantage of the vector space structure of $\mathbb{R}^n$ in our algorithm to compute the Temporal 1-Form described below. However, it should in principle be possible to adapt the computation of the Temporal 1-Form to other cases in which $\mathbf{X}$ is a non-Euclidean smooth manifold.

Noting that in all dimensions above 3 the limit cycle cannot be knotted (see, e.g., the corollary to Theorem 1 in [Hae61]), there is a choice of smooth isotopy that "flattens" the limit cycle to a planar, unknotted circle. The circulation 2-form of such a plane spanned by the orthonormal pair (v_1, v_2) is $v_1^\flat \wedge v_2^\flat$, where $v_1^\flat, v_2^\flat$ are the duals of v_1, v_2 via the Euclidean inner product (i.e., $v_i^\flat(\cdot) = \langle v_i, \cdot \rangle$). This allows us to define $\mathbf{d}\theta$ at any point $x \in \mathbf{X}$ by pulling back $v_1^\flat \wedge v_2^\flat$, suitably normalized, via the isotopy and contracting this pulled-back form with x minus a suitably chosen center (identifying this difference with a tangent vector via the Euclidean structure).

Since the θ form is of non-trivial cohomology and the first de Rham cohomology of $\mathbf{X}$

is one-dimensional, we can be sure that the Temporal 1-Form is in the span of the θ form's cohomology class. If the magnitude of the θ form is chosen correctly, the residual will be of trivial cohomology and can (in principle) be approximated by a suitable basis of exact forms.

In most practical applications we have encountered in the past, the first two principal components v_1 and v_2 of the data in phase space defined a 2-plane within which the data was restricted to an annulus with a pronounced average flow around it. We thus forgo the preceding generality by taking the isotopy to be given by orthogonal projection. This allows rapid computation of $\mathbf{d}\theta(x)$ using two pre-computed matrices $S := v_1^\flat \otimes v_2^\flat - v_2^\flat \otimes v_1^\flat$ and $\Pi := v_1 \otimes v_1^\flat + v_2 \otimes v_2^\flat$, giving

$$\mathbf{d}\theta(x) = \frac{1}{C} \frac{\langle S, x - x_0 \rangle}{\|\Pi\,(x - x_0)\|^2}, \tag{3.41}$$

where $C \in \mathbb{R}$ and $\overline{x} \in \mathbb{R}^n$ are to be determined, and $\langle S, x \rangle$ is tensor contraction on the first index of S.

As a result of this choice the θ form has three parameters: the constant C, and $v_1^\flat(\overline{x}), v_2^\flat \overline{x})$ representing the location of the singularity of the θ form in the first two coordinates. The latter two parameters can be obtained heuristically by simply setting $\overline{x}$ equal to the arithmetic mean of points in the data set. We can approximate C by computing a Fourier series approximation of the limit cycle via standard numerical methods, temporarily setting $C = 1$ in (3.41), integrating the resulting 1-form over the Fourier approximation of the limit cycle, and defining C to be the value of this integral.

3.5.1.2 Fitting the residual exact form

Let $v_1, \ldots, v_n$ and $\overline{x}$ be the principal components and arithmetic mean of the data set, respectively. Let $y_1, \ldots, y_n$ denote coordinates of the data with respect to the v_i, after been translated by the mean. Explicitly, $y_i \coloneqq v_i^\flat(x - \overline{x})$. Let (r, α) be polar coordinates on the plane spanned by v_1 and v_2, so that $y_1 = r\cos(\alpha)$ and $y_2 = r\sin(\alpha)$. Assuming that orthogonal projection onto $\mathrm{span}\{v_1, v_2\}$ restricted to the limit cycle is injective, the limit cycle may be parametrized by α. Hence we may write our Fourier series approximation of the limit cycle as $(\gamma_r(\alpha), \gamma_3(\alpha), \ldots, \gamma_n(\alpha))$. This enables us to define a system of coordinates $(\alpha, z_2, \ldots, z_n)$ on a tubular neighborhood of the limit cycle by

$$\begin{aligned} z_2 &\coloneqq r - \gamma_r(\alpha) \\ \forall i \geq 3\colon\ z_3 &\coloneqq y_i - \gamma_i(\alpha). \end{aligned} \tag{3.42}$$

Finally, we define $z \coloneqq (z_2, \ldots z_n)$. Here the z_i are a system of coordinates transverse to the limit cycle, such that the limit cycle corresponds to $\{z = 0\}$.

Since the constant C of (3.41) was chosen so that $\mathbf{d}\theta$ lies in the same cohomology class as ϕ, $\phi - \mathbf{d}\theta = \mathbf{d}\tilde{V}$ for some C^1 function $\tilde{V}$. We define a function V approximating $\tilde{V}$ using a Taylor-Fourier series about the limit cycle, using our new coordinates:

$$V(\alpha, z) \coloneqq \sum_j \sum_k e^{ik\alpha} A_{jk} \cdot z^{\otimes j}. \tag{3.43}$$

Here A_{jk} is a symmetric j-linear map to be determined and $\otimes$ is the tensor product. The exterior derivative $\mathbf{d}V$ can be computed in closed form:

$$\mathbf{d}V(\alpha, z) = \sum_j \sum_k \left[\left(ike^{ik\alpha} A_{jk} \cdot z^{\otimes j}\right) \mathbf{d}\alpha + jA_{jk} \cdot z^{\otimes(j-1)} \otimes \mathbf{d}z\right]. \tag{3.44}$$

Here we use the shorthand notation $\mathbf{d}z$ to mean the vector-valued 1-form whose contraction with any vector w is given by $(\mathbf{d}z_2(w), \ldots, \mathbf{d}z_n(w))$.

Next, writing our data set explicitly as a set of pairs $(x_\ell, \dot{x}_\ell)_{\ell=1}^N$, we define residuals

$$\epsilon_\ell \coloneqq 1 - \langle \mathbf{d}\theta(x_\ell), \dot{x}_\ell \rangle \,. \tag{3.45}$$

Hence for all $1 \leq \ell \leq N$ we want $\langle \mathbf{d}V, \dot{x}_\ell \rangle = \epsilon_\ell$, or

$$\sum_j \sum_k \left[\left(ike^{ik\alpha_\ell} A_{jk} \cdot z_\ell^{\otimes j} \right) \mathbf{d}\alpha(\dot{x}_\ell) + jA_{jk} \cdot z_\ell^{\otimes(j-1)} \otimes \mathbf{d}z(\dot{x}_\ell) \right] = \epsilon_\ell, \tag{3.46}$$

where (α_ℓ, z_ℓ) is the representative of x_ℓ in our new coordinates. Equation (3.46) is linear in the A_{jk} terms, and therefore we choose to view (3.46) as an ordinary least squares problem for these terms, which we obtain via a standard numerical algorithm (SciPy implementation `lstsq`). Defining our estimate $\widehat{\phi} \coloneqq \mathbf{d}\theta + \mathbf{d}V$ completes the description of our algorithm to estimate ϕ. Hereafter, we refer to this estimate of phase as *form phase*.

An example illustration of the contributions to $\widehat{\phi}$ of the terms in our series expansion of $\mathbf{d}V$ is shown in Figure 3.1.

3.6 Empirical performance of the new algorithm

We produced three noisy simulation systems of dimensions two, three and eight[8]. The deterministic limit of all systems have known asymptotic phase. The first is easy to visualize; the second is the minimal dimension in which general complex eigenvalues can appear in the Floquet structure; the third is more typical of the dimensionality of biological systems for which we developed this estimation method. We examined the performance of three different estimates of the phase of these systems and compared the estimated phase to the analytically known asymptotic phase.

[8]Specifically, we generated data by numerically integrating Stratonovich SDEs of the form (3.34) representing a perturbation of an underlying deterministic oscillator.

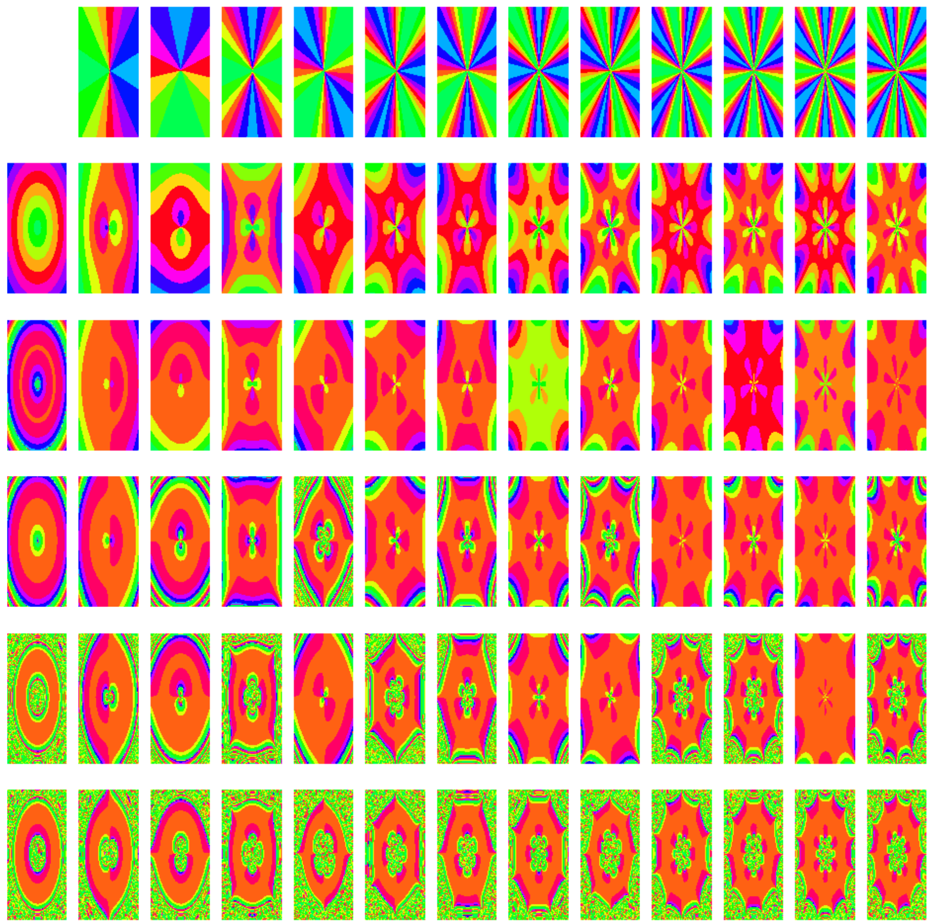


Figure 3.1: Shown is a plot of the contribution of each term in our form phase estimate for an example planar system having the unit circle as a limit cycle. We show the contributions for polynomial and Fourier terms up to sixth order. From left to right the order of the Fourier term increases and the cosine and sine terms alternate, with the first column containing the constant term. The order of the transverse polynomial term increases from top to bottom. The top left square is blank, because the constant term in the expansion of V vanishes after taking the exterior derivative, and therefore does not contribute to the phase estimate.

The first phase estimator we used is comparable to "event-based" phase estimates commonly used in biological research. Researchers often use distinguished events, such as voltage levels in the nervous system [DHF+15], footfall [TBF94, JF02], or anterior extreme position of a limb [CS88] to identify the beginning of a cycle and presume that phase evolves uniformly in time between these events. As our event, we used the zero crossing of the first principal component of the data. The second phase estimator we used was the previously published *Phaser* algorithm [RG08]. The third phase estimator, described previously in this chapter, was derived from the Temporal 1-Form and we henceforth refer to this new estimate as *form phase*.

We trained our phase estimation methods on one set of noisy simulated paths and tested them on another, generated from the same underlying equations. We calculated a residual by subtracting the estimated phases from the ground truth phases for each method. We removed the globally indeterminate phase difference by taking the circular mean of this residual to zero by subtracting a constant.

For each simulation we considered three different conditions corresponding to different levels of initial condition variability and system noise. The ratios of initial condition variance to system noise variance for the three conditions were $1 : 9 : 25$. The simulations along with the code used to generate them are available at `http://github.com/BIRDSlab/temporal1form`. The results show that the form phase method consistently has lower mean-square residual than the *Phaser* method, which in turn has lower mean-square residual than the event-based method; see residuals in Table 3.1 and Figure 3.2.

3.6.1 Isochrons and phase response curves

For the two-dimensional system of the previous section, we estimated the isochrons in the case of low system noise and highest initial condition noise. We plotted points in the plane in 20 different phase intervals and compared them to the ground truth phase. The

Table 3.1: Variance of residual phase for difference phase estimation techniques with different initial condition noise and system noise levels.

	noise			res. var.		
D	initial	system	phase	event	*Phaser*	form
2	0.1	0.1	0.1	0.110	0.0469	0.0185
2	0.2	0.1	0.1	0.107	0.0461	0.0182
2	0.1	0.2	0.1	0.113	0.0467	0.0173
2	0.1	0.1	0.2	0.214	0.0881	0.0397
3	0.066	0.0066	0.066	0.0733	0.130	0.0356
3	0.133	0.0066	0.066	0.0768	0.141	0.0174
3	0.066	0.0133	0.066	0.0873	0.132	0.0483
3	0.066	0.0066	0.133	0.143	0.104	0.0253
8	0.025	0.0025	0.025	0.0446	0.0327	0.0258
8	0.05	0.0025	0.025	0.0451	0.0344	0.0290
8	0.025	0.005	0.025	0.0683	0.0540	0.0486
8	0.025	0.0025	0.05	0.0646	0.0344	0.0264

results are depicted in Figure 3.3. We found that event-based estimates of the isochrons are poor, with a large spread in position and poor agreement with the ground truth. *Phaser* performed well on the limit cycle, but the angle of the limit cycle to the isochrons was incorrect. The form phase estimate matched the true isochrons far more closely. Since the angle of the isochrons to the limit cycle corresponds to the *infinitesimal Phase Response Curve (iPRC)* [Erm96], this last observation suggests that in particular, the form phase estimates can provide superior estimates of the iPRC, which is often used in the study of neuronal coupling.

Formally, the iPRC is the Lie derivative of an angle-valued representation of asymptotic phase (see §3.3.1) along some vector field in state space (typically one of the (constant) coordinate vector fields), subsequently restricted to the limit cycle. Hence the iPRC is equivalently given by the contraction of the Temporal 1-Form with such a vector field on the limit cycle, so if the vector field is a (constant) coordinate vector field, then the iPRC is simply given by the corresponding component of the Temporal 1-Form as a function of points on the limit cycle. We estimated the iPRC (iPRC) of a two dimensional system to

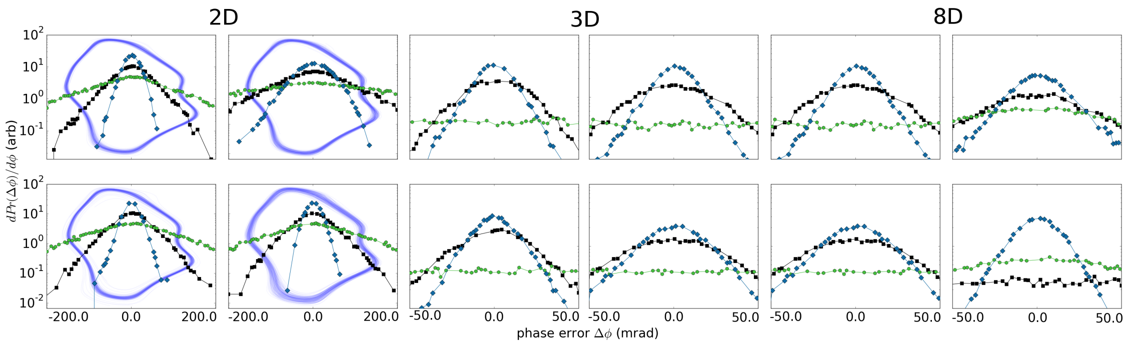


Figure 3.2: Comparison of phase estimators on 2D, 3D, and 8D systems. Comparison of event phase (green) *Phaser* (black) and form phase (turquoise) phase noise distributions on simulated trajectories. Plots show four different conditions for each dimension: one baseline condition (top left), an increase in the magnitude of the SDE diffusion term (top right), an increase in the levels of variability in the initial conditions (bottom left) and an increase in noise on the coordinate corresponding to phase in the equivalent deterministic system (bottom right; see Table 3.1 for details).

perturbation in the first coordinate. The result is depicted in Figure 3.4.

3.6.2 Phase estimation from partial data

Because our algorithm requires only state and state velocity pairs to calculate a phase estimate, it can be trained using short, disjointed training examples – even if no individual example contains more than a small portion of a cycle. This is not true of any event-based phase, since the phase of all segments that do not contain the event cannot be determined at all. Similarly, *Phaser* cannot determine the phase of any data coming from a segment that does not cross the zero phase Poincaré section [RG08]. Furthermore, because it uses the Hilbert transform to produce a protophase, *Phaser* requires each individual time series in its training data to be multiple cycles long – otherwise the Hilbert transform exhibits ringing artifacts.

However, to recover the isochrons our algorithm does effectively require *in theory* that the union of training data from all individual trials is "dense" in some positively invariant tubular neighborhood of the limit cycle. A sufficient formal condition is to assume that

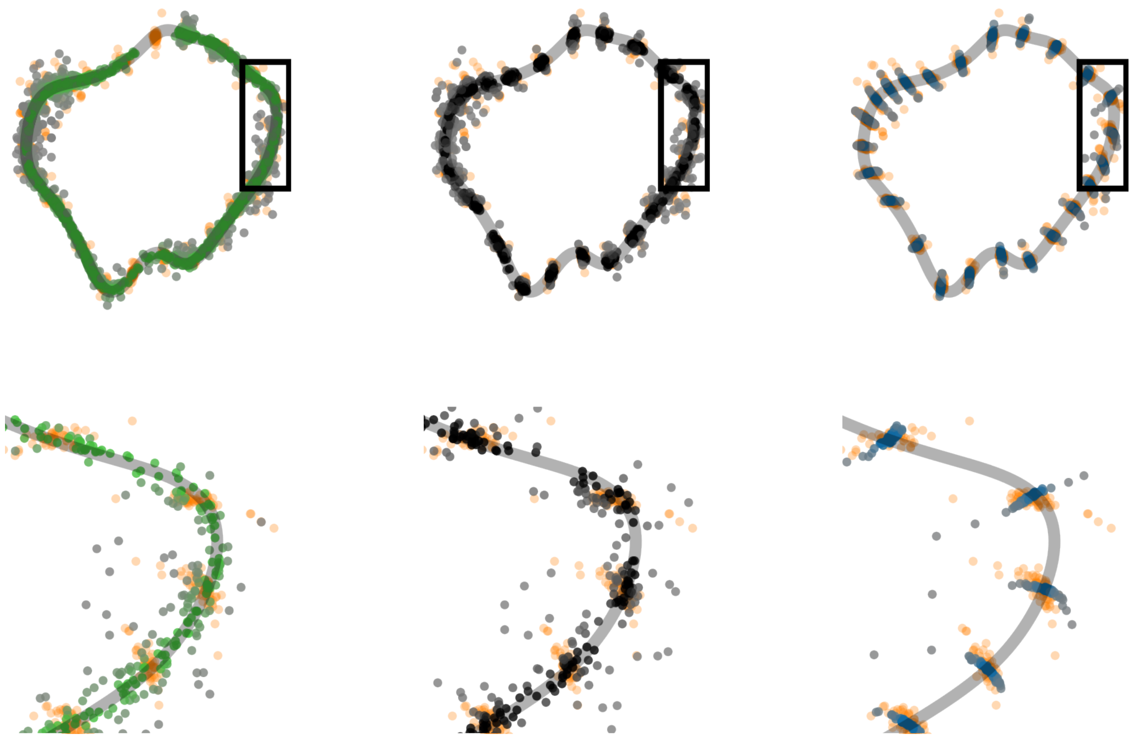


Figure 3.3: Comparison of the isochrons obtained from the form-based, *Phaser* and event-based phase estimates. Plots show the limit cycle (gray curve) and scatter plots of trajectory points falling in 50 equally spaced intervals of width 0.02 periods in “ground-truth” phase (orange). Scatter plots also show trajectory points falling into the same intervals from event ([A], green), *Phaser*([B], black), and form-based phase estimates ([C] , gray), with corresponding boxes zoomed in in each sub-figure.

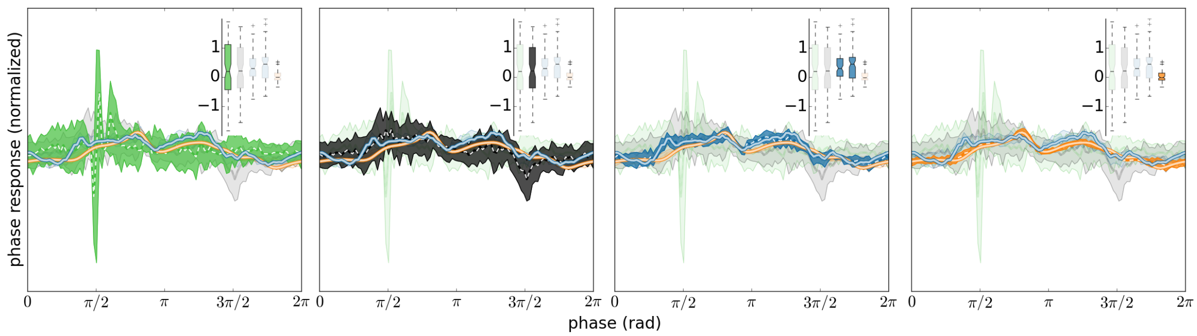


Figure 3.4: Normalized phase response curves to variation in the x_1 coordinate using four different methods. For phase derived from each estimation method, we estimated the PRC by dividing the limit cycle into 100 locations, taking the 5000 points closest to each location, and taking from the 5000 points the 500 samples whose y coordinate was closest to that of the location on the cycle. We then calculated the exterior derivative of the phase with respect to this coordinate by least squares regression. Additionally, we obtained the phase response curve on the limit cycle directly from the form-phase by taking the y-component of the Temporal 1-Form at each location. All methods are shown in each plot, highlighting a different method in each one (left to right): PRC from event phase (green); *Phaser* (black); Form phase (blue); ground truth (orange). We calculated the PRCs and their confidence intervals by bootstrapping a linear fit to points in restricted phase ranges. In addition, we show a PRC calculated analytically from the deterministic system (solid orange line) and the x_1-component of the form phase (solid blue line). We show the distribution of residuals with respect to ground truth (inset box-plots). Two box plots are plotted for form phase: one from bootstrapping regression as in the other estimation methods (left), and one from the x_1-component of the form (right).

the data is drawn from an appropriate continuous probability density so that Theorem III.6 applies. This can be seen as a manifestation of Theorem III.1, which asserts uniqueness of the Temporal 1-Form on any positively invariant tubular neighborhood of the limit cycle. On the other hand, it is easy to find examples of a neighborhood of a proper subset of the limit cycle on which there exist infinitely many invariant foliations and therefore infinitely many Temporal 1-Forms. Because of this, we heuristically expect that we should not be able to compute the Temporal 1-Form or isochrons using a data set which does not fill a tubular neighborhood of the cycle. However, *in practice* we have often found that somehow isochrons can reasonably recovered using only data lying in such partial tubular neighborhoods! An example of this surprising finding is illustrated in using guineafowl data in Figure 3.5, and an explanation remains a topic for future work.

3.7 Discussion

We have introduced the Temporal 1-Form and shown that, for smooth exponentially stable oscillators, it agrees with the classical definition of asymptotic phase. The properties of the Temporal 1-Form makes evident the fact that when asymptotic phase exists (a decidedly global requirement), the Temporal 1-Form is nevertheless uniquely determined by purely local conditions.

These theoretical properties allowed us to develop an algorithm for approximating the Temporal 1-Form of an oscillator from which noisy trajectory segments are available. We have shown that this algorithm performs well on both simulated and experimental data. The computation of the Temporal 1-Form does not require the dynamics themselves to be integrated.

Our algorithms offers two significant improvements over previously known methods for phase estimation.

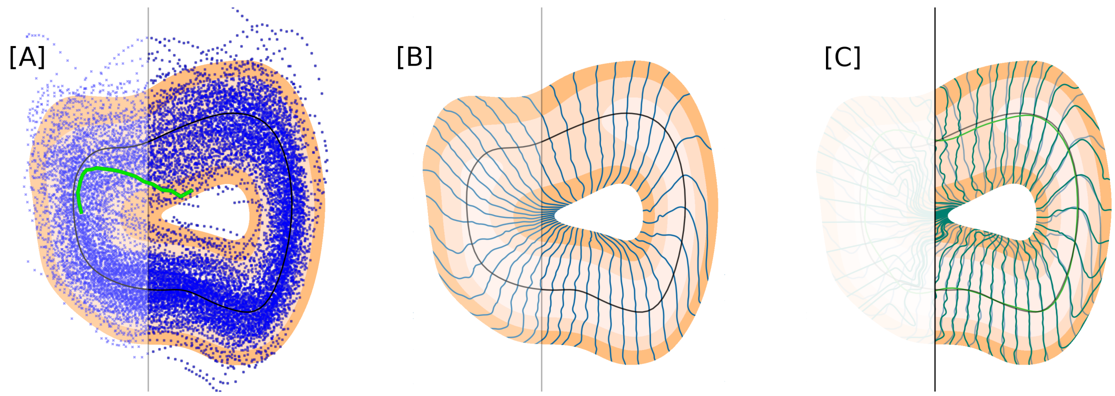


Figure 3.5: A mysterious result: estimating phase from partial data (c.f. §3.6.2). We applied our technique to 292 cycles of foot positions taken at 250Hz from *Numida meleagris* gineafowl running at 3Hz. We split trials into 20 sample long segments with gaps of at least 40 samples in between. Each segment was far shorter than a cycle ([A] blue dots all data; example segment wide solid green). We differentiated each segment of position data to obtain velocity data using finite differences. We used the form phase estimator to estimate phase for these data, and plotted a Fourier series approximation of the limit cycle ([B], solid black) and isochrons ([B] light teal, every $\pi/20$ radians in phase). We then removed all data left of the line $x = -0.5$ (the x coordinate is in arbitrary units since the data has been z-scored), eliding about one third of every cycle ([A,B,C] vertical line; elided data faded to its left). We recomputed phase, plotted the isochrons at the same points as before ([C] dark teal; full data isochrons in faded light teal) and showed the updated Fourier approximation of the limit cycle (solid green). We indicated the amount of available training data in each area using contours of a kernel smoothed density plot of the training data points (50%, 70% and 90% of max density, going from strong orange to pale orange respectively). Even with fractional cycles and systematic inability to observe a large fraction of the cycle, form phase recovers isochrons closely resembling those using the full data set in [B]. As discussed in §3.6.2, this result is surprising, and an explanation remains a topic for future work.

1. Our algorithm has the ability to estimate phase from state and state velocity pairs $(x, \dot{x})$. Therefore, our algorithm can use collections of time series having arbitrarily small lengths, as long as the union of data from all time series fills an appropriate tubular neighborhood of the cycle (Theorem III.6). In particular, this capability makes our algorithm attractive for modeling systems for which we are unable to obtain records of full periods of oscillation.

2. Our algorithm has the ability to estimate phase far from the limit cycle, in any region from which trajectories convergent to the limit cycle could be constructed from observed flow data. This allows for the study of oscillatory systems undergoing large perturbations or undergoing a qualitative change in dynamics (e.g. gait transitions in animals).

We are not aware of any other data-driven phase estimation method which possesses either of these important properties.

The Temporal 1-Form allows one to detect the phase response to perturbations of the underlying deterministic dynamics. The phase response describes the only effect of a perturbation that persists after a long time, and even the infinitesimal phase response approximation – the infinitesimal phase response curve (iPRC) – allows the observed system to be modeled in the weakly coupled oscillator approximation. This iPRC is available directly from the Temporal 1-Form phase estimation procedure.

Experimental data frequently contains partial cycles due to missed observations. We have shown that our form phase estimator can be successfully trained on short fragments of cycles, as long as the union of data from all trials fills a suitable neighborhood of the limit cycle. Event-based methods such as anterior extreme position detection [CS88] or heel-strike [TBF94, JF02] detection cannot use such fragmented data since the distinguished event need not appear in every fragment of data. The *Phaser* algorithm cannot solve this problem as it performs a Hilbert transform to calculate its initial proto-phases, introducing transients on the order of 1 to 3 cycles. Methods based on forward integration [Win80, MM12] cannot handle partial cycles in data either, as these methods require collections of individual time series each having length exceeding several periods. In contrast, our algorithm readily handles this situation as long as the union of data over all trials fills an appropriate neighborhood of the cycle (see 1 above). Furthermore, we have presented preliminary evidence (the oscillator associated with the guineafowl data) that our form phase estimator may still accurately produce isochrons using data lying in only a neighborhood of a proper subset of the limit cycle. This preliminary finding is surprising, but the evidence is compelling, and we plan to investigate this in future work.

For scientists studying oscillatory systems, a phase-based model allows the experimentalist to compare the outcome of experimental treatment (e.g., a mechanical perturbation to locomotion) to that of the counterfactual unperturbed system by modeling the responses to be deterministic functions of phase [RKF08]. The results in §3.6 indicate that for systems of different dimensions and levels of noise, approximating phase using a Temporal 1-Form produced by the form phase algorithm presented here is superior to event-based phase estimates and the *Phaser* algorithm. The mean-square error of the phase estimate from the form-based technique is smaller. The inferred isochrons are closer to the ground truth. Finally, the phase response curves correspond better to those computed from the ground-truth systems.

Future work may extend our algorithm by judicious application of domain-specific knowledge, in particular by augmenting our series expansion for the residual ($\phi - \mathbf{d}\theta$) with new basis functions. Such extensions would no doubt improve the speed and accuracy of our algorithm. Another future direction could include the incorporation of more general large-scale numerical solvers for approximating the Temporal 1-Form, without resorting to explicit basis function expansions. Additionally, we have restricted ourselves in this chapter to the computation of classical asymptotic phase from noisy data. An interesting direction for future work would be to develop algorithms to compute generalized notions of asymptotic phase that have been developed for stochastic oscillators [SP10, SP13, TL14].

The Temporal 1-Form offers a new perspective on asymptotic phase for nonlinear oscillators. Its applications to the analysis of dynamical systems, its power in organizing empirical data, and its extension to systems with other attractor topologies remain to be explored.

CHAPTER IV

Data-driven Gait Optimality Testing in the Perturbed Stokes Regime

4.1 Acknowledgements

This chapter is based on joint work with Brian Bittner and Shai Revzen. In particular, Bittner deserves the credit for Figures 4.2, 4.3 and 4.4, as well as for writing the code used to generate these figures. All authors were supported by ARO grants W911NF-14-1-0573 and W911NF-17- 1-0306 to Revzen. Kvalheim would like to thank Jaap Eldering for introducing him to the relevance of NHIM theory to locomotion, for helpful comments and suggestions regarding the global asymptotic stability of the slow manifold of Theorem IV.9, and for other useful suggestions.

4.2 Introduction

In this chapter, we study how animals and robots move through space by deforming the internal "shape" of their body — typically in a cyclic fashion — to produce forces that propel the body. We call such cyclic shape deformations *gaits*. We study a class of locomotion which includes swimming and crawling in viscous media, in which the viscous damping forces are large compared to the inertia of the body. A classic exposition of such locomotors "living life at low Reynold's number" is given in [Pur77]. An important aspect of our work is that we consider the *perturbed Stokes regime* [EJ16] in which the inertia-damping ratio (or Reynolds number) is small but nonzero, as opposed to previous geometric mechanics literature addressing only the viscous or *Stokesian limit* which formally assumes the inertia-damping ratio is zero [KM96, KM95, HC11, BHR18]. We note that our methods are related to the realization of nonholonomic constraints as a limit of friction forces [Bre81, Kar81, Eld16].

For both scientific and engineering purposes, it is often of interest to ask whether a particular gait is is optimal with respect to a goal function. For animal locomotion, explicit equations of motion are nigh impossible to come by, and therefore directly testing animal gait optimality via analytic tools like the calculus of variations is not an option. However, if equations of motion can be obtained from data for the local dynamics on a tubular neighborhood of the gait cycle, then local optimality tests can be formulated and evaluated from the data. Such an approach was taken in [BHR18], which introduced an algorithm informed by geometric mechanics and utilizing data-driven techniques for studying oscillators [RG08, Rev09, RK15].

One limitation of [BHR18] was the assumption that motion was entirely kinematic, effectively assuming that the inertia-damping ratio is zero by assuming a *viscous connection*-based model as introduced by [KM95] and to be discussed more below. The real-world systems we are interested in have small — but always nonzero — inertia-damping ratio, and therefore we are interested in the extent to which the algorithm of [BHR18] can be improved.

By applying normally hyperbolic invariant manifold (NHIM) theory [Fen71, Fen74, Fen77, HPS77, Fen79] in a singular perturbation context, we show that an exponentially stable invariant *slow manifold* exists for small inertia-damping ratio (this was also shown in [EJ16]). Furthermore, this slow manifold is close to the viscous connection (viewed geometrically as a subbundle — hence as a submanifold — of state space), and therefore the dynamics restricted to the slow manifold are close to those assumed in the purely viscous case [KM96, KM95, HC11, BHR18], and reduces to those in the zero inertia-damping ratio limit. Aside from being nice theoretically, this result also has practical implications: it is possible to explicitly compute "correction terms" which, when added to the ideal viscous connection model, yield the dynamics restricted to the slow manifold. These dynamics are provably more accurate than those of the idealized viscous connection model alone, yet they still enjoy similar useful properties: (i) reduced dimensionality of the model and (ii) symmetry of the model are preserved. The computation of such correction terms is a fundamental technique in geometric singular perturbation theory [Fen79, Jon95], and has been used, e.g., to compute reduced-order models of robots with flexible joints [SKK87].

After laying the theoretical foundations needed to describe our class of models, we introduce an algorithm for estimating a reduced model of the dynamics of a locomoting body near its gait cycle. Such an algorithm would enable variational tests for local optimality of a gait cycle with respect to any cost functional that the model allows us to evaluate. We have in mind two applications: (i) verification of whether a postulated goal function is optimized for an observed animal gait, and (ii) optimization of robot gaits with "hardware-in-the-loop" — learning effective gaits without the need for precise models of the robot or its interactions with the environment.

The main obstacle to hardware-in-the-loop optimization is that a gait, being generated from a loop in the control inputs, is an infinite-dimensional object undergoing optimization. Any gradient calculation or optimization of a gait must deal with the inherently high dimensionality of the gait parameterization, and thus requires many experiments. Combined with the high practical cost of hardware experiments in terms of time and robot wear, this renders hardware-in-the-loop optimization nigh infeasible. A tractably computable local model can resolve this problem, by allowing the high-dimensional gradients to be computed by simulating the model, instead of directly using the hardware.

It is our hope that, through a combination of geometric mechanics and NHIM theory, we can develop an algorithm which can serve the purposes of both biologists and robotics engineers.

4.3 Background

In studying locomotion, we will consider dissipative Lagrangian mechanical systems on a product configuration space $Q = S \times G$ with coordinates (r, g), and with Lagrangian of the form kinetic minus potential energy. Here S is the *shape space* of the locomoting body, and G is a Lie group (typically a subgroup of the Euclidean group $\mathsf{SE}(3)$ of rigid motions) representing the body's position and orientation in the world. We assume throughout this chapter that S is compact. We will also assume that this system is subjected to external viscous drag forces which are linear in velocity[1].

If the physics of locomotion are independent of the body's position and orientation, then the Lagrangian $L(r, g)$ and viscous drag force $F_R(r, g)$ are independent of g. Under this symmetry assumption, [KM96] derived general equations of motion satisfied by g and by the *body momentum*[2] $p \in \mathfrak{g}^*$. For a detailed statement and derivations of these equations, see §4.7.

Let us suppose that the kinetic energy metric of the body is scaled by an inertial parameter $m > 0$, that the viscous drag force F_R is scaled by a damping parameter $c > 0$, and define $\epsilon \coloneqq \frac{m}{c}$ a Reynolds number-like ratio of the two. [KM96] showed that in the limit

[1] We make this assumption for simplicity. In principle, it should be possible to relax this assumption to derive modified but similar results for a force depending nonlinearly on velocities, as long as the linear approximation (with respect to velocities) of this force satisfies the same assumptions that we impose on our assumed linear force.

[2] Here $\mathfrak{g}^*$ is just a real vector space; for the knowledgable reader, it is the dual of the Lie algebra $\mathfrak{g}$ of G.

$\epsilon \to 0$, the equation of motion for g becomes independent of p and takes the simple form

$$\overset{\circ}{g} = -A_{\text{visc}}(r) \cdot \dot{r}, \tag{4.1}$$

where A_{visc} is called the *local viscous connection*, and $\overset{\circ}{g} \coloneqq \mathsf{D}\mathrm{L}_{g^{-1}}\dot{g}$ is the *body velocity*.[3]

On the other hand, [EJ16] studied the *perturbed Stokes regime*, in which ϵ is assumed to be small but nonzero. They showed that for ϵ sufficiently small, there is an exponentially stable invariant *slow manifold* M_ϵ for the full dynamics; we derive similar results tailored for our applications in §4.8. In §4.8 we also prove that the equations of motion for trajectories within M_ϵ take the form given by Theorem IV.1 below. Hence trajectories of the full dynamics satisfy Equation 4.41 below, possibly after a brief transient phase of convergence to M_ϵ. An illustration of this situation is in Figure 4.1 (in the figure, $\Omega = \mathbb{I}_{\text{loc}}^{-1} p$ is a linear function of p, where $\mathbb{I}_{\text{loc}}$ is defined in 4.7).

Theorem IV.1. *Assume that the shape space S is compact. For sufficiently small $\epsilon > 0$, there exist smooth fields of linear maps $B(r)$ and $(1,2)$ tensors $G(r)$ such that the dynamics restricted to the slow manifold M_ϵ satisfy*

$$\overset{\circ}{g} = -A_{\text{visc}}(r) \cdot \dot{r} + \epsilon B(r) \cdot \ddot{r} + \epsilon G(r) \cdot (\dot{r}, \dot{r}) + \mathcal{O}(\epsilon^2). \tag{4.2}$$

Remark IV.2. The (1,2) tensors $G(r)$ are *not*, in general, symmetric: e.g., they are unlike Hessians.

[BHR18] developed a data-driven algorithm for approximating the equations of motion of a locomotion system assuming the model (4.1). In this chapter, we extend the domain of application of that algorithm to approximate equations of motion of a locomoting system operating in the perturbed Stokes regime, wherein ϵ is allowed to be small but nonzero. We

[3] The body velocity is often written $g^{-1}\dot{g}$ by an abuse of notation which makes literal sense only for matrix Lie groups. In general, the body velocity is given by $\mathsf{D}\mathrm{L}_{g^{-1}}\dot{g} \in \mathsf{T}_e G \cong \mathfrak{g}$.

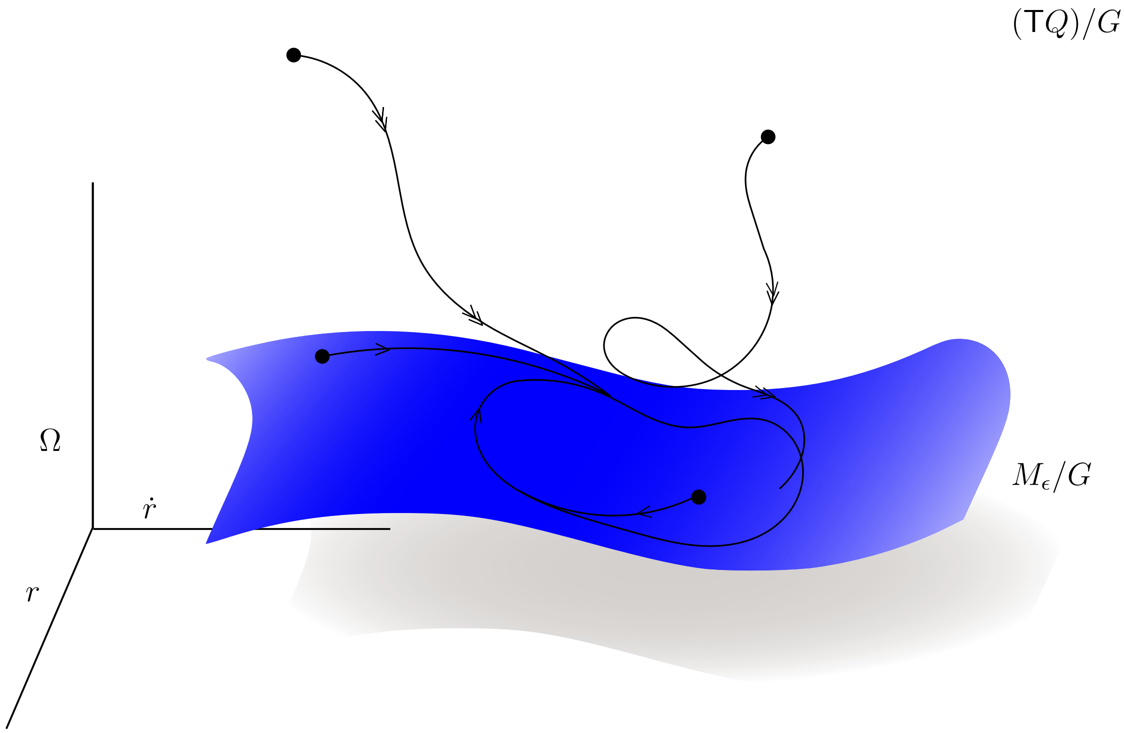


Figure 4.1: For small but nonzero inertia-damping ratio ϵ, an exponentially stable invariant "slow manifold" M_ϵ exists. Shown here is the invariant manifold M_ϵ/G (blue) for the reduced dynamics written on $(\mathsf{T}Q)/G$ in the case that the shape variables satisfy autonomous dynamics and have an attracting limit cycle (representing a gait). Note that in the special case that $Q = S \times G$, we have $(\mathsf{T}Q)/G \cong \mathsf{T}S \times \mathfrak{g}$.

accomplish this by assuming a model of the form (4.2).

4.4 Estimating data-driven models in the perturbed Stokes regime

In this section, we develop a data-driven algorithm for estimating the dynamics in a neighborhood of an exponentially stable limit cycle (gait) γ, based on the perturbed Stokes model (4.2). We do this in §IV.1, after first reviewing the previous algorithm of [BHR18] based on (4.1) for estimating the dynamics in the Stokesian limit. Our algorithm builds on the previous one by augmenting its regressors with additional ones motivated by (4.2), thereby extending predictive power into the perturbed Stokes regime. We conclude this section with a discussion in §4.4.2 of the applications we have in mind for our algorithm.

4.4.1 Determination of regressors for estimation of the dynamics

We first give an overview of the algorithm of [BHR18] for estimating dynamics in the Stokesian limit, given an input consisting of an ensemble of noisy trajectories near a limit cycle (for an underlying deterministic system). In the following, we will assume for simplicity that the gait cycle is contained within a single chart of shape space, so that we can use a single set of coordinates below; this is often the case in practice, such as the situation that shape space is a torus and the gait is contained in a contractible subset (this is also the case with our example in §4.5). Note that in this section we use the Einstein summation convention.

First, we assign a "phase" value[4] to each data point using an algorithm such as that of [RG08] or Chapter III, and then use these phase values to numerically compute a Fourier series model of the limit cycle as a function of phase. Here the utility of assigning phase values is that it enables each point $r(t)$ from the trajectory ensemble to be assigned to a point $\gamma(t) \in \Gamma$ (where $\Gamma \coloneqq \gamma(\mathbb{R})$). Hence in local coordinates we have the observation $r(t) = \gamma(t) + \delta(t)$, with $\delta(\cdot)$ representing a perturbation off of the limit cycle which is small in a Sobolev sense. By this we mean that the (e.g., Euclidean) norms of δ, $\dot{\delta}$ and $\ddot{\delta}$ are uniformly bounded by some $\kappa > 0$. Introducing coordinates and Taylor expanding (4.1) gives:

$$\mathring{g}^k \approx -\underbrace{A_i^k \dot{\gamma}^i}_{C_0^k} - \underbrace{A_i^k}_{C_1^k} \dot{\delta}^i - \underbrace{\frac{\partial A_i^k}{\partial r^j} \dot{\gamma}^i}_{C_2^k} \delta^j - \underbrace{\frac{\partial A_i^k}{\partial r^j}}_{C_3^k} \delta^j \dot{\delta}^i, \tag{4.3}$$

where we have omitted higher-order terms, the subscript of A_{visc}, and the nonlinear γ dependence of the local expression A_i^k. From (4.3), we formulate a least-squares problem for each $\gamma \in \Gamma$ with the regressors $1, \delta,\ \dot{\delta}, \delta \otimes \dot{\delta}$ coming from the terms r assigned the same phase as

[4] This value could represent the *asymptotic phase* of $r(t)$ (Chapter III), but we do not require this here.

γ. The least squares solution of the regression yields us estimates of $C_0, \ldots, C_3$ as a function of γ, which we then model as a Fourier series function of phase, providing an approximation of the dynamics of (4.1) around the limit cycle Γ for any possible δ.

Now, Theorem IV.1 tells us that additional terms appear in the (reduced) equations of motion in the perturbed Stokes regime. We model them in the same way as we did A, giving a functional form of a more accurate approximation of the dynamics.

Similar to the above, we Taylor expand the right hand side of (4.2) at $\gamma(t)$, setting $r(t) = \gamma(t) + \delta(t)$, to obtain

$$\begin{aligned}
\mathring{g}^k \approx & -A^k_i \dot{\gamma}^i - A^k_i \dot{\delta}^i - \frac{\partial A^k_i}{\partial r^j} \delta^j \dot{\gamma}^i - \frac{\partial A^k_i}{\partial r^j} \delta^j \dot{\delta}^i + \epsilon \left(B^k_i \ddot{\gamma}^i + B^k_i \ddot{\delta}^i + \frac{\partial B^k_i}{\partial r^j} \delta^j \ddot{\gamma}^i + \frac{\partial B^k_i}{\partial r^j} \delta^j \ddot{\delta}^i \right. \\
& \ldots + G^k_{ij} \dot{\gamma}^i \dot{\gamma}^j + G^k_{ij} \dot{\gamma}^i \dot{\delta}^j + G^k_{ij} \dot{\delta}^i \dot{\gamma}^j + G^k_{ij} \dot{\delta}^i \dot{\delta}^j \\
& \left. \ldots + \frac{\partial G^k_{ij}}{\partial r^\ell} \delta^\ell \dot{\gamma}^i \dot{\gamma}^j + \frac{\partial G^k_{ij}}{\partial r^\ell} \delta^\ell \dot{\gamma}^i \dot{\delta}^j + \frac{\partial G^k_{ij}}{\partial r^\ell} \delta^\ell \dot{\delta}^i \dot{\gamma}^j + \frac{\partial G^k_{ij}}{\partial r^\ell} \delta^\ell \dot{\gamma}^i \dot{\delta}^j + \frac{\partial G^k_{ij}}{\partial r^\ell} \delta^\ell \dot{\delta}^i \dot{\delta}^j \right).
\end{aligned} \tag{4.4}$$

Partitioning these terms according to their dependence on δ,$\dot{\delta}$, and $\ddot{\delta}$, we formulate a least squares regression for (4.4) for each $\gamma \in \Gamma$ with regressors 1, δ, $\dot{\delta}$, $\delta \otimes \dot{\delta}$, $\ddot{\delta}$, $\delta \otimes \ddot{\delta}$, $\dot{\delta} \otimes \dot{\delta}$, and $\delta \otimes \dot{\delta} \otimes \dot{\delta}$. Because it is the only term of order κ^3, we find that in practice this last term can often be omitted if $\kappa > 0$ is sufficiently small. In the remainder of this chapter, we refer to these regressors as the "perturbed Stokes regressors", and refer to those used in the old algorithm as the "Stokes regressors."

Remark IV.3. All tensors appearing in Equations (4.3) and (4.4) above are generally not symmetric, and therefore the order of terms matters.

Remark IV.4. Notice that even in (4.3), there are some constraints that the regression does not enforce. Namely, $C_0 = [C_1]_i \dot{\gamma}^i$ and $C_2 = [C_3]_i \dot{\gamma}^i$. Ignoring these implicit constraints and solving the least squares problem, one finds that these constraints are not respected in practice. However, an important consequence of Theorem IV.1 and Equation (4.4) is that statistics alone is likely *not* to blame for this inconsistency phenomenon for systems operating in the perturbed Stokes regime, rather than in the Stokesian limit — this is because some independent new terms appear in $C_1, \ldots, C_3$ which would break these constraints.

4.4.2 Utility of the estimated models

The data-driven models computed by the process described above have predictive power locally, in a neighborhood of a gait cycle. By supplying a shape trajectory in this neighborhood, the local model will predict a corresponding trajectory for the body in the world. We assume that we are interested in some $\mathbb{R}$-valued goal functional $\tilde{\phi}(\gamma, g_\gamma)$ defined on an appropriate space of trajectories. Here the group trajectory $g_\gamma(t)$ is determined by the gait $\gamma(t)$ via (4.2), and therefore we may consider the goal functional $\phi(\gamma) \coloneqq \tilde{\phi}(\gamma, g_\gamma)$ to be a function of γ alone. Knowledge of a local model of (4.2) enables two primary applications.

1. First, the gait of an organism can be tested for optimality by checking that

$$\frac{\partial}{\partial s}\phi(\gamma_s)|_{s=0} = 0$$

for all smooth variations γ_s of a gait γ (where $\gamma_0 = \gamma$). This condition is necessary for optimality, but depending on the choice of ϕ it might be possible to argue on physical grounds that its satisfaction is also sufficient for optimality (alternatively, second derivative tests can be formulated). Now there are infinitely many such smooth variations, but in practice this problem is rendered computationally tractable by considering a finite-dimensional approximation consisting of an appropriate family γ_p with $p \in \mathbb{R}^N$ and N sufficiently large, and numerically computing the gradient $\nabla_p \phi(\gamma_p)$. If the Euclidean norm of this gradient is sufficiently small at some parameter value p^*, then it might be possible to argue that the gait is nearly extremal (or possibly optimal) with respect to[5] ϕ. The value here is that this procedure can be carried out *directly from observation* and without need for any general model of body-environment interactions beyond the assumptions needed to justify Theorem IV.1.

2. The same gradient $\nabla_p \phi(\gamma_p)$ can be used in gradient ascent algorithms applied to hardware-in-the-loop optimization of a robot. As discussed in §4.2, the main obstacle to hardware-in-the-loop optimization is the high dimensionality of the parameter space $\mathbb{R}^N \ni p$ used to approximate an infinite-dimensional space of gaits. This is because the high-dimensionality requires many experiments in order to numerically compute $\nabla_p \phi(\gamma_p)$, and this imposes a high practical cost related to time expenditure, robot wear, etc. However, our algorithm allows this gradient calculation to be obtained via simulation rather than physical experiment. For each gait, we only require enough experimental data for building a good local model, which is subsequently simulated to compute gradients. Such decoupling of hardware-in-the-loop optimization into data collection and simulation components could enable rapid adaptation of robot motion in the face of foreign environments or mechanical failures.

4.5 Performance comparison of the data-driven models

One of the primary contributions of this chapter is the introduction of new regressors based on Theorem IV.1, which we use to augment the regressors used in the algorithm of [BHR18] for estimating the dynamics near a gait. This allows us to extend the domain of validity of their algorithm — the Stokesian limit — to the perturbed Stokes regime. In this section, we will compare the accuracy of the new and old approaches on a simple swimming

[5] It is possible to justify this finite-dimensional approximation procedure assuming that the family γ_p satisfies certain properties, and furthermore it is possible to argue that such suitable finite-dimensional families always exist using techniques of [Mil69, Sec. 16] on finite-dimensional approximations of path spaces. Though quite interesting, we do not wish to get bogged down by these technicalities here.

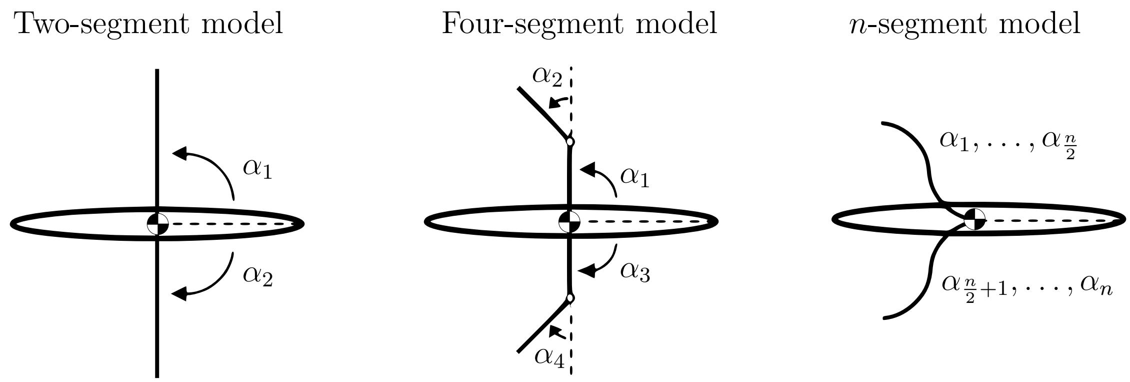


Figure 4.2: We study a swimmer with an even number of segments, which are divided among two paddles. The body has a mass m, moment inertia $I = m\bar{I}$, and length L (we explicitly write the factor $m > 0$ in the moment of inertia for scaling purposes). Each segment has length $\frac{L}{n}$, where n is the number of segments.

model.

4.5.1 Modeling the swimmer

We test the prediction quality of both models on a simple swimming model. The system shown in Figure 4.2 has uniformly distributed mass along a central link, with two chains of massless links extending from the center of the body. Each paddle can be broken up into an arbitrary number n of equally spaced links, which sum to the same total length. This allows us to vary the properties of the system from behaving like a boat with oars (for $n = 1$) to a bacterial cell with flagella (for n large).

The system moves in the plane, with configuration space the product $\mathbb{T}^{2n} \times \mathsf{SE}(2)$ of the $2n$-torus and the special Euclidean group of planar rigid motions $\mathsf{SE}(2)$, and the system has $\mathsf{SE}(2)$ symmetry. The group element $g \in \mathsf{SE}(2)$ provides the position and orientation of the central link in world coordinates, i.e., with respect to some fixed inertial reference frame. In what follows, we represent g as a column vector $g = [x, y, \theta]^T$, and similarly represent $\dot{g}$ as a column vector. We define the body velocity

$$\mathring{g} = R^{-1}(\theta)\dot{g} = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \dot{g}. \tag{4.5}$$

We treat the link at the main body (length L) and the links corresponding to the flippers (length d) as slender members, and model their drag forces according to Cox theory [Cox70] using the drag matrices

$$C_D = c \begin{bmatrix} C_x d & 0 & 0 \\ 0 & C_y d & 0 \\ 0 & 0 & \frac{1}{12} d^3 C_y \end{bmatrix}, \quad C_L = c \begin{bmatrix} C_x L & 0 & 0 \\ 0 & C_y L & 0 \\ 0 & 0 & \frac{1}{12} L^3 C_y \end{bmatrix}, \tag{4.6}$$

where the factor $c > 0$ is explicitly written for later scaling purposes. The drag coefficient ratio C_y/C_x has a maximum value of 2 corresponding to the limit of infinitesimally thin segments, and we will assume this limiting ratio here (c.f. [HC13, Sec. 2.B]). Given these drag matrices, the drag wrench (force and moment in body coordinates) on the ith segment can be written as

$$F_i = c\overline{F}_i - R(\alpha_i) C_D \begin{bmatrix} \cos(\alpha_i) & \sin(\alpha_i) & 0 & 0 \\ -\sin(\alpha_i) & \cos(\alpha_i) & \frac{d}{2} & \frac{d}{2} \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mathring{g} \\ \dot{\alpha}_i \end{bmatrix}, \tag{4.7}$$

and the wrench on the central link can be written as

$$F_{body} = c\overline{F}_{body} = -C_L \mathring{g}. \tag{4.8}$$

These forces act on the body (which has uniformly distributed mass m and moment of inertia $I = m\overline{I}$ about its midpoint) yielding the following equations of motion in world coordinates:

$$\ddot{g} = \begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{\theta} \end{bmatrix} = \frac{1}{\epsilon} \begin{bmatrix} \frac{1}{m} & 0 & 0 \\ 0 & \frac{1}{m} & 0 \\ 0 & 0 & \frac{1}{\overline{I}} \end{bmatrix} R(\theta) \left(\overline{F}_{d_1} + \overline{F}_{d_2} + \overline{F}_{body} \right), \tag{4.9}$$

where $\epsilon := \frac{m}{c}$ is the inertia-damping ratio.

Upon inspection of Equation (4.9), we see that by modifying $\epsilon > 0$ we can directly adjust the ratio of inertial to viscous forces in the swimming model. The Stokesian limit corresponds to $\epsilon \to 0$; on the other hand, the $\epsilon \to \infty$ limit corresponds to a fully "momentum dominated" regime, wherein viscous effects are negligible and motion is governed by conservation of momentum via Noether's theorem (see Corollary IV.6 in §4.7.1). In the following section §4.5.2 we simulate the swimming model at a variety of ϵ values, and compare the performance of the old and new algorithms for estimating the dynamics near a gait cycle.

4.5.2 Comparison of accuracy of estimated models

In all simulations in this section, we used the parameter values $L = 1$, $d = 0.5$, $C_x = 1$, $C_y = 2$, $m = 1$, and $\overline{I} = 1$. The procedure we used for generating simulations for experiments in this section is identical to that described in [BHR18]. Briefly, an experiment consists of 30 cycles of a numerically integrated stochastic differential equation (SDE) representing shape space dynamics consisting of a deterministic oscillator perturbed by system noise (see [BHR18, Sec. 6.2] for precise details on the SDE, parameter values used, etc.). These noisy shape dynamics are used to drive the body momentum and group dynamics via the full equations of motion (4.23) derived in §4.7.3. During each simulation, a "ground truth" body velocity trajectory $\mathring{g}_G$ is recorded and stored as a benchmark for the accuracy of the body velocities predicted by the new and old algorithms. The body velocities computed using the

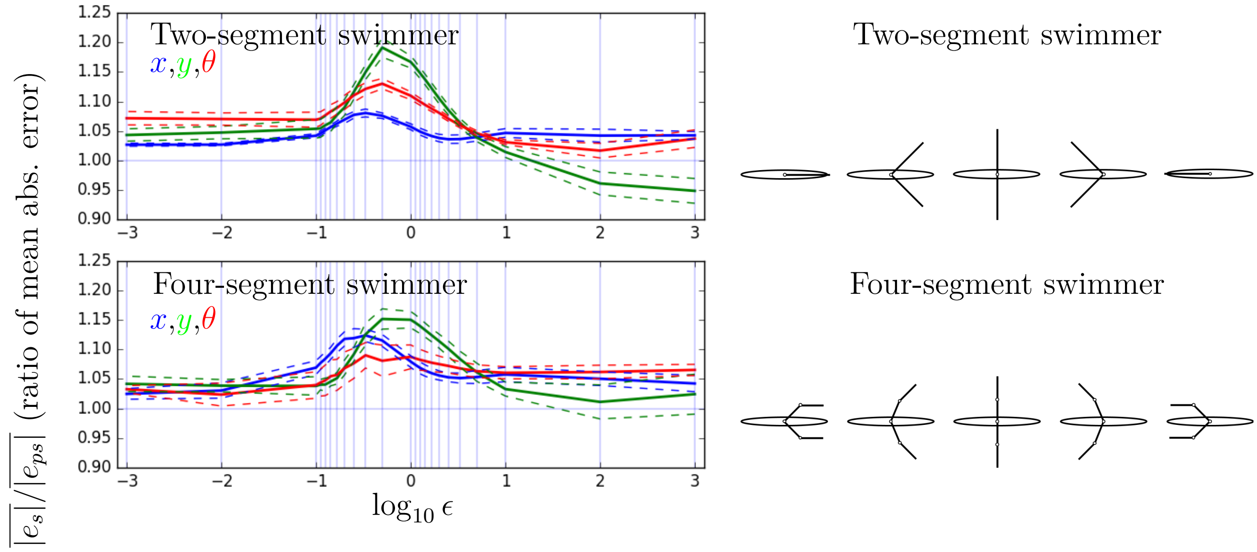


Figure 4.3: The two-segment and four-segment swimmers approximately follow the symmetric gaits depicted on the right (modulo noise). At each value of ϵ depicted by a blue vertical line, 8 experiments were conducted with each experiment consisting of 30 simulated cycles. A local model of the dynamics of each swimmer was constructed using the old algorithm (with "Stokes regressors") (s) and using the new algorithm (with "perturbed Stokes regressors") (ps). Using the body velocity recorded during simulation as "ground truth", the prediction accuracy of both models are compared by plotting the ratio of the mean absolute prediction errors, which are computed on an individual trial basis. Qualitatively, it seems clear that the additional regressors present in the new algorithm provide a nontrivial improvement in predictive power in the region $\log_{10}\epsilon \in [-1, 1]$.

new algorithm (with the "perturbed Stokes regressors") and using the old algorithm (with the "Stokes regressors") will be denoted $\mathring{g}_{ps}$ and $\mathring{g}_s$, respectively. Given an experiment with k data points, the mean absolute errors for each of the two algorithms are denoted by:

$$\overline{|e_{ps}|} \coloneqq \frac{1}{k}\sum_{i=1}^{k} |\mathring{g}_{ps,i} - \mathring{g}_{G,i}| \qquad \overline{|e_s|} \coloneqq \frac{1}{k}\sum_{i=1}^{k} |\mathring{g}_{s,i} - \mathring{g}_{G,i}| \tag{4.10}$$

Figure 4.3 shows the results from an experiment in which we tested the respective prediction quality of both models across a spectrum of dynamical regimes corresponding to

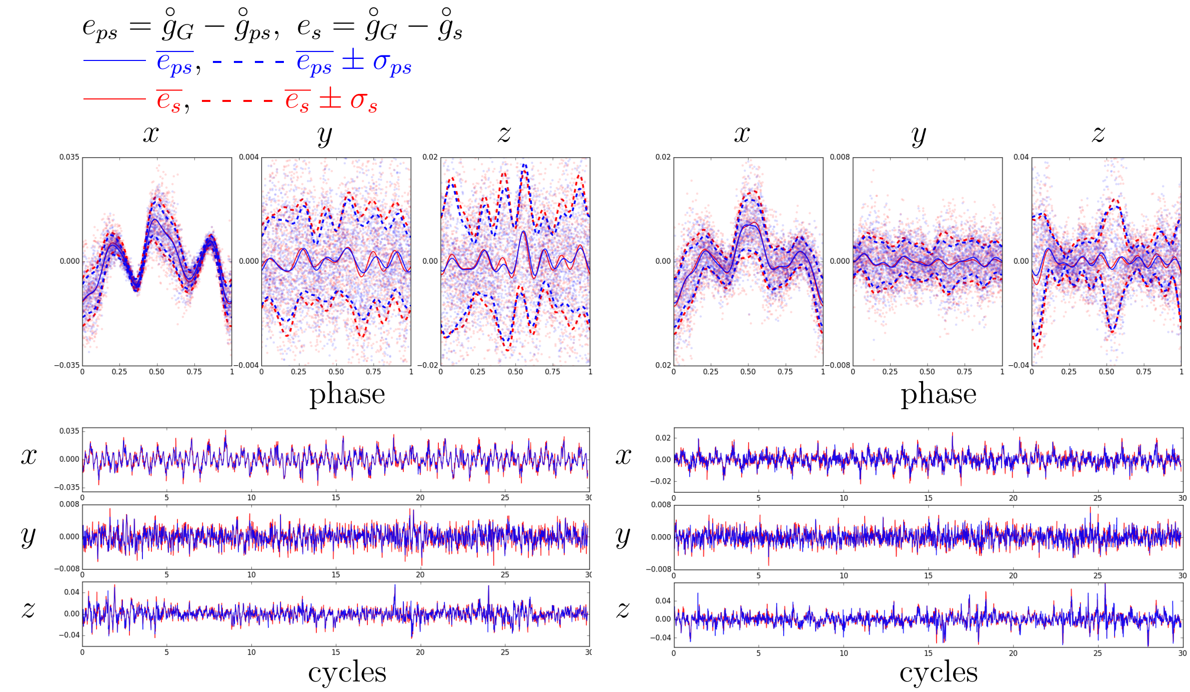


Figure 4.4: Here, each platform (two segment on the left, four segment on the right) is cycled at $\epsilon = 0.5$. Each experiment is taken from Figure 4.3. The modeling error is plotted with respect to cycles (bottom) and as a scatter plot vs. phase (top). It is visually clear that the perturbed Stokes regressors used in the new algorithm provide a more accurate prediction of body velocity compared to that of the old algorithm using the Stokes regressors.

different values of ϵ. The gait for each swimmer involves symmetric sinusoidal oscillations of the joints, such that the relative angles between adjacent segments are equal and such that the amplitude of the oscillation is π. By simulating the dynamics for a broad range of ϵ values, we are able to clearly discern a regime in which the perturbed Stokes regressors display a significant advantage. The additional regressors used in the new algorithm appear to improve prediction the most approximately near $\epsilon = 0.5$ sec. Figure 4.4 illustrates the difference in performance at $\epsilon = 0.5$ sec specifically.

4.5.3 Discussion

The results show that for the versions of the simple swimming models that we tested (two-link and four-link) depicted in Figure 4.2, there exists a sizable window of ϵ values wherein the new algorithm (using the perturbed Stokes regressors) provides models of superior quality when compared to prior work (using the Stokes regressors). For very small values of ϵ, the system is nearing the Stokesian limit and therefore we do not expect much improvement from the new algorithm in this regime; this is consistent with our experiments. For very large values of ϵ, predictive quality of both algorithms are hindered by three factors, although only the first two can be observed here. First, the $\mathcal{O}(\epsilon^2)$ term in Theorem IV.1 becomes more significant as ϵ increases. This issue is insurmountable if we restrict ourselves to Stokes regressors, but it is possible to compute correction terms which are higher order in ϵ and which can inform the selection of additional regressors for addition to our algorithm. We do not pursue this here, but it is one possible direction for future work. The second issue hindering the predictive quality of both algorithms is more serious: for ϵ sufficiently large, we expect a bifurcation in which the slow manifold (whose existence is guaranteed by Theorem IV.9 in §4.8) ceases to exist. For such values of ϵ, the hypotheses of Theorem IV.1 are not satisfied, its conclusions are violated, and a reduced-order model does not exist. Finally, the third issue is relevant for real-world systems. For sufficiently large values of ϵ, we expect the full complications of fluid dynamics to come to the fore, eventually invalidating our linear viscous friction model (and thereby even rendering ϵ ill-defined). We suspect that this third issue should not occur until after ϵ is already sufficiently large for the slow manifold to have disappeared, but this is pure speculation and it would be interesting to explore this issue further.

4.6 Discussion

We have shown that the accuracy of data-driven models motivated from geometric mechanics can be improved by using a collection of regressors derived from an asymptotic series approximation of an attracting invariant manifold in the small parameter representing the ratio of inertial to viscous forces, (a Reynolds number). The existence of such an invariant manifold was previously known in similar situations[6], as were the approximation techniques we employed, but the combination of these together for producing data-driven models of locomotion is a novel contribution. In a simulation where we tested geometrically similar motions over a large range of Reynolds numbers, we obtained improvements of 20% over several orders of magnitude of the Reynolds number when compared to previous work, suggesting that these better-informed models can indeed capture the perturbed Stokes regime more accurately.

Future work will include application of our algorithm to questions of locomotion optimality in animals, and to hardware-in-the-loop optimization of robot motions. An additional direction for future work is the selection of regressors and regression techniques for hybrid dynamical systems, and for non-viscous dissipation models.

4.7 Derivation of the equations of motion

In this and the following section we consider systems more general than those considered earlier, and in so doing assume that the reader is familiar with some basic concepts in geometric mechanics and differential geometry: Lie groups, group actions, and principal bundles. Basic concepts and definitions from the theory of principal bundles are reviewed in Appendix C; we refer the reader to [KN63, MR94, Lee13, Blo15] for the relevant standard definitions related to Lie groups and group actions, and we refer the reader to [KN63, MOWZ91, Mar93, Blo15] for further material on bundles.

We consider a mechanical system on a configuration space Q whose Lagrangian is of the form kinetic minus potential energy. We will also consider this system to be subjected to external viscous forcing arising from a Rayleigh dissipation function, and also subjected to

[6] But see the discussion preceding Theorem IV.9 in §4.8, which details how our result differs from that of [EJ16].

an external force exerted by the locomoting body. We are interested in the situation that we have a smooth action $\theta\colon G \times Q \to Q$ of a Lie group G on Q, such that the Lagrangian, viscous forces, and external force are all invariant under the action. In this case, we say that G is a *symmetry group*.

In §4.7.1, we will define some geometric quantities on Q which encode information about the symmetry and the dynamics. Working in coordinates induced by a local trivialization, in §4.7.2 we derive the equations of motion in terms of these quantities. In §4.7.3, we recall how the equations become governed by the so-called viscous connection in the Stokesian limit [KM96, EJ16], which will set the stage for our derivation in §4.8 of a corrected reduced-order model for the perturbed Stokes regime.

4.7.1 The mechanical and viscous connections

In this section, we define the mechanical and viscous (or Stokes) connections, roughly following [KM96]. We consider a Lagrangian $L\colon \mathsf{T}Q \to \mathbb{R}$ which is invariant under the lifted action $\mathsf{D}\theta_g$ of G on $\mathsf{T}Q$ (here D denotes the derivative or pushforward). We assume the Lagrangian to be of the form kinetic minus potential energy, where kinetic energy is given by $\frac{m}{2}k$, where $m > 0$ is a mass parameter, k is a smooth symmetric bilinear form, and mk is the *kinetic energy metric*. In what follows, we assume that k is positive definite when restricted to tangent spaces to G orbits, but *not* necessarily that k is positive definite on all tangent vectors[7]. Denoting by $\mathfrak{g}$ the Lie algebra of G and $\mathfrak{g}^*$ its dual, we define the (Lagrangian) *momentum map* $J\colon \mathsf{T}Q \to \mathfrak{g}^*$ via

$$\langle J(v_q), \xi \rangle = \langle \mathbb{F}L(v_q), \xi_Q(q) \rangle = mk_q(v_q, \xi_Q(q)), \tag{4.11}$$

[7]This does not affect any of the following derivations and results. However, this generality is merely a convenience ensuring that our results apply to certain idealized examples, e.g., linkages with some links having zero mass. Of course such examples are not physical and, e.g., must be supplemented with assumptions to ensure that the massless links have well-defined dynamics.

where $v \in \mathsf{T}_qQ$ and $\xi \in \mathfrak{g}$. Here $\mathbb{F}L\colon \mathsf{T}Q \to \mathsf{T}^*Q$ is the *fiber derivative* of L given by $\mathbb{F}L(v_q)(w_q) := \frac{\partial}{\partial s}|_{s=0} L(v_q + sw_q)$, and the smooth vector field ξ_Q on Q is the *infinitesimal generator* defined by $\xi_Q(q) := \frac{\partial}{\partial s}|_{s=0} \theta_{\exp(s\xi)}(q)$. We define the *mechanical connection* $\Gamma_{\text{mech}}\colon \mathsf{T}Q \to \mathfrak{g}$ via $\Gamma_{\text{mech}}(v_q) := \mathbb{I}^{-1}(q)J(v_q)$, where $\mathbb{I}(q)\colon \mathfrak{g} \to \mathfrak{g}^*$ is the *locked inertia tensor* defined via

$$\langle \mathbb{I}(q)\xi, \eta \rangle := \langle \mathbb{F}L(\xi_Q(q)), \eta_Q(q) \rangle = mk_q(\xi_Q(q), \eta_Q(q)), \tag{4.12}$$

where $\xi, \eta \in \mathfrak{g}$.

We now follow an analogous procedure to define the viscous connection $\Gamma_{\text{visc}}\colon \mathsf{T}Q \to \mathbb{R}$. We consider a Rayleigh dissipation function $R\colon \mathsf{T}Q \to \mathbb{R}$ defined in terms of a G-invariant smooth symmetric bilinear form ν on Q: $R(v_q) := \frac{c}{2}\nu_q(v_q, v_q)$, where $c > 0$ is a parameter representing the amount of damping or dissipation in the system due to viscous forces. As with k, we assume that ν is positive definite when restricted to tangent spaces to G orbits, but *not* necessarily that ν is positive definite on all tangent vectors[8]. The corresponding force field $F_R\colon \mathsf{T}Q \to \mathsf{T}^*Q$ is given by minus the fiber derivative of R, $F_R := \mathbb{F}(-R)$. We define a map $K\colon \mathsf{T}Q \to \mathfrak{g}^*$, analogous to the momentum map J, via

$$\langle K(v_q), \xi \rangle = \langle F_R(v_q), \xi_Q(q) \rangle = -c\nu_q(v_q, \xi_Q(q)), \tag{4.13}$$

where $v \in \mathsf{T}_qQ$ and $\xi \in \mathfrak{g}$. We define the *viscous connection* or *Stokes connection* $\Gamma_{\text{visc}}\colon \mathsf{T}Q \to \mathfrak{g}$ via $\Gamma_{\text{visc}}(v_q) := \mathbb{V}^{-1}(q)K(v_q)$, where $\mathbb{V}(q)\colon \mathfrak{g} \to \mathfrak{g}^*$ is defined via

$$\langle \mathbb{V}(q)\xi, \eta \rangle := \langle F_R(\xi_Q(q)), \eta_Q(q) \rangle = -c\nu_q(\xi_Q(q), \eta_Q(q)), \tag{4.14}$$

where $\xi, \eta \in \mathfrak{g}$.

[8] This generality simply allows for, e.g., the situation of a linkage in which not all links are subject to viscous forces.

Using the G-invariance of L and ν, a calculation shows that Γ_{mech} and Γ_{visc} are equivariant with respect to the adjoint action of G on $\mathfrak{g}$:

$$\forall g \in G \colon \Gamma_{\text{mech}} \circ \mathsf{D}\theta_g = \mathrm{Ad}_g \circ \Gamma_{\text{mech}}, \quad \Gamma_{\text{visc}} \circ \mathsf{D}\theta_g = \mathrm{Ad}_g \circ \Gamma_{\text{visc}} \tag{4.15}$$

Hence if the natural projection $\pi_Q \colon Q \to Q/G$ from Q to the space of orbits Q/G of points in Q is a principal G-bundle, then the mechanical and viscous connections Γ_{mech} and Γ_{visc} are indeed principal connections (see Definition C.15); this justifies their titles.

Now in order for our system to move itself through space, we also allow there to be a G-invariant external force $F_E \colon \mathbb{R} \times \mathsf{T}Q \to \mathsf{T}^*Q$ exerted by the locomoting body, which physically implies that F_E takes values in the annihilator of $\ker \mathsf{D}\pi_Q$, the distribution generated by the group orbits (c.f. [EJ16, Sec. 3.3]). For future use, we now prove the following

Proposition IV.5. *The derivative of J along trajectories of the G-invariant mechanical system is given by*

$$\dot{J} = K, \tag{4.16}$$

making the canonical identifications $\mathsf{T}_J\mathfrak{g} \cong \mathfrak{g}$.

Proof. We compute in a local trivialization on $\mathsf{T}Q$ induced by a chart for Q, so that we may write a trajectory as $(q, \dot{q})$. Note that in such local coordinates, $\mathbb{F}L(q, \dot{q})(v_q) = \frac{\partial L(q,\dot{q})}{\partial \dot{q}} v_q$. Hence

$$\begin{aligned} \langle \dot{J}(q, \dot{q}), \xi \rangle &= \frac{d}{dt}\left(\frac{\partial L(q(t), \dot{q}(t))}{\partial \dot{q}} \xi_Q(q(t)) \right) \\ &= \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} \right) \xi_Q(q) + \frac{\partial L}{\partial \dot{q}} \mathsf{D}\xi_Q(q)\dot{q} \\ &= \left(\frac{\partial L}{\partial q} + F_R + F_E \right) \xi_Q(q) + \frac{\partial L}{\partial \dot{q}} \mathsf{D}\xi_Q(q)\dot{q}, \end{aligned} \tag{4.17}$$

where we obtained the last line using $\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = F_R + F_E$, which follows from the Lagranged'Alembert principle [Blo15, p. 8]. Since F_E annihilates tangent vectors to group orbits, $\langle F_E, \xi_Q(q)\rangle = 0$. Hence rearranging and letting Φ_ξ^s denote the flow of ξ_Q, we find

$$\begin{aligned}\langle \dot{J}(q,\dot{q}), \xi\rangle &= \frac{\partial}{\partial s} L\left(\Phi_\xi^s(q(t)), \mathsf{D}\Phi_\xi^s(q(t))\dot{q}(t)\right) + \langle F_R(q,\dot{q}), \xi_Q(q)\rangle \\ &= \frac{\partial}{\partial s} L\left(\Phi_\xi^s(q(t)), \mathsf{D}\Phi_\xi^s(q(t))\dot{q}(t)\right) + \langle K(q,\dot{q}), \xi\rangle.\end{aligned}$$

The derivative term is zero due to the invariance of L under the action of G, so from the arbitrariness of $\xi \in \mathfrak{g}$ we obtain the desired result. □

As a corollary, we obtain a slight generalization of the classical Noether's theorem.

Corollary IV.6 (Noether's theorem)**.** *Consider a mechanical system given by a G-invariant Lagrangian of the form kinetic minus potential energy. Assume that the only external forces take values in the annihilator of the distribution tangent to the G orbits. Then the derivative of the momentum map J along trajectories satisfies*

$$\dot{J} = 0.$$

Proof. Set $K = 0$ in Proposition IV.5. □

4.7.2 Local form of the equations of motion

Assuming that the action of G on Q is free and proper [Lee13, Ch. 21] so that $\pi_Q\colon Q \to Q/G$ is a principal G-bundle, we now derive the equations in a local trivialization, following [KM96]. In a local trivialization $U \times G$, π_Q simply becomes projection onto the first factor and the G action is given by left multiplication on the second factor. We define $S \coloneqq Q/G$ to be the *shape space* representing all possible shapes of a locomoting body, and we write a point in the local trivialization as $(r, g) \in U \times G$ where $U \subset S$. We assume that U is the domain of a chart for S, so that we have induced coordinates $(r, \dot{r})$ for $\mathsf{T}U$.

The equivariance property (4.15) of the connection forms $\Gamma_{\mathrm{mech}}, \Gamma_{\mathrm{visc}}$ imply that they may be written in the trivialization as[9]

$$\begin{aligned} \Gamma_{\mathrm{mech}}(r,g)\cdot(\dot{r},\dot{g}) &= \mathrm{Ad}_g\left(g^{-1}\dot{g} + A_{\mathrm{mech}}(r)\cdot\dot{r}\right) \\ \Gamma_{\mathrm{visc}}(r,g)\cdot(\dot{r},\dot{g}) &= \mathrm{Ad}_g\left(g^{-1}\dot{g} + A_{\mathrm{visc}}(r)\cdot\dot{r}\right), \end{aligned} \tag{4.18}$$

where $A_{\mathrm{mech}}\colon \mathsf{T}U \to \mathfrak{g}$ and $A_{\mathrm{visc}}\colon \mathsf{T}U \to \mathfrak{g}$ are respectively the *local mechanical connection* and *local viscous connection*. We define a diffeomorphism $(r,\dot{r},g,\dot{g}) \mapsto (r,\dot{r},g,p)$, with p the *body momentum* defined by

$$p \coloneqq \mathrm{Ad}_g^* J \in \mathfrak{g}^*. \tag{4.19}$$

Here Ad_g^* is the dual of the adjoint action Ad_g of G on $\mathfrak{g}$. We additionally define

$$\begin{aligned} \mathbb{I}_{\mathrm{loc}} &\coloneqq \mathrm{Ad}_g^* \mathbb{I} \mathrm{Ad}_g \colon \mathfrak{g} \to \mathfrak{g}^* \\ \mathbb{V}_{\mathrm{loc}} &\coloneqq \mathrm{Ad}_g^* \mathbb{V} \mathrm{Ad}_g \colon \mathfrak{g} \to \mathfrak{g}^* \end{aligned} \tag{4.20}$$

to be the local forms of $\mathbb{I}$ and $\mathbb{V}$. We note that the invariance of the Lagrangian L and Rayleigh dissipation function R under G, together with the general identity $\mathsf{D}\theta_g \xi_Q(q) = (\mathrm{Ad}_g\xi)_Q(\theta_g(q))$, imply that $\mathbb{I}_{\mathrm{loc}}(r), \mathbb{V}_{\mathrm{loc}}(r)$ depend on the shape variable r only.

Rearranging (4.18), using the expressions (4.19), (4.20), and using Proposition IV.5, we obtain the equations of motion

$$\begin{aligned} g^{-1}\dot{g} &= -A_{\mathrm{mech}}\cdot\dot{r} + \mathbb{I}_{\mathrm{loc}}^{-1}p \\ \dot{p} &= \mathbb{V}_{\mathrm{loc}}(A_{\mathrm{visc}} - A_{\mathrm{mech}})\cdot\dot{r} + \mathbb{V}_{\mathrm{loc}}\mathbb{I}_{\mathrm{loc}}^{-1}p + \mathrm{ad}^*_{\mathbb{I}_{\mathrm{loc}}^{-1}p}\,p - \mathrm{ad}^*_{A_{\mathrm{mech}}\cdot\dot{r}}\,p, \end{aligned} \tag{4.21}$$

where we have suppressed the r-dependence of $A_{\mathrm{mech}}, A_{\mathrm{visc}}, \mathbb{I}_{\mathrm{loc}}, \mathbb{V}_{\mathrm{loc}}$ for readability. Notice

[9] As mentioned earlier, the notation $g^{-1}\dot{g}$ makes literal sense as matrix multiplication for matrix Lie groups, but in general this standard abuse of notation should be interpreted as shorthand for $\mathsf{D}\mathrm{L}_{g^{-1}}\dot{g}$, with $\mathrm{L}_g\colon G \to G$ left multiplication. However, in these final technical sections of the chapter we embrace this abuse of notation, forgoing the less traditional notation $\mathring{g}$ used earlier in the paper.

that the $\dot{p}$ equation is completely decoupled from g.

In this chapter, we are interested in the effect of shape changes on body motion, and not on the generation of shape changes themselves. Hence we have suppressed the equations for $\dot{r}, \ddot{r}$ from (4.21), simply viewing $r, \dot{r}$ as inputs in those equations, but see [BKMM96] for more details on the specific form of the equations. We merely note that, if the kinetic energy metric is positive-definite, then the Lagrangian is hyperregular and our assumption of G-invariance of the exerted force F_E implies that

$$\ddot{r} = f(t, r, \dot{r}, \mathbb{I}_{\mathrm{loc}}^{-1} p) \tag{4.22}$$

for some function f which depends on the local trivialization. If the kinetic energy metric is not positive-definite (for use in toy examples; see the precise assumptions in §4.7.1, and the footnote there), then we *assume* that $\ddot{r}$ is given by (4.22).

4.7.3 Reduction in the Stokesian limit

From the definitions (4.12), (4.14) of $\mathbb{I}_{\mathrm{loc}}, \mathbb{V}_{\mathrm{loc}}$, we see that we may define $\overline{\mathbb{I}}_{\mathrm{loc}}, \overline{\mathbb{V}}_{\mathrm{loc}}$ by

$$\mathbb{I}_{\mathrm{loc}}(r) =: m\overline{\mathbb{I}}_{\mathrm{loc}}(r) \qquad \mathbb{V}_{\mathrm{loc}}(r) =: c\overline{\mathbb{V}}_{\mathrm{loc}}(r).$$

Defining the parameter $\epsilon := \frac{m}{c}$ and multiplying both sides of (4.21) by $\mathbb{I}_{\mathrm{loc}}\mathbb{V}_{\mathrm{loc}}^{-1}$, we obtain the rewritten equations of motion

$$\begin{aligned} g^{-1}\dot{g} &= -A_{\mathrm{mech}} \cdot \dot{r} + \frac{1}{m}\overline{\mathbb{I}}_{\mathrm{loc}}^{-1} p \\ \epsilon \overline{\mathbb{I}}_{\mathrm{loc}} \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \dot{p} &= m\overline{\mathbb{I}}_{\mathrm{loc}}(A_{\mathrm{visc}} - A_{\mathrm{mech}}) \cdot \dot{r} + p + \epsilon \overline{\mathbb{I}}_{\mathrm{loc}} \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \mathrm{ad}^*_{\mathbb{I}_{\mathrm{loc}}^{-1} p} p - \epsilon \overline{\mathbb{I}}_{\mathrm{loc}} \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \mathrm{ad}^*_{A_{\mathrm{mech}} \cdot \dot{r}} p. \end{aligned} \tag{4.23}$$

In considering the limit in which viscous forces dominate the inertia of the locomoting body, [KM96] formally set $\epsilon = 0$ in (4.23) to obtain $p = m\overline{\mathbb{I}}_{\mathrm{loc}}(A_{\mathrm{mech}} - A_{\mathrm{visc}}) \cdot \dot{r}$ from the second equation. Substituting this into the first equation of (4.23), they derive the following form of the equations of motion:

$$g^{-1}\dot{g} = -A_{\text{visc}} \cdot \dot{r}. \tag{4.24}$$

In the language of differential geometry, (4.24) states that in the Stokesian limit trajectories are *horizontal* with respect to the viscous connection (Definition C.16). We will see in the next section that this reduction can be extended away from the $\epsilon \to 0$ limit.

4.8 Reduction in the perturbed Stokes regime

In [EJ16], the argument of [KM96] was explained in more detail using the theory of normally hyperbolic invariant manifolds (NHIMs) in the context of geometric singular perturbation theory [Fen79, Jon95, Kap99]. The idea is to show that for $\epsilon > 0$ sufficiently small, the dynamics (4.23) possess an exponentially attractive invariant *slow manifold* M_ϵ, such that the dynamics restricted to M_ϵ approach (4.24) as $\epsilon \to 0$. This situation is depicted in Figure 4.1. We give an alternative argument which yields a result differing from that of [EJ16] in two ways.

1. [EJ16] give an argument for general mechanical systems without symmetry under the assumption that the configuration space Q is compact, although they do indicate that compactness can be replaced with uniformity conditions using noncompact NHIM theory [Eld13]. Our argument assumes symmetry but allows G to be noncompact, though we do require that $S \coloneqq Q/G$ be compact. This enables application of our result to locomotion systems with noncompact symmetry groups, such as the Euclidean group of planar rigid motions $\mathsf{SE}(2)$.

2. [EJ16] consider the limit $m \to 0$ while holding c and the force exerted by the locomoting body fixed. This makes sense, because if the exerted force were held fixed while taking

$c \to \infty$, then trivial dynamics would result in the singular limit: the system would not move at all. Rather than holding the exerted force fixed, we will consider the differential equation prescribing the *dynamics* of the shape variable to be fixed[10]. Under this assumption, we show that the dynamics depend only on the *ratio* $\epsilon = m/c$, and in particular the dynamics obtained in the two singular limits $m \to 0$ and $c \to \infty$ are the same.

Before stating Theorem IV.9, we need the following definition.

Definition IV.7 (C_b^k time-dependent vector fields)**.** Let M be a compact manifold with boundary, and let $f\colon \mathbb{R} \times M \to \mathsf{T}M$ a $C^{k\geq 0}$ time-dependent vector field. Let $(U_i)_{i=1}^n$ be a finite open cover of M and $(V_i, \psi_i)_{i=1}^n$ be a finite atlas for M such that $\overline{U}_i \subset V_i$ for all i, and for each i define $f_i \coloneqq (\mathsf{D}\psi_i \circ f \circ (\mathrm{id}_{\mathbb{R}} \times \psi_i^{-1}))$. We define an associated C^k norm $\|f\|_k$ of f via

$$\|f\|_k \coloneqq \max_{1\leq i\leq n} \max_{\substack{0\leq j\leq k \\ x\in\psi_i(\overline{U}_i)}} \|\mathsf{D}^j f_i(x)\|, \tag{4.25}$$

where $\|\mathsf{D}^j f_i(x)\|$ denotes the norm of a k-linear map; here $\mathsf{D}^j f$ includes partial derivatives with respect to time as well as the spatial variables. If $\|f\| < \infty$, we say that f is C^k-bounded and write $f \in C_b^k$. The norm $\|\cdot\|_k$ makes the C_b^k time-dependent vector fields into a Banach space. The norms induced by any two such finite covers of M are equivalent, and thereby induce a canonical C_b^k *topology* on the space of C_b^k time-dependent vector fields.

Remark IV.8. Definition IV.7 defines the C_b^k topology on the space of C_b^k time-dependent vector fields on a compact manifold. As discussed in [Eld13, Sec. 1.7], this C_b^k topology is finer than the C^k weak Whitney topology and coarser than the C^k strong Whitney topology [Hir94, Ch. 2], but all of these topologies induce the same topology on the subspace of time-independent vector fields due to compactness. Definition IV.7 is a special case of the

[10]This implicitly assumes that the locomoting body is capable of exerting $\mathcal{O}(c)$ forces.

definition in [Eld13, Ch. 2] for the C_b^k topology on C_b^k vector fields on Riemannian manifolds of bounded geometry, and on C_b^k maps between such manifolds.

The following theorem concerns a G-symmetric dynamical system on $\mathsf{T}Q$ whose equations of motion are consistent with our assumptions so far: i.e., they are given in local trivializations by (4.23) and an equation of the form (4.22).

Theorem IV.9. *Assume that $S = Q/G$ is compact. Let $2 \leq k < \infty$, and let X^ϵ be a C^k family of G-symmetric time-dependent vector fields on $\mathsf{T}Q$ with the following properties:*

1. *For every compact neighborhood with C^k boundary $K_0 \subset \mathsf{T}Q$ and $\epsilon > 0$, $X^\epsilon|_{\mathbb{R}\times K_0} \in C_b^k$ (Definition IV.7).*
2. *There exists a compact connected neighborhood $K \subset \mathsf{T}S$ of the zero section of $\mathsf{T}S$ with C^k boundary, such that $N := \mathsf{D}\pi_Q^{-1}(K) \subset \mathsf{T}Q$ is positively invariant for X^ϵ, for all sufficiently small $\epsilon > 0$.*
3. *X^ϵ is given in each local trivialization $\mathsf{T}(U \times G)$, where U is a chart for S, by (4.22) and (4.23):*

$$\begin{aligned} \ddot{r} &= f\left(t, r, \dot{r}, \frac{1}{m}\overline{\mathbb{I}}_{\text{loc}}^{-1} p\right) \\ \epsilon \overline{\mathbb{I}}_{\text{loc}} \overline{\mathbb{V}}_{\text{loc}}^{-1} \dot{p} &= m \overline{\mathbb{I}}_{\text{loc}} (A_{\text{visc}} - A_{\text{mech}}) \cdot \dot{r} + p + \epsilon \overline{\mathbb{I}}_{\text{loc}} \overline{\mathbb{V}}_{\text{loc}}^{-1} \text{ad}^*_{\overline{\mathbb{I}}_{\text{loc}}^{-1} p} p - \epsilon \overline{\mathbb{I}}_{\text{loc}} \overline{\mathbb{V}}_{\text{loc}}^{-1} \text{ad}^*_{A_{\text{mech}} \cdot \dot{r}} p \\ g^{-1}\dot{g} &= -A_{\text{mech}} \cdot \dot{r} + \frac{1}{m} \overline{\mathbb{I}}_{\text{loc}}^{-1} p \end{aligned} \tag{4.26}$$

   for some function f which depends on the local trivialization but is independent of ϵ.

Then for all sufficiently small $\epsilon > 0$, there exists a C^k normally hyperbolic invariant manifold with boundary $M_\epsilon \subset \mathbb{R} \times N \subset \mathbb{R} \times \mathsf{T}Q$ for the extended dynamics given by the extended vector field $(1, X_\epsilon)$ on $\mathbb{R} \times \mathsf{T}Q$. Additionally, M_ϵ is uniformly (in time and space) globally asymptotically stable and uniformly locally exponentially stable (with respect to the distance *induced by any complete G-invariant Riemannian metric on $\mathsf{T}Q$) for the extended dynamics restricted to $\mathbb{R} \times N$. Finally, there exists $\epsilon_0 > 0$ such that, for each local trivialization $U \times G$, there exists a C^k map $h_\epsilon \colon \mathbb{R} \times (\mathsf{T}U \cap K) \times (0, \epsilon_0) \to \mathfrak{g}^*$ such that $M_\epsilon \cap \mathsf{D}\pi_Q^{-1}(\mathsf{T}U \cap K)$ corresponds to*

$$\{(t, r, \dot{r}, p, g) : p = h_\epsilon(t, r, \dot{r}, \epsilon)\}, \tag{4.27}$$

$$h_\epsilon(t, r, \dot{r}, \epsilon) = \mathbb{I}_{\text{loc}}\left[(A_{\text{mech}}(r) - A_{\text{visc}}(r)) \cdot \dot{r} + \mathcal{O}(\epsilon)\right]$$

(with p defined by (4.19)), and h_ϵ together with its partial derivatives of order k or less are bounded uniformly in time. If $f(t, r, \dot{r}, \mathbb{I}_{\text{loc}}^{-1} p)$ is independent of t, then h_ϵ and M_ϵ are independent of t, and M_ϵ can be interpreted as a NHIM for the (non-extended) dynamics restricted to N.

Remark IV.10. Note that even if we assume $f \in C^\infty$, we can generally only obtain C^k NHIMs M_ϵ for k finite. This is because we obtain M_ϵ as a perturbation of a NHIM M_0, and perturbations of C^∞ NHIMs are generally only finitely smooth because the maximum perturbation size ϵ required to obtain degree of smoothness k for M_ϵ generally depends on k in such a way that $\epsilon \to 0$ as $k \to \infty$. See [Eld13, Rem. 1.12] and [vS79] for more discussion.

Remark IV.11. By replacing compactness of Q/G with uniformity conditions, it should be possible to generalize Theorem IV.9 to the situation of Q noncompact where either Q/G is noncompact, or where there is no symmetry at all. This was pointed out in [EJ16, App. 1]. This observation seems important for the consideration of dissipative mechanical systems which are only *approximately* invariant under a symmetry group G, which seems to be a more realistic assumption.

Remark IV.12. By taking $\epsilon \to 0$ in Theorem IV.9, we find that $p = \mathbb{I}_{\text{loc}}(A_{\text{mech}} - A_{\text{visc}}) \cdot \dot{r}$ in the limit. Substituting this into the first equation of (4.29), we obtain Equation (4.21) as in [KM96].

Proof.

Preparation of the equations of motion. Throughout the proof, we consider the dynamics in local trivializations of the form $U \times G$ for Q, where U is the domain of a chart for S, so that we have induced coordinates $(r, \dot{r})$ for $\mathsf{T}U$. In such a local trivialization we would like to use (4.26) to analyze the dynamics, but there are two (related) problems with this. First, the definition of p depends on m, and this will cause difficulties in verifying Definition IV.7 to check that certain vector fields are close in the C_b^k topology. Second, we would like to analyze (4.26) in a singular perturbation framework, but this is difficult to do directly because m explicitly appears, and the size of m may or may not be commensurate with the size of ϵ. To remedy this situation, we change variables via the diffeomorphism $(r, \dot{r}, p, g) \mapsto (r, \dot{r}, \Omega, g)$ of $\mathsf{T}U \times \mathfrak{g}^* \times G \to \mathsf{T}U \times \mathfrak{g} \times G$ where $\Omega \in \mathfrak{g}$ is defined by

$$\Omega := \mathbb{I}_{\text{loc}}^{-1} p = \operatorname{Ad}_{g^{-1}} \Gamma_{\text{mech}}(\dot{g}, \dot{r}) = g^{-1}\dot{g} + A_{\text{mech}} \cdot \dot{r}. \tag{4.28}$$

Sometimes Ω is referred to as the *(body) locked angular velocity* [BKMM96, p. 61]. Differentiating $\mathbb{I}_{\text{loc}}\Omega = p$, using (4.26), and rearranging yields

$$\begin{aligned}
\dot{t} &= 1 \\
\dot{r} &= v \\
\dot{v} &= f(t, r, v, \Omega) \\
\epsilon\dot{\Omega} &= -\epsilon \overline{\mathbb{I}}_{\text{loc}}^{-1} \left(\frac{d}{dt}\overline{\mathbb{I}}_{\text{loc}}\right)\Omega + \overline{\mathbb{I}}_{\text{loc}}^{-1}\overline{\mathbb{V}}_{\text{loc}}(A_{\text{visc}} - A_{\text{mech}}) \cdot v + \overline{\mathbb{I}}_{\text{loc}}^{-1}\overline{\mathbb{V}}_{\text{loc}}\Omega + \epsilon\overline{\mathbb{I}}_{\text{loc}}^{-1}\operatorname{ad}^*_{g^{-1}\dot{g}}\overline{\mathbb{I}}_{\text{loc}}\Omega,
\end{aligned} \tag{4.29}$$

where we have introduced the variable $v := \dot{r}$. We have written $\operatorname{ad}^*_{g^{-1}\dot{g}}$ for space reasons, but note that the $\dot{\Omega}$ equation is independent of g since

$$g^{-1}\dot{g} = -A_{\text{mech}} \cdot \dot{r} + \Omega, \tag{4.30}$$

and this implies that $\mathrm{ad}^*_{g^{-1}\dot{g}} = \mathrm{ad}^*_\Omega - \mathrm{ad}^*_{A_{\mathrm{mech}}\cdot\dot{r}}$. We see that (4.29) is split into slow (t, r, v) and fast (Ω) variables, which is the appropriate setup for a singular perturbation analysis. The remainder of the proof consists of two parts: (i) proving that the NHIM M_ϵ exists, and (ii) establishing the stability properties of M_ϵ.

Proof that M_ϵ exists. Introducing the "fast time" $\tau \coloneqq \frac{1}{\epsilon}t$ and denoting a derivative with respect to τ by a prime, after the time-rescaling we obtain the regularized equations

$$
\begin{aligned}
t' &= \epsilon \\
r' &= \epsilon v \\
v' &= \epsilon f(t, r, v, \Omega) \\
\Omega' &= -\epsilon \overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\left(\frac{d}{dt}\overline{\mathbb{I}}_{\mathrm{loc}}\right)\Omega + \overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\overline{\mathbb{V}}_{\mathrm{loc}}(A_{\mathrm{visc}} - A_{\mathrm{mech}})\cdot v + \overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\overline{\mathbb{V}}_{\mathrm{loc}}\Omega + \epsilon\overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\mathrm{ad}^*_{g^{-1}\dot{g}}\overline{\mathbb{I}}_{\mathrm{loc}}\Omega.
\end{aligned}
\tag{4.31}
$$

This rescaling of time is equivalent to replacing the vector field $(1, X_\epsilon)$ on $\mathbb{R}\times\mathsf{T}Q$ by $(\epsilon, \epsilon X_\epsilon)$. We see from (4.30) and (4.31) that there is a well-defined C^k time-dependent vector field $\tilde{X}_0$ given by the pointwise limit $\tilde{X}_0 \coloneqq \lim_{\epsilon\to 0}\epsilon X_\epsilon$. Given any G-symmetric time-dependent vector field Y on $\mathsf{T}Q$, we let Y/G denote the corresponding reduced vector field on $(\mathsf{T}Q)/G$. Hence (4.31) shows that the extended vector field $(1, \tilde{X}_0/G)$ has a smooth embedded submanifold (M_0/G) of critical points whose intersection with a locally trivializable neighborhood is given by

$$
\{(r, v, \Omega) \in \mathsf{T}U \times \mathfrak{g} : \Omega = (A_{\mathrm{mech}} - A_{\mathrm{visc}})\cdot v\}, \tag{4.32}
$$

and it is readily seen that M_0/G is described globally as the quotient of the Ehresmann connection $M_0 \coloneqq \ker\Gamma_{\mathrm{visc}}$ by the lifted action of G on $\mathsf{T}Q$.

Furthermore, M_0/G is a globally exponentially stable NHIM for the $\epsilon = 0$ system. To see this, first note that in any local trivialization t, r, v are constants when $\epsilon = 0$, and hence Ω' is of the form $\Omega' = \overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\overline{\mathbb{V}}_{\mathrm{loc}}\Omega + b$ for a constant b, and therefore has a globally exponentially stable equilibrium provided that all eigenvalues of $\overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\overline{\mathbb{V}}_{\mathrm{loc}}$ have negative real part. To see that this is the case, fix a basis of $\mathfrak{g}$ and corresponding dual basis for $\mathfrak{g}^*$, and first consider the product $\mathbb{I}^{-1}\mathbb{V}$. With respect to our chosen basis, $\mathbb{I}, \mathbb{V}$ and their inverses $\mathbb{I}^{-1}, \mathbb{V}^{-1}$ are respectively represented by r-dependent matrices I_{ij}, V_{ij} and their inverses I^{ij}, V^{ij}. It is immediate from the definitions (4.12) and (4.14) that I_{ij} and V_{ij} are respectively positive definite and negative definite symmetric matrices (this is why we required the bilinear forms k, ν to be positive definite when restricted to vectors tangent to G orbits). Since I_{ij} is symmetric positive definite, we may let $(\sqrt{I})_{ij}$ be a matrix square root of I_{ij} and let $(\sqrt{I})^{ij}$ be its inverse. But then the product $I^{ik}V_{kj}$ is similar to the symmetric negative definite matrix $(\sqrt{I})^{ik}V_{k\ell}(\sqrt{I})^{\ell j}$ (Einstein summation implied). Hence $\mathbb{I}^{-1}\mathbb{V}$ has only eigenvalues with negative real part, and the same is true of $\mathbb{I}_{\mathrm{loc}}^{-1}\mathbb{V}_{\mathrm{loc}}$ because of the similarity $\mathbb{I}_{\mathrm{loc}}^{-1}\mathbb{V}_{\mathrm{loc}} = \mathrm{Ad}_g^{-1}\mathbb{I}^{-1}\mathbb{V}\mathrm{Ad}_g$.

Let $\tilde{\pi}\colon (\mathsf{T}Q)/G \to \mathsf{T}S$ denote the projection induced by $\mathsf{D}\pi_Q$. Equation (4.32) implies that M_0/G is the image of a section $\sigma_0\colon \mathsf{T}S \to (\mathsf{T}Q)/G$ of $\tilde{\pi}$. Hence $(M_0/G) \cap \tilde{\pi}^{-1}(K) = \sigma_0(K)$ is compact, and M_0/G intersects $\tilde{\pi}^{-1}(\partial K)$ transversely. Furthermore, the assumption that $X^\epsilon|_{\mathbb{R}\times K_0} \in C_b^k$ for any compact neighborhood with C^k boundary $K_0 \subset \mathsf{T}Q$ implies that all partial derivatives of f are bounded on compact sets uniformly in time. This makes it clear that for any compact $K_1 \subset (\mathsf{T}Q)/G$, $(\epsilon X_\epsilon/G)|_{\mathbb{R}\times K_1}$ can be made arbitrarily close to $(\tilde{X}_0/G)|_{\mathbb{R}\times K_1}$ in the C_b^k topology (Definition IV.7) by taking $\epsilon > 0$ sufficiently small. Hence by the noncompact NHIM results of [Eld13, Sec. 4.1-4.2], it follows that $(M_0/G) \cap \tilde{\pi}^{-1}(K)$ persists in extended state space $\mathbb{R} \times N$ to a nearby attracting NHIM[11] M_ϵ/G with boundary for $(\epsilon, \epsilon X_\epsilon/G)$. Furthermore, M_ϵ/G is the image of a section $\sigma_\epsilon\colon \mathbb{R}\times K \to (\mathsf{T}Q)/G$ of $\tilde{\pi}$, and is given in each local trivialization of $(\mathsf{T}Q)/G$ by the graph of a function $\Omega = \tilde{h}_\epsilon(t, r, \dot{r}, \epsilon)$ which is C^k bounded uniformly in time. By symmetry, the preimage $M_\epsilon = \pi_{\mathsf{T}Q}^{-1}(M_\epsilon/G)$ of

[11] M_ϵ/G is unique up to the choice of a cutoff function used to modify the dynamics near the boundary of a slightly enlarged neighborhood of $\tilde{\pi}^{-1}(K)$, used in order to render a slightly enlarged version of $(M_0/G)\cap \tilde{\pi}^{-1}(K)$ overflowing invariant [Eld13, Sec. 4.3]. See [EKR18, Sec. 5] and [Jos00, Sec. 2] for more details on such boundary modifications.

M_ϵ/G via the quotient $\pi_{\mathsf{T}Q}\colon \mathsf{T}Q \to (\mathsf{T}Q)/G$ yields a NHIM M_ϵ for $(\epsilon, \epsilon X_\epsilon)$ (and hence also for $(1, X_\epsilon)$) on the subset $\mathbb{R} \times N$ of $\mathbb{R} \times \mathsf{T}Q$, and M_ϵ is given in each local trivialization by the graph of the same function $\Omega = \tilde{h}_\epsilon$ as M_ϵ/G but augmented with trivial dependence on g. The function h_ϵ from the theorem statement is given by $h_\epsilon = \mathbb{I}_{\text{loc}}\tilde{h}_\epsilon$.

Proof of the stability properties of M_ϵ. Fix any complete G-invariant Riemannian metric on[12] $\mathsf{T}Q$, so that it descends to a metric on $(\mathsf{T}Q)/G$ making $\pi_{\mathsf{T}Q}\colon \mathsf{T}Q \to (\mathsf{T}Q)/G$ into a Riemannian submersion [dC92, p. 185]. We have distance functions $d_{\mathsf{T}Q}$ and $\tilde{d}$ on $\mathsf{T}Q$ and $(\mathsf{T}Q)/G$ induced by these metrics. For $t \in \mathbb{R}$, we let $M_\epsilon(t) \coloneqq M_\epsilon \cap (\{t\} \times N)$ and $M_\epsilon(t)/G \coloneqq \pi_{\mathsf{T}Q}(M_\epsilon(t))$. Given $w \in \mathsf{T}Q$ and its orbit $\pi_{\mathsf{T}Q}(w) \in (\mathsf{T}Q)/G$, it follows that for all[13] $t \in \mathbb{R}$, $d(w, M_\epsilon(t)) = d(\pi_{\mathsf{T}Q}(w), M_\epsilon(t)/G)$. Hence it suffices to prove that M_ϵ/G is uniformly globally asymptotically stable and locally exponentially stable for the vector field $(1, X_\epsilon/G)$ on $\mathbb{R} \times \tilde{\pi}^{-1}(K) = \mathbb{R} \times \pi_{\mathsf{T}Q}(N)$, and to do this it suffices to prove the same for $(\epsilon, \epsilon X_\epsilon/G)$.

Fixing an inner product $\langle \cdot\,, \cdot \rangle$ and associated norm $\|\cdot\|$ on $\mathfrak{g}$, we accomplish this in two steps. First, we show that there exists a compact neighborhood $K_0 \subset \pi_{\mathsf{T}Q}(N)$ of M_ϵ/G such that K_0 is positively invariant for the time-dependent flow of X_ϵ, and such that any other compact neighborhood $K_1 \subset \pi_{\mathsf{T}Q}(N)$ of M_ϵ/G flows into K_0 after some finite time depending on K_1 but independent of the initial time. Second, we show that all trajectories in K_0 converge to M_ϵ/G at a uniform exponential rate. To achieve this second step, we show that in the intersection of each local trivialization with K_0, $\|\Omega - \tilde{h}_\epsilon(t, r, v)\|$ decreases at an exponential rate. Since $(\mathsf{T}Q)/G$ is covered by finitely many local trivialization (by compactness of S), and since all Riemannian metrics are uniformly equivalent on compact

[12]For example, take the Sasaki metric on $\mathsf{T}Q$ induced by any complete G-invariant metric on Q.

[13]To prove this, first note that $\tilde{d}(w, M_\epsilon(t)) \leq d(\pi_{\mathsf{T}Q}(w), M_\epsilon(t)/G)$ because the length $\ell(\tilde{\gamma})$ of any curve $\tilde{\gamma}\colon [0,1] \to \mathsf{T}Q$ satisfies $\ell(\pi_{\mathsf{T}Q} \circ \tilde{\gamma}) \leq \ell(\tilde{\gamma})$. But if $\gamma : [0,1] \to (\mathsf{T}Q)/G$ is any curve joining $\pi_{\mathsf{T}Q}(w)$ to M_ϵ/G, then its horizontal lift $\tilde{\gamma}$ is a curve joining w to M_ϵ such that $\ell(\tilde{\gamma}) = \ell(\gamma)$. Taking the infimum over all such γ shows that $d(w, M_\epsilon(t)) = d(\pi_{\mathsf{T}Q}(w), M_\epsilon(t)/G)$.

sets[14], this will establish uniform exponential convergence of points in K_0 with respect to the distance induced by any Riemannian metric, and in particular the distance $\tilde{d}$.

Consider a local trivialization $U \times G$ of Q and the associated form (4.31) of the dynamics restricted to $\tilde{\pi}^{-1}(K \cap \mathsf{T}U)$. Differentiating $\|\Omega\|^2$ using the last equation of (4.31), it is easy to check that $\frac{d}{d\tau}\|\Omega\|^2 \to -\infty$ as $\|\Omega\|^2 \to \infty$, uniformly in (t, r, v, ϵ) for ϵ sufficiently small. (This follows from the negative definiteness of $\mathbb{I}_{\text{loc}}^{-1}\mathbb{V}_{\text{loc}}$ and the compactness of K.) Hence we see that there exists $k_0 > 0$ such that for all ϵ sufficiently small, $\frac{d}{d\tau}\|\Omega\|^2 \leq -1$ when $\|\Omega\|^2 \geq k_0^2$. Now k_0 depends on the local trivialization, but we can replace k_0 with the largest such constant selected from finitely many fixed local trivializations covering Q. Hence there exists a compact subset $K_0 \subset \pi_{\mathsf{T}Q}(N)$ given by $\{\|\Omega\| \leq k_0\}$ in each of these fixed local trivializations, such that K_0 is positively invariant for the time-dependent flow of X_ϵ and such that any other compact neighborhood $K_1 \subset \pi_{\mathsf{T}Q}(N)$ of M_ϵ/G flows into K_0 after some finite time independent of the initial time.

It remains only to establish the uniform exponential rate of convergence of trajectories in K_0 to M_ϵ. For each local trivialization $U \times G$ of Q, we define the translated variable $\tilde{\Omega} \coloneqq \Omega - \tilde{h}_\epsilon(t, r, v, \epsilon)$. Since M_ϵ/G is invariant, we must have $\tilde{\Omega}' = 0$ whenever $\tilde{\Omega} = 0$. Differentiating $\tilde{\Omega}$ using (4.31), we therefore find that

$$\begin{aligned}\tilde{\Omega}' &= \left[-\epsilon\overline{\mathbb{I}}_{\text{loc}}^{-1}\left(\frac{d}{dt}\overline{\mathbb{I}}_{\text{loc}}\right) + \epsilon\overline{\mathbb{I}}_{\text{loc}}^{-1}\text{ad}^*_{g^{-1}\dot{g}}\overline{\mathbb{I}}_{\text{loc}} + \epsilon\zeta(t, r, v, \tilde{\Omega}) + \overline{\mathbb{I}}_{\text{loc}}^{-1}\overline{\mathbb{V}}_{\text{loc}}\right]\tilde{\Omega} \\ &=: \left[\epsilon A(t, r, v, \tilde{\Omega}) + \overline{\mathbb{I}}_{\text{loc}}^{-1}\overline{\mathbb{V}}_{\text{loc}}(r)\right]\tilde{\Omega},\end{aligned} \tag{4.33}$$

since all of the terms which do not vanish when $\tilde{\Omega} = 0$ must cancel. Here ζ is defined via

[14]Let $\|\cdot\|, \|\cdot\|'$ denote the Finslers (norms) induced by two Riemannian metrics, and K_0 our compact set. Since all norms are equivalent on finite-dimensional vector spaces, we have that the restrictions of these norms to the tangent space of a single point x satisfy $\frac{1}{c(x)}\|\cdot\| \leq \|\cdot\|' \leq c(x)\|\cdot\|$. Defining $\bar{c} \coloneqq \sup_{x\in K_0} c(x)$, we obtain the uniform equivalence $\frac{1}{\bar{c}}\|\cdot\| \leq \|\cdot\|' \leq \bar{c}\|\cdot\|$ on all of K_0. If K_0 is a connected submanifold and we give it the restricted metrics, then by considering the lengths of curves in K_0 this implies the uniform bound $\frac{1}{\bar{c}}d \leq d' \leq \bar{c}d$ on the Riemannian distances between points in K_0 with respect to the restricted metrics.

Hadamard's lemma [Nes03, Lemma 2.8]:

$$\zeta(t,r,v,\tilde{\Omega}) \coloneqq \frac{\partial}{\partial v}\tilde{h}_\epsilon(t,r,v)\int_0^1 \frac{\partial}{\partial \Omega} f(t,r,v,\tilde{h}_\epsilon(t,r,v)+s\tilde{\Omega})\,ds, \tag{4.34}$$

so that $\zeta(t,r,v,\tilde{\Omega})\tilde{\Omega} = \tilde{h}_\epsilon(t,r,v)f(t,r,v,\tilde{h}_\epsilon+\tilde{\Omega})$. As previously mentioned, the C^k boundedness of X_ϵ on compact subsets of $\mathsf{T}Q$ implies that $\tilde{h}_\epsilon$, f, and their first k partial derivatives are uniformly bounded on sets of the form $\mathbb{R}\times K_2$ with K_2 compact. Hence whenever $\Omega \in K_0$ and $(r,v)\in U\cap K$, $\|A(t,r,v,\tilde{\Omega}\| \leq L$ for some constant L depending on the local trivialization; we replace L with the largest such constant chosen from finitely many local trivializations covering Q. Integrating both sides of (4.33), taking norms using the triangle inequality, and applying Grönwall's Lemma (Corollary E.3) therefore yields

$$\begin{aligned}\|\tilde{\Omega}(\tau)\| &\leq e^{-\lambda(\tau-\tau_0)}e^{\int_{\tau_0}^{\tau}\epsilon\|A(t(s),r(s),v(s),\tilde{\Omega}(s)\|\,ds}\|\tilde{\Omega}(\tau_0)\| \\ &\leq e^{[-\lambda+\epsilon L](\tau-\tau_0)}\|\tilde{\Omega}(\tau_0)\|.\end{aligned} \tag{4.35}$$

where $-\lambda<0$ is defined via $-\lambda \coloneqq \sup_{r\in S}\max\operatorname{spec}(\overline{\mathbb{I}}_{\mathrm{loc}}^{-1}\overline{\mathbb{V}}_{\mathrm{loc}}(r))$, and is strictly negative since S is compact. By the previous discussion, taking $\epsilon>0$ sufficiently small so that $-\lambda+\epsilon L<0$ completes the proof. □

Theorem IV.9 and Remark IV.12 show that, to zeroth order in ϵ, the dynamics restricted to the slow manifold M_ϵ are given by the viscous connection model (4.24). The following theorem shows that the dynamics restricted to M_ϵ can be explicitly computed to higher order in ϵ. We compute the restricted dynamics to first order in ϵ. Higher order terms in ϵ can also be computed recursively, but we choose not to pursue this here.

Theorem IV.13. *Assume the same hypotheses as in Theorem IV.9. Then the dynamics* *restricted to the slow manifold M_ϵ are given in a local trivialization by*

$$g^{-1}\dot{g} = -A_{\text{visc}} \cdot \dot{r} + \epsilon \overline{\mathbb{V}}_{\text{loc}}^{-1} \left(\left(\frac{\partial}{\partial_r} \overline{h}_0 \right) \dot{r} + \left(\frac{\partial}{\partial \dot{r}} \overline{h}_0 \right) \ddot{r} - \operatorname{ad}^*_{g^{-1}\dot{g}}(\overline{h}_0) \right) + \mathcal{O}(\epsilon^2), \tag{4.36}$$

where

$$\overline{h}_0(r, \dot{r}) \coloneqq \frac{1}{m} h_0(r, \dot{r}) = \overline{\mathbb{I}}_{\text{loc}}(A_{\text{mech}}(r) - A_{\text{visc}}(r)) \cdot \dot{r},$$

where we are using the definition $\overline{\mathbb{I}}_{\text{loc}} \coloneqq \frac{1}{m}\mathbb{I}_{\text{loc}}$. Alternatively, we may write

$$g^{-1}\dot{g} = -A_{\text{visc}} \cdot \dot{r} + \epsilon \overline{\mathbb{V}}_{\text{loc}}^{-1} \left(\left(\frac{\partial}{\partial r} \overline{h}_0 \right) \dot{r} + \left(\frac{\partial}{\partial \dot{r}} \overline{h}_0 \right) f(t, r, \dot{r}, \overline{\mathbb{I}}_{\text{loc}}^{-1} \overline{h}_0) - \operatorname{ad}^*_{g^{-1}\dot{g}}(\overline{h}_0) \right) + \mathcal{O}(\epsilon^2), \tag{4.37}$$

for a different $\mathcal{O}(\epsilon^2)$ term.

Remark IV.14. Notice the presence, in the second term of (4.36), of $\overline{h}_0$ rather than h_0 of (4.27). This is important because the expression for h_0 contains an $\mathbb{I}_{\text{loc}} = m\overline{\mathbb{I}}_{\text{loc}}$ factor. Because of the possibility that the size of m is commensurate with ϵ, this means that h_0 could be $\mathcal{O}(\epsilon)$. However, $\overline{h}_0$ is $\mathcal{O}(1)$, ensuring that the second term is $\mathcal{O}(\epsilon)$ but not $\mathcal{O}(\epsilon^2)$.

Remark IV.15. Equations (4.36) and (4.37) can be viewed as adding $\mathcal{O}(\epsilon)$ correction terms to the viscous connection model (4.24), valid in the limit $\epsilon \to 0$, to account for the more realistic situation that the inertia-damping ratio $\frac{m}{c} = \epsilon$ is small but nonzero.

Proof of Theorem IV.13. Consider the function

$$\tilde{h}_\epsilon(t, r, \dot{r}, \epsilon) \coloneqq \mathbb{I}_{\text{loc}}^{-1} h_\epsilon = (A_{\text{mech}}(r) - A_{\text{visc}}(r)) \cdot \dot{r} + \mathcal{O}(\epsilon)$$

from the proof of Theorem IV.9, and define $\overline{h}_\epsilon \coloneqq \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_\epsilon = \frac{1}{m} h_\epsilon$. Since $\overline{h}_\epsilon, \tilde{h}_\epsilon \in C^k$, we may expand them as asymptotic series

$$\begin{aligned}\overline{h}_\epsilon &= \overline{h}_0 + \epsilon \overline{h}_1 + \ldots + \epsilon^k \overline{h}_k + \mathcal{O}(\epsilon^{k+1}) \\ \tilde{h}_\epsilon &= \tilde{h}_0 + \epsilon \tilde{h}_1 + \ldots + \epsilon^k \tilde{h}_k + \mathcal{O}(\epsilon^{k+1}),\end{aligned} \tag{4.38}$$

where for all i, $\overline{h}_i = \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_i$. We also already know from Theorem IV.9 that $\tilde{h}_0 = (A_{\text{mech}} - A_{\text{visc}}) \cdot \dot{r}$, and therefore $\tilde{h}_0(t, r, \dot{r}) \equiv \tilde{h}_0(r, \dot{r})$ has no explicit t-dependence. We now compute $\tilde{h}_1$ via a standard technique [Jon95]. Differentiating both sides of the equation $\Omega = \tilde{h}_\epsilon(t, r, \dot{r}, \epsilon)$ with respect to time (using (4.29) to differentiate the left hand side), substituting the second equation of (4.38) for Ω in the resulting expression, and retaining terms only up to $\mathcal{O}(\epsilon)$ we obtain

$$-\epsilon \overline{\mathbb{I}}_{\text{loc}}^{-1} \left(\frac{d}{dt} \overline{\mathbb{I}}_{\text{loc}} \right) \tilde{h}_0 + \overline{\mathbb{I}}_{\text{loc}}^{-1} \overline{\mathbb{V}}_{\text{loc}} (A_{\text{visc}} - A_{\text{mech}}) \cdot \dot{r} + \overline{\mathbb{I}}_{\text{loc}}^{-1} \overline{\mathbb{V}}_{\text{loc}} \left(\tilde{h}_0 + \epsilon \tilde{h}_1 \right) + \epsilon \overline{\mathbb{I}}_{\text{loc}}^{-1} \text{ad}^*_{g^{-1}\dot{g}} \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_0 = \epsilon \dot{\tilde{h}}_0 + \mathcal{O}(\epsilon^2).$$

Equating the coefficients of ϵ yields

$$\begin{aligned}\tilde{h}_1 &= \overline{\mathbb{V}}_{\text{loc}}^{-1} \left(\frac{d}{dt} \overline{\mathbb{I}}_{\text{loc}} \right) \tilde{h}_0 + \overline{\mathbb{V}}_{\text{loc}}^{-1} \overline{\mathbb{I}}_{\text{loc}} \dot{\tilde{h}}_0 - \overline{\mathbb{V}}_{\text{loc}}^{-1} \text{ad}^*_{g^{-1}\dot{g}} \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_0 \\ &= \overline{\mathbb{V}}_{\text{loc}}^{-1} \frac{d}{dt} \left(\overline{\mathbb{I}}_{\text{loc}} \tilde{h}_0 \right) - \overline{\mathbb{V}}_{\text{loc}}^{-1} \text{ad}^*_{g^{-1}\dot{g}} \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_0.\end{aligned}$$

Since $h_1 = \mathbb{I}_{\text{loc}} \tilde{h}_1$ and $\overline{h}_0 = \overline{\mathbb{I}}_{\text{loc}} \tilde{h}_0$, we find

$$h_1 = \mathbb{I}_{\text{loc}} \overline{\mathbb{V}}_{\text{loc}}^{-1} \frac{d}{dt} \left(\overline{h}_0 \right) - \mathbb{I}_{\text{loc}} \overline{\mathbb{V}}_{\text{loc}}^{-1} \text{ad}^*_{g^{-1}\dot{g}} \left(\overline{h}_0 \right), \tag{4.39}$$

and therefore (substituting $\ddot{r} = f(t, r, \dot{r}, \mathbb{I}_{\text{loc}}^{-1} p) = f(t, r, \dot{r}, \tilde{h}_0) + \mathcal{O}(\epsilon)$ and differentiating $\overline{h}_0(r, \dot{r})$

via the chain rule),

$$
\begin{aligned}
h_\epsilon(t,r,\dot r,\epsilon) &= \mathbb{I}_{\mathrm{loc}}(A_{\mathrm{mech}} - A_{\mathrm{visc}}) \cdot \dot r \\
&\quad + \epsilon \mathbb{I}_{\mathrm{loc}} \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \left(\left(\frac{\partial}{\partial r} \overline{h}_0 \right) \dot r + \left(\frac{\partial}{\partial \dot r} \overline{h}_0 \right) f(t,r,\dot r,\tilde h_0) - \mathrm{ad}^*_{g^{-1}\dot g}(\overline{h}_0) \right) + \mathbb{I}_{\mathrm{loc}} \mathcal{O}(\epsilon^2).
\end{aligned}
\tag{4.40}
$$

Notice that, since $\tilde h_0$ is a function of $r, \dot r$ only, the $\mathcal{O}(\epsilon)$ portion of the right hand side of (4.40) is a function of $t, r, \dot r$ alone and not p. This is required since h_ϵ is required to be a function of $t, r, \dot r, \epsilon$ alone, and is the reason that we needed to replace $\ddot r$ by $f(t,r,\dot r,\tilde h_0)$ in the $\mathcal{O}(\epsilon)$ term. Substituting (4.40) into the first equation of (4.23) yields Equation (4.37). Finally, making the substitution $f(t,r,\dot r,\tilde h_0) = \ddot r + \mathcal{O}(\epsilon)$ in Equation (4.37) yields Equation (4.36). □

The following theorem makes clearer the functional form of the dynamics (4.36), and it removes the $g^{-1}\dot g$ dependence of the right hand side of (4.36).

Theorem IV.1′. *Assume the hypotheses of Theorem IV.9. For sufficiently small $\epsilon > 0$, then for each local trivialization there exist smooth fields of linear maps $B(r)$ and $(1,2)$ tensors $G(r)$ such that the dynamics restricted to the slow manifold M_ϵ in the local trivialization satisfy*

$$
g^{-1}\dot g = -A_{\mathrm{visc}}(r) \cdot \dot r + \epsilon B(r) \cdot \ddot r + \epsilon G(r) \cdot (\dot r, \dot r) + \mathcal{O}(\epsilon^2). \tag{4.41}
$$

Remark IV.16. The (1,2) tensors $G(r)$ are *not* generally symmetric, which is clear from Equation (4.43) below.

Proof. Using the properties of ad^*, we may write $\mathrm{ad}^*_{g^{-1}\dot g}(\overline{h}_0) = (C \cdot \overline{h}_0) \cdot (g^{-1}\dot g)$ for an appropriate (r-independent) linear map $C \colon \mathfrak{g}^* \to \mathrm{End}(\mathfrak{g})$, and hence we may rewrite (4.36) as

$$
(\mathrm{id}_{\mathfrak{g}} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1}(C \cdot \overline{h}_0)) \cdot (g^{-1}\dot g) = -A_{\mathrm{visc}} \cdot \dot r + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \left(\left(\frac{\partial}{\partial r} \overline{h}_0 \right) \dot r + \left(\frac{\partial}{\partial \dot r} \overline{h}_0 \right) \ddot r \right) + \mathcal{O}(\epsilon^2).
$$

For sufficiently small ϵ, we may use the identity

$$(\mathrm{id}_{\mathfrak{g}} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1}(C \cdot \overline{h}_0))^{-1} = \mathrm{id}_{\mathfrak{g}} - \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1}(C \cdot \overline{h}_0) + \mathcal{O}(\epsilon^2)$$

to obtain

$$g^{-1}\dot{g} = -A_{\mathrm{visc}} \cdot \dot{r} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1}(C \cdot \overline{h}_0) \cdot A_{\mathrm{visc}} \cdot \dot{r} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \left(\frac{\partial}{\partial r} \overline{h}_0 \right) \dot{r} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \left(\frac{\partial}{\partial \dot{r}} \overline{h}_0 \right) \ddot{r} + \mathcal{O}(\epsilon^2). \quad (4.42)$$

Since $\overline{h}_0(r, \dot{r}) = \overline{\mathbb{I}}_{\mathrm{loc}}(r)(A_{\mathrm{mech}}(r) - A_{\mathrm{visc}}(r)) \cdot \dot{r}$ is linear in $\dot{r}$, it follows that the second and third terms are bilinear in $\dot{r}$, and the fourth term is linear in $\ddot{r}$. Hence we may take $B(r) \coloneqq \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \left(\frac{\partial}{\partial \dot{r}} \overline{h}_0 \right)$ and

$$G(r) \cdot (\dot{r}, \dot{r}) \coloneqq \overline{\mathbb{V}}_{\mathrm{loc}}^{-1}(C \cdot \mathbb{I}_{\mathrm{loc}}(A_{\mathrm{mech}} - A_{\mathrm{visc}}) \cdot \dot{r}) \cdot A_{\mathrm{visc}} \cdot \dot{r} + \epsilon \overline{\mathbb{V}}_{\mathrm{loc}}^{-1} \frac{\partial}{\partial r} (\mathbb{I}_{\mathrm{loc}}(A_{\mathrm{mech}} - A_{\mathrm{visc}}) \cdot \dot{r}) \cdot \dot{r}. \quad (4.43)$$

□

APPENDICES

APPENDIX A

Smoothness of linear parallel transport covering an inflowing invariant manifold

In this appendix, we show that a C^r flow on an inflowing invariant manifold $M \subset Q$ can always be lifted to a C^r linear flow on E, where $\pi\colon E \to M$ is any C^r subbundle of $TQ|_M$. For the definition of a fiber metric [KN63, p. 116] see Def. C.12 in Appendix C.

Lemma A.1. *Let M be a C^r inflowing invariant submanifold of $Q \in C^\infty$ for the flow Φ^t generated by a C^r vector field. Let $\pi\colon E \to M$ be a C^r subbundle of $TQ|_M$ equipped with any fiber metric g. Then there exists a C^r fiber metric h on E arbitrarily close to g and a C^r flow Π^t on E such that for all $t > 0$, $\Pi^t \coloneqq \Pi(t, \cdot\,)$ is an isometry with respect to the fiber metric h, covering $\Phi^t|_M$.*

Proof. We define a C^r submanifold (without boundary) M_ϵ by the formula $M_\epsilon \coloneqq \Phi^{-\epsilon}(\operatorname{int} M)$, with $\operatorname{int} M$ denoting the manifold interior of M. Because M is inflowing invariant, $M \subset M_\epsilon$. We extend E arbitrarily to a C^r subbundle $E_\epsilon \supset E$ of $TQ|_{M_\epsilon}$. In [PT77, App. 1] it is shown that M_ϵ has a compatible C^{r+1} differentiable structure with respect to which the vector field f restricted to M_ϵ is[1] C^r.

[1] The theorem in [PT77, App. 1] is stated for a C^1 invariant manifold and C^1 vector field, but the same

Denote M_ϵ with this C^{r+1} structure by $\widetilde{M}_\epsilon$, and let $I\colon \widetilde{M}_\epsilon \to M_\epsilon$ be the C^r diffeomorphism which is the identity map when viewed as a map of sets. Thus the pullback bundle I^*E_ϵ is a C^r vector bundle over $\widetilde{M}_\epsilon$ which is C^r isomorphic to E_ϵ via a vector bundle isomorphism $G_1\colon I^*E_\epsilon \to E_\epsilon$ covering I [Hir94, p. 97]. Furthermore, a standard argument using a universal bundle shows that there exists a C^{r+1} vector bundle $\widetilde{I^*E_\epsilon}$ over $\widetilde{M}_\epsilon$ and a C^r vector bundle isomorphism $G_2\colon \widetilde{I^*E_\epsilon} \to I^*E_\epsilon$ covering the identity [Hir94, p. 101, Thm 3.5]. This situation is depicted in the following diagram.

$$\begin{array}{ccccc} \widetilde{I^*E_\epsilon} & \xrightarrow{G_2} & I^*E_\epsilon & \xrightarrow{G_1} & E_\epsilon \\ \downarrow P & & \downarrow P & & \downarrow \pi \\ \widetilde{M}_\epsilon & \xrightarrow{\mathrm{id}_{\widetilde{M}_\epsilon}} & \widetilde{M}_\epsilon & \xrightarrow{I} & M_\epsilon \end{array} \tag{A.1}$$

Now pull back the fiber metric g on E_ϵ to $\widetilde{I^*E_\epsilon}$. That is, define $\tilde{g}$ through

$$\tilde{g}(v,w) = G^*(g)(v,w) = g(G(v), G(w)),$$

where $G = G_1 \circ G_2$. Now choose a C^{r+1} fiber metric $\tilde{h}$ on $\widetilde{I^*E_\epsilon}$ that is close to $\tilde{g}$. Let $\widetilde{\nabla}\colon \Gamma(\mathsf{T}\widetilde{M}_\epsilon \otimes \widetilde{I^*E_\epsilon}) \to \Gamma(\widetilde{I^*E_\epsilon})$ be a C^r affine connection compatible with the metric $\tilde{h}$ [KN63, Chap. 3]. Then the map of parallel transport along solution curves of f, $\widetilde{\Pi}^t\colon \widetilde{I^*E_\epsilon} \to \widetilde{I^*E_\epsilon}$, is an isometry since $\widetilde{\nabla}$ is compatible with $\tilde{h}$, and it is C^r because with respect to local coordinates $x_1, \ldots, x_{n_m}$ and any local frame $(\sigma_1, \ldots, \sigma_{n_s})$, the parallel transport equation takes the form

$$\sum_k \left(\frac{d}{dt} v^k \circ \Phi^t(x) + \sum_{i,j} \Gamma^k_{i,j}(v^i f^j) \circ \Phi^t(x) \right) \sigma_k \circ \Phi^t(x) = 0, \tag{A.2}$$

proof works, mutatis mutandis, for a *locally* invariant C^r manifold and C^r vector field, which is our situation here.

where the Christoffel symbols $\Gamma^k_{i,j}$ defined by

$$\widetilde{\nabla}_{\frac{\partial}{\partial x^i}} \sigma_j = \sum_k \Gamma^k_{i,j} \sigma_k$$

are C^r functions $\Gamma^k_{i,j} \colon \widetilde{M}_\epsilon \to \mathbb{R}$. Since f is a C^r vector field with respect to the smooth structure of $\widetilde{M}_\epsilon$, it follows that (A.2) defines a C^r ODE for v in local coordinates. The ODE theorems on existence, uniqueness, and smooth dependence on parameters imply that the solution to (A.2) depends smoothly on $x, v(x)$, and t. Thus $\widetilde{\Pi} \colon \mathbb{R} \times \widetilde{I^* E_\epsilon} \to \widetilde{I^* E_\epsilon}$ is indeed C^r.

Next, define the fiber metric h on E_ϵ by setting $h \coloneqq (G_2^{-1} \circ G_1^{-1})^* \tilde{h}$ and define $\Pi \colon \mathbb{R} \times E_\epsilon \to E_\epsilon$ via

$$\Pi^t(v) \coloneqq \Pi(t, v) \coloneqq (G_1 \circ G_2) \circ \widetilde{\Pi}^t \circ (G_1 \circ G_2)^{-1}(v). \tag{A.3}$$

Since $\tilde{h}$ was arbitrarily close to $\tilde{g}$, the same holds for h and g. The map Π is C^r because it is the composition of smooth functions. For any $t \in \mathbb{R}$, $\widetilde{\Pi}^t$ is an isometry of $(\widetilde{I^* E_\epsilon}, \tilde{h})$, and our choice of the pullback metric h on E_ϵ implies that $G_1 \circ G_2$ is an isometry into (E_ϵ, h). Thus for any $t \in \mathbb{R}$, Π^t is a composition of vector bundle isometries and is thus an isometry of vector bundles, hence preserves h. By construction Π^t covers $\Phi^t|_{M_\epsilon}$.

Now M is positively invariant under Φ^t since M is inflowing invariant, hence also E is positively invariant under Π^t. We therefore obtain a well-defined restriction of Π^t to $M \subset M_\epsilon$ and also restrict h to E, completing the proof. □

APPENDIX B

Inflowing NAIMs: Reduction to the boundaryless case

In this appendix, we prove a result which shows roughly that any compact inflowing NAIM can always be viewed as a subset of a compact boundaryless NAIM. In particular, this result allows the application of various linearization theorems from the literature [PS70, BK94, HPS77, Sak94, Rob71, PT77, Sel84, Sel83, BK94] to inflowing NAIMs as in Corollaries II.24 and II.25 and in §2.5, despite the fact that in the literature these theorems are formulated only for boundaryless invariant manifolds. We use this result in §2.6 to derive a linear normal form result for singular perturbation problems in which the critical manifold is a NAIM.

First, let us describe the intuition behind our construction. Let $M \subset Q$ be a compact, inflowing NAIM for some vector field on Q, and $N \supset M$ a slight extension along the backward flow. We rip a hole in our space Q by removing a small neighborhood U_0 of ∂N. Then we glue two copies of $Q \setminus U_0$ together at their boundaries (thought of as a "wormhole") creating a total space $\widehat{Q}$. We modify the copies of N slightly such that they connect through the wormhole as a smooth, compact submanifold $\widehat{N} \subset \widehat{Q}$. Finally, we carefully modify the vector field near the wormhole so that $\widehat{N}$ is a NAIM again for the modified vector field.

This procedure is made precise in the proof of Proposition B.1 below, but let us already introduce some more details using Figure B.1. A family of smooth tubular neighborhoods $U_0 \subset \ldots \subset U_3$ of ∂N are chosen so that the vector field f points inward at each $N \setminus U_i$, and so that each $W^s_{\text{loc}}(M) \cap U_i = \varnothing$. We smoothly rescale f inside U_3 to create a vector field $\tilde{f}$ such that $\tilde{f}$ is zero on $\overline{U}_2$, and we smoothly approximate N inside U_2 to create a submanifold $\widetilde{N}$ such that $\widetilde{N} \cap U_1$ is a C^∞ submanifold. We next create a copy of Q, remove the subset U_0 from each copy to form two copies of $Q' := Q \setminus U_0$, and let $\widehat{Q}$ be the double of Q' obtained by glueing the two copies of Q' along $\partial Q' = \partial U_0$, forming a "wormhole" between the two spaces. Using a standard technique from differential topology, we give $\widehat{Q}$ a C^∞ differential structure such that $\widehat{N} \subset \widehat{Q}$ is a C^r submanifold, where $\widehat{N}$ is comprised of the two copies of $\widetilde{N}$ (this step is the reason why we needed to approximate N by $\widetilde{N}$). We give $\widehat{Q}$ a Riemannian metric which agrees with the original metric on each copy of Q' except on an arbitrarily small neighborhood of $\partial Q'$. The vector field $\hat{f}_0$, defined to be equal to $\tilde{f}$ on each copy of Q', is automatically C^r since it is zero on a neighborhood of $\partial Q'$. Finally, we modify $\hat{f}_0$ inside each copy of U_3 to create a vector field $\hat{f}$ on $\widehat{Q}$ such that $\widehat{N}$ is an r-NAIM for $\hat{f}$. We show that the resulting global stable foliation $\widehat{W}^s(M)$ for $\hat{f}$ over a copy of M agrees with the global stable foliation $W^s(M)$ for f, and that certain asymptotic rates for f are preserved by $\hat{f}$.

Proposition B.1. *Let $M, N \subset Q$ be compact inflowing r-NAIMs, with M a proper subset of the manifold interior of N, for the $C^{r\geq 1}$ flow Φ^t generated by the $C^{r\geq 1}$ vector field f on Q. Let U_0 be an arbitrarily small tubular neighborhood of ∂N, having smooth boundary ∂U_0 and disjoint from $W^s_{\text{loc}}(M)$. Define $\widehat{Q}$ to be the double of $Q \setminus U_0$.*

Then there exists a C^∞ differential structure on $\widehat{Q}$ and a C^r vector field $\hat{f} \colon \widehat{Q} \to T\widehat{Q}$ such that

1. *$\hat{f}$ is equal to f on each copy of $Q \setminus U_0$, except on an arbitrarily small neighborhood of*

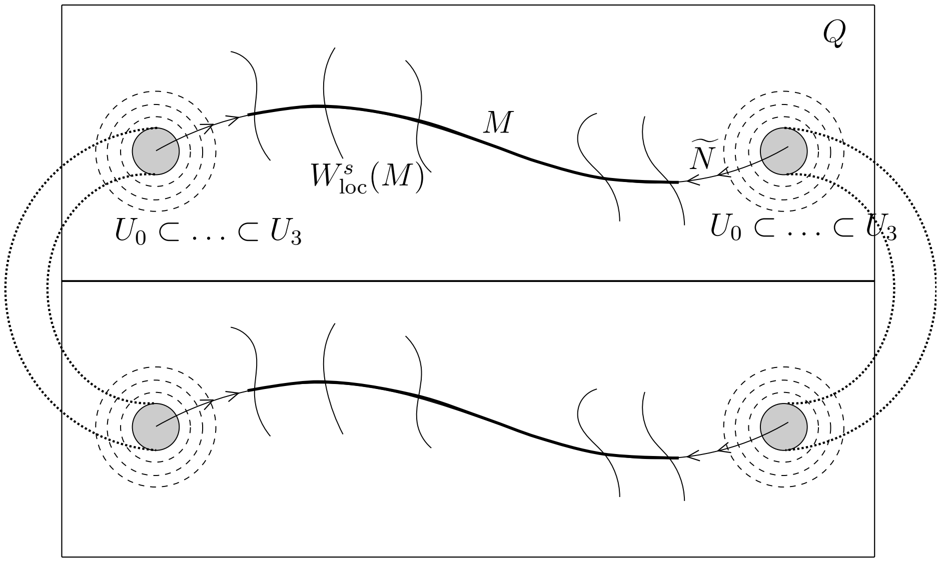


Figure B.1: A schematic figure of the constructions used in Prop. B.1.

∂U_0.

2. *There exists a compact and boundaryless r-NAIM $\widehat{N}$ for $\hat{f}$, with $\widehat{N}$ equal to N on each copy of $Q \setminus U_0$, except on an arbitrarily small neighborhood of ∂U_0.*

3. *The global stable foliation of M for f does not intersect U_0, and it coincides with the global stable foliation of M for $\hat{f}$, when M and $W^s(M)$ are identified via inclusion with subsets of a copy of $Q \setminus U_0$ in $\widehat{Q}$.*

Let $\widehat{\Phi}^t$ be the C^r flow generated by $\hat{f}$, and let $\widehat{E}^s$ be the $D\widehat{\Phi}^t|_{\widehat{N}}$-invariant stable vector bundle for the NAIM $\widehat{N}$. If, additionally, there exist constants $K > 0$ and $\alpha < 0$ such that for all $m \in M$, $t \geq 0$ and $0 \leq i \leq k$ the k-center bunching condition

$$\|\mathsf{D}\Phi^t|_{T_mM}\|^i \|\mathsf{D}\Phi^t|_{E^s_m}\| \leq Ke^{\alpha t} \llbracket \mathsf{D}\Phi^t|_{T_mM} \rrbracket \tag{B.1}$$

is satisfied for the original system on Q, then (B.1) will also be satisfied with M, E^s, and Φ^t replaced by $\widehat{N}, \widehat{E}^s$, and $\widehat{\Phi}^t$, and with α replaced by some different constant $\hat{\alpha} < 0$.

Similarly, if additionally there exist constants $0 < \delta < -\alpha < -\beta$ *and* $K \geq 1$ *such that for all* $t \geq 0$

$$K^{-1}e^{-\delta t} \leq \lfloor\!\lfloor \mathsf{D}\Phi^t|_{TM} \rfloor\!\rfloor \leq \|\mathsf{D}\Phi^t|_{TM}\| \leq Ke^{\delta t},$$
$$K^{-1}e^{-\delta t} \leq \lfloor\!\lfloor (\mathsf{D}\Phi^t|_{TM})^{-1} \rfloor\!\rfloor \leq \|(\mathsf{D}\Phi^t|_{TM})^{-1}\| \leq Ke^{\delta t}, \tag{B.2}$$
$$K^{-1}e^{\beta t} \leq \lfloor\!\lfloor \mathsf{D}\Phi^t|_{E^s} \rfloor\!\rfloor \leq \|\mathsf{D}\Phi^t|_{E^s}\| \leq Ke^{\alpha t}$$

uniformly on TM *and* E^s*, then we can choose* $\hat{f}$ *appropriately, such that the same will be true for* $\widehat{\Phi}^t$*,* $T\widehat{N}$*, and* $\widehat{E}^s$ *with modified constants* $0 < \hat{\delta} < -\hat{\alpha} < -\hat{\beta}$ *arbitrarily close to* δ, α, β*.*

Remark B.2. It is not an additional hypothesis to require the existence of the manifold N in Proposition B.1. This is because given any compact inflowing NAIM M, then for any sufficiently small $\epsilon > 0$, $N := \Phi^{-\epsilon}(M)$ will be a compact inflowing NAIM containing M. We mention N explicitly only to highlight the fact that *any* compact inflowing NAIM N containing M in its manifold interior will do.

Proof. Let $\epsilon > 0$ be any small positive number, and let U_0, U_1, U_2, U_3 be arbitrarily small tubular neighborhoods of ∂N, disjoint from $W^s_{\text{loc}}(M)$, satisfying $\partial N \subset U_0 \subset \overline{U}_0 \subset U_1 \subset \overline{U}_1 \subset U_2 \subset \overline{U}_2 \subset U_3 \subset \Phi^{-\epsilon}(W^s_{\text{loc}}(N))$, and such that all U_i have C^∞ boundary ∂U_i. See Figure B.1. Since N is inflowing, we may further construct the U_i so that f is strictly inward pointing at the boundary of each $N \setminus U_i$, and so that all points in U_3 leave U_3 in a uniformly finite time.

It follows that $W^s(M) \cap U_3 = \varnothing$. We have chosen U_3 to be disjoint from $W^s_{\text{loc}}(M)$, so to see this, suppose that there exists $m \in M$ and $y \in (W^s(m) \setminus W^s_{\text{loc}}(M)) \cap U_3$. Since $\Phi^\epsilon(U_3) \subset W^s_{\text{loc}}(N)$ by construction, by continuity there exists $t_0 > 0$ such that $\Phi^{t_0}(y) \in W^s_{\text{loc}}(N) \setminus W^s_{\text{loc}}(M)$. Let $n \in N \setminus M$ be the unique point such that $\Phi^{t_0}(y) \in W^s_{\text{loc}}(n)$. Since $y \in W^s(m)$ and since $W^s(M)$ is Φ^t-invariant, it follows also that $\Phi^{t_0}(y) \in W^s(\Phi^{t_0}(m))$. By uniqueness of the stable fibers, this implies that $n = \Phi^{t_0}(m)$. But $n \in N \setminus M$ and

$\Phi^{t_0}(m) \in M$ by positive invariance of M, so we have obtained a contradiction.

Rescaling f with a smooth cutoff function supported in $Q \setminus U_2$ and identically equal to 1 on $Q \setminus U_3$, we replace f with a C^r vector field $\tilde{f}$ which is equal to f on $Q \setminus U_3$ and zero on $\overline{U}_2$. By continuity of Df, $\overline{U}_2$ consists entirely of nonhyperbolic critical points for $\tilde{f}$.

We next approximate N by a C^r manifold $\widetilde{N}$ such that $\widetilde{N}$ coincides with N on $Q \setminus U_2$ and such that $\widetilde{N} \cap U_1$ is a C^∞ submanifold intersecting ∂U_0 transversely. This can be achieved by giving N a C^∞ differential structure and then approximating the inclusion $N \hookrightarrow Q$ relative to $N \setminus U_2$ in the C^r topology, see [Hir94, Ch. 2] for approximation theory details. The fact that $\tilde{f}|_{\overline{U_2}} = 0$ implies that $\tilde{N}$ is invariant under $\tilde{f}$.

We define $Q' := Q \setminus U_0$, a C^∞ manifold with boundary $\partial Q' = \partial U_0$. Recall that a C^∞ collar for $\partial Q'$ in Q' is a C^∞ embedding $h\colon \partial Q' \times [0, \infty) \to Q'$ such that $h(x, 0) \equiv x$ [Hir94, p. 113]. We choose a C^∞ collar h for $\partial Q'$ which restricts to a collar of $\partial \widetilde{N}$ in $\widetilde{N}$, i.e., $h|_{\partial \widetilde{N} \times [0,\infty)} \to \widetilde{N}$ is a collar [Hir94, Thm 6.2]. Now let $\widehat{Q}$ be the double of Q', the topological space obtained by first forming the disjoint union of two copies of Q', then identifying corresponding points in $\partial Q'$. We use the collar h to henceforth endow $\widehat{Q}$, in the usual way, with a C^∞ differential structure (see, e.g., [Hir94, p. 184] or [Lee13, p. 226]), and we let S denote the common image of $\partial Q'$ in $\widehat{Q}$.

Let $\widehat{N}$ denote the image of $\widetilde{N}$ in $\widehat{Q}$. Since h was chosen to restrict to a collar for $\partial \widetilde{N}$ in $\widetilde{N}$, it follows that $\widehat{N}$ is a C^r submanifold of $\widehat{Q}$. Letting $\hat{f}_0$ be the vector field on $\widehat{Q}$ which is equal to $\tilde{f}$ on each copy of Q', it is immediate that $\hat{f}_0 \in C^r$ since $\hat{f}_0$ is zero on a neighborhood of $\partial Q'$. Finally, using a partition of unity, we give $\widehat{Q}$ a C^∞ Riemannian metric which coincides with the original metric on each copy of Q', except on an arbitrarily small neighborhood of S.

Next, we modify $\hat{f}_0$ near S to make $\widehat{N}$ normally attracting. Let X be a C^∞ manifold which is C^1-close to $\widehat{N}$. Let $\varphi\colon E' \to \widehat{Q}$ be a C^∞ tubular neighborhood of X. I.e., $\pi'\colon E' \to X$ is a C^∞ vector bundle and φ is an open C^∞ embedding with $\varphi|_X$ the inclusion map, identifying

X with the zero section of E'. If X approximates $\widehat{N}$ sufficiently closely, then $\varphi^{-1}(\widehat{N})$ is the image of a C^r section $h\colon X \to E'$. Let $V_2 \subset \widehat{Q}$ denote the open set which is the image of the two copies of U_2 in $\widehat{Q}$, and define V_3 similarly. Let $\chi\colon E' \to [0,\infty)$ be a C^∞ compactly supported bump function such that $\chi \equiv 1$ on $\overline{V_2}$ and $\operatorname{supp}\chi \subset V_3$. We define a C^r vector field $\hat{f}$ on $\varphi(E')$ by

$$\hat{f} \circ \varphi(v_x) := \mathsf{D}\varphi_{v_x}\left[(\varphi^* \hat{f}_0)(v_x) - \rho\chi(v_x)(v_x - h(x))\right],$$

where $\pi'(v_x) = x$, $\varphi^*\hat{f}_0 := (\varphi^{-1})_*\hat{f}_0$, and $\rho := (\alpha+\beta)/2$ if (B.2) holds and $\rho := 1$ otherwise. Since χ is compactly supported, it follows that $\hat{f}(v_x)$ is equal to $f(v_x)$ for sufficiently large $\|v_x\|$, hence we may extend $\hat{f}$ to a C^r vector field on $\widehat{Q}$ (still denoted $\hat{f}$) by defining $\hat{f}$ to be equal to f on $\widehat{Q} \setminus \varphi(E)$. Let $\widehat{\Phi}_1^t$ denote the flow of $\hat{f}$.

Define a subbundle E of $TQ|_{\widehat{N}}$ by $E := \mathsf{D}\varphi(\mathsf{V}E'|_{h(X)})$, where $\mathsf{V}E' := \ker \mathsf{D}\pi' \subset TE'$ is the vertical bundle. Since h is a section of E', it follows that $\mathsf{T}E'|_{h(X)} = \mathsf{T}h(X) \oplus \mathsf{V}E'|_{h(X)}$, and since φ is a local diffeomorphism it follows that $\mathsf{D}\varphi$ preserves this splitting: $\mathsf{T}\widehat{Q}|_{\widehat{N}} = T\widehat{N} \oplus E$. Let $\Pi^E\colon T\widehat{Q}|_{\widehat{N}} \to E$ be the projection $T\widehat{N} \oplus E \to E$. We now argue that $\widehat{N}$ is an r-NAIM for $\hat{f}$; it suffices to show that $\widehat{N}$ is an r-NAIM for the linear flow $\Pi^E \circ \mathsf{D}\widehat{\Phi}_1^t|_E$ [Fen71, Prop. 1, Thm 6]. To do this, by the Uniformity Lemma [Fen71] it suffices to show that for each $n \in \widehat{N}$ there exist $C_n > 0$ and $a_n < 0$ such that for any $t \geq 0$ and $0 \leq i \leq r$,

$$\|\Pi^E \circ \mathsf{D}\widehat{\Phi}_1^t|_{E_n}\| \leq C_n e^{a_n t} \llbracket \mathsf{D}\widehat{\Phi}_1^t|_{\mathsf{T}_n\widehat{N}} \rrbracket^i. \tag{B.3}$$

First note that $\hat{f}$ is equal to f on $\widehat{Q} \setminus V_3$, $\widehat{N} \setminus V_3$ is positively invariant, and N is an r-NAIM for f. It follows that for each $n \in \widehat{N} \setminus V_3$, we can find a_n and C_n such that (B.3) holds. Next, let $n \in \widehat{N}$ be any point with $\hat{f}_0(n) \neq 0$. Since $\widehat{\Phi}_1^t$ takes n into $\widehat{N} \setminus V_3$ in finite time, in this case we can also find a_n, C_n such that (B.3) holds. Finally, if $n \in \widehat{N}$ is any point with

$\hat{f}_0(n) = 0$, then E_n is invariant under $\Pi^E \circ \mathsf{D}\widehat{\Phi}_1^t|_{E_n}$. The definition of $\hat{f}$ and E imply that n is an exponentially stable fixed point for the restriction of this flow to E_n, so we again find a_n, C_n such that (B.3) holds. Hence, by the Uniformity Lemma, $\widehat{N}$ is indeed an r-NAIM — in particular, there exists a $\mathsf{D}\widehat{\Phi}_1^t$-invariant stable bundle $\widehat{E}_1^s$ over $\widehat{N}$.

Now suppose additionally that either the k-center bunching conditions (B.1) or (B.2) held for the original system on Q. Considering now the flow $\mathsf{D}\widehat{\Phi}_1^t|_{\widehat{E}_1^s}$ on $\widehat{E}_1^s$ and repeating the argument in the preceding paragraph — using a different version of the Uniformity Lemma [Fen74, Lem. 16] for the case of center bunching conditions — shows that the corresponding condition still holds for $\hat{f}$ on $\widehat{Q}$.

It remains only to show that the global stable foliation $\widehat{W}^s(M)$ for $\hat{f}$ agrees with $W^s(M)$ when we identify M and $W^s(M)$ with either copy of their images in $\widehat{Q}$ — for definiteness, let us fix one such copy of M and $W^s(M)$ in what follows (with the former copy a subset of the latter). To accomplish this, we first consider the local foliations $\widehat{W}^s_{\text{loc}}(M)$ and $W^s_{\text{loc}}(M)$. Both local foliations are $\widehat{\Phi}^t$-invariant, the latter because $\hat{f}$ is equal to f on $W^s(M)$. But since M is a compact and inflowing NAIM, the standard Hadamard graph transform [Fen74] shows that there exists a *unique* local invariant foliation transverse to M. More precisely, this means that there exists a sufficiently small neighborhood J of M such that $\forall m \in M : J \cap W^s_{\text{loc}}(m) = J \cap \widehat{W}^s_{\text{loc}}(m)$. Now for any $m \in M$ and any $y \in W^s(m)$, there exists $t_0 > 0$ such that $\Phi^{t_0}(y) \in J \cap W^s_{\text{loc}}(\Phi^{t_0}(m)) = J \cap \widehat{W}^s_{\text{loc}}(\Phi^{t_0}(m))$. It follows that $y \in \Phi^{-t_0}(\widehat{W}^s_{\text{loc}}(\Phi^{t_0}(m))) \subset \widehat{W}^s(m)$ and therefore that $W^s(m) \subset \widehat{W}^s(m)$. A symmetric argument shows that $\widehat{W}^s(m) \subset W^s(m)$, and since $m \in M$ was arbitrary it follows that the leaves of $W^s(M)$ and $\widehat{W}^s(M)$ coincide. This completes the proof. □

APPENDIX C

Fiber bundles

In this appendix, we review the basic notions from the theory of fiber bundles that we use. Our definition of C^k fiber bundles follows [Nee10, Def. 1.1.1]. Other useful references for the topological and C^∞ cases include [Ste51, Hus66, Blo15], with a self-contained and brief introduction appearing in [Blo15, Ch. 2]. We also mention [Lee13, Hir94, KN63] as containing nice introductions to vector bundles, and [Blo15, KN63] as containing nice introductions to principal bundles.

Definition C.1 (Fiber bundles)**.** A C^k fiber bundle, with $1 \leq k \leq \infty$, is a quadruple (E, B, F, π) consisting of C^k manifolds E, B, and F and a C^k map $\pi \colon E \to B$ with the following property of local triviality: each point $b \in B$ has an open neighborhood $U \subset B$ for which there exists a C^k diffeomorphism

$$\varphi_U \colon \pi^{-1}(U) \to U \times F,$$

satisfying

$$\mathrm{pr}_1 \circ \varphi_U = \pi,$$

where $\mathrm{pr}_1 \colon U \times F \to U$ is the projection onto the first factor. A C^0 fiber bundle is defined by replacing all C^k manifolds and diffeomorphisms above with arbitrary topological spaces and homeomorphisms, respectively. Often we abuse terminology and simply refer to E or to $\pi \colon E \to B$ as the fiber bundle when the other data is understood.

Remark C.2. The following terminology is common. E is called the *total space*, B is called the *base space*, F is the *model fiber* or *fiber type*, and π is called the *bundle projection*. Sets of the form $E_b := \pi^{-1}(b)$ are called the *fibers* of the bundle or of π. The map φ_U is called a *local trivialization*. (E, B, F, π) is sometimes called an F*-bundle* over B.

Example C.3 (Disk bundles)**.** A C^k *disk bundle* is a C^k fiber bundle (E, B, F, π) with $F = \mathbb{R}^n$, for some $n \in \mathbb{N}$.

Definition C.4 (Vector bundles)**.** A (finite-dimensional) C^k *vector bundle* is a C^k disk bundle (E, B, F, π) with the following additional requirement. For any open sets $U, V \subset B$ with $U \cap V \neq \varnothing$, the *transition map*

$$\varphi_{U,V} := \varphi_U^{-1} \varphi_V|_{(U \cap V) \times F} \colon (U \cap V) \times F \to (U \cap V) \times F$$

is given by

$$\varphi_{U,V}(b, v) = (b, A(b)v),$$

where $A \colon U \cap V \to \mathrm{GL}(n, \mathbb{R})$ is a C^k invertible matrix-valued map.

Example C.5 (The tangent bundle)**.** Let Q be a smooth (C^∞) n-manifold. Then its tangent bundle $\pi \colon TQ \to Q$ is a smooth vector bundle. To see this, let (U, ψ_U) be a smooth chart for Q. Identifying $\mathsf{T}\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$, then $\varphi_U := (\pi, \mathrm{pr}_2 \circ \mathsf{D}\psi_U) \colon TQ|_U \to U \times \mathbb{R}^n$ satisfies $\mathrm{pr}_1 \circ \varphi_U = \pi$, where $TQ|_U := \pi^{-1}(U)$ and pr_i is projection onto the i-th factor. If (V, ψ_V) is another chart, then

$$\varphi_{U,V}\left(b, \sum_i v^k e_k\right) = \left(b, \sum_i \frac{\partial \psi^i_{U,V}}{\partial x^j} v^j e_i\right),$$

where (e_k) is the standard basis of $\mathbb{R}^n$ and $\psi_{U,V} \coloneqq \psi_U^{-1} \circ \psi_V|_{U\cap V}$. Hence we may take the Jacobian of $\psi_{U,V}$ to play the role of A from Definition C.4, so TQ is indeed a smooth vector bundle.

Definition C.6. A C^k isomorphism $\psi\colon E \to E'$ of C^k fiber bundles (E, B, F, π) and (E', B', F', π') covering a map $\rho\colon B \to B'$ is a C^k fiber-preserving diffeomorphism ψ (homeomorphism if $k = 0$): each fiber $\pi^{-1}(b)$ is bijectively mapped to the fiber $\pi'^{-1}(\rho(b))$. This is equivalent to requiring that ψ is a C^k diffeomorphism and

$$\pi' \circ \psi = \pi \circ \rho.$$

A C^k fiber bundle (E, B, F, π) is *trivial* as a C^k bundle if it is C^k isomorphic to the fiber bundle $(B \times F, B, F, \mathrm{pr}_2)$, where pr_2 is projection onto the second factor.

Example C.7. Let $\pi\colon E \to B$ be any C^k map, and assume there exists a fiber bundle (E', B, F, π') and a C^k diffeomorphism $\psi\colon E \to E'$ with the property that

$$\pi' \circ \psi = \mathrm{id}_B.$$

Then (E, B, F, π) is also a C^k fiber bundle. This is because if $\varphi_U\colon \pi'^{-1}(U) \to U \times F$ is a C^k local trivialization of $\pi'\colon E' \to B$, then $\varphi_U \circ \psi\colon \pi^{-1}(U) \to U \times F$ is a C^k local trivialization of $\pi\colon E \to B$. In particular, if (E', B, F, π') is a disk bundle, then so is (E, B, F, π). (recall that all vector bundles are disk bundles — we use this in the proof of Theorem II.5).

Definition C.8. A map $X\colon B \to E$ is a C^k *section* of a fiber bundle (E, B, F, π) if $X \in C^k$ and $\pi \circ X = \mathrm{id}_B$.

Example C.9. A C^k section X of the tangent bundle $\mathsf{T}Q$ is the same thing as a C^k vector field X on Q. The requirement $\pi \circ X = \mathrm{id}_Q$ simply means that $X(q)$ is a tangent vector based at $q \in Q$.

We next give definitions of vector subbundles and the Whitney sum of vector bundles. These concepts are fundamental to the very definition of normal hyperbolicity, see §2.3.

Definition C.10 (Vector subbundle)**.** A vector bundle (E', B, F, π) is a C^k *vector subbundle* or *linear subbundle* of a C^k vector bundle (E, B, F, π) if every $b \in B$ has a neighborhood U such that there exist pointwise linearly independent C^k sections $(X_1, \ldots, X_d)$ (called a *local frame*) which span E'_c for all $c \in U$.

Definition C.11 (Whitney sum)**.** The *Whitney sum* of two C^k vector bundles $(E, B, \mathbb{R}^n, \pi)$ and $(E', B, \mathbb{R}^m, \pi')$ is the C^k vector bundle $(E \oplus E', B, \mathbb{R}^{n+m}, \tilde{\pi})$ whose fiber $(E \oplus E')_b$ is given by $E_b \oplus E'_b$. The C^k vector bundle structure is determined as follows. If φ_U and φ'_U are local trivializations for E and E' over U, then $(\varphi_U, \mathrm{pr}_2 \circ \varphi'_U)$ is a local trivialization for $E \oplus E'$ over U.

Following [KN63, p. 116], we now give the definition of a *fiber metric* on a vector bundle, which generalizes the notion of a Riemannian metric on the tangent bundle of a manifold.

Definition C.12 (Fiber metric)**.** A C^k fiber metric on a C^k vector bundle $\pi \colon E \to M$ is an assignment, to each $m \in M$, of an inner product g_m on the fiber $\pi^{-1}(m)$, such that for any C^k sections $X, Y \colon M \to E$, the map $m \mapsto g_m(X(m), Y(m))$ is C^k.

Using a partition of unity, it easy to show that fiber metrics always exist on any vector bundle over a paracompact base. A fiber metric defines a norm on each fiber $\pi^{-1}(m)$ via $\|X\|_m \coloneqq \sqrt{g_m(X, X)}$ for $X \in E_m$. We will often suppress the subscript m and simply write $\|X\|$.

We next define *principal bundles*, for which we need only the smooth case. We define principal bundles such that the transition maps are given in terms of right multiplication by an element of the relevant Lie group (see below). The opposite convention seems to be more commonly used, but our convention applies to our needs in Chapter IV. See also [BKMM96], where the same convention is adopted for similar reasons.

Definition C.13 (Principal bundles)**.** Let G be a Lie group. A smooth principal G-bundle (E, B, G, π) is a smooth fiber bundle (E, B, G, π) with the following additional requirement. For any open sets $U, V \subset B$ with $U \cap V \neq \varnothing$, the transition map

$$\varphi_{U,V} \coloneqq \varphi_U^{-1}\varphi_V|_{(U\cap V)\times F} \colon (U \cap V) \times F \to (U \cap V) \times F$$

is given by

$$\varphi_{U,V}(b, g) = (b, gt(b)),$$

where $t\colon U \cap V \to G$ is a smooth G-valued.

Example C.14 (Principal bundles from group actions)**.** Let E be a smooth manifold and let $\theta\colon E \times G \to N$ be a smooth left action of the Lie group G on N. The group action is *free* if for any nonidentity group element $g \neq e$, $\theta_g\colon E \to E$ has no fixed points. The group action is *proper* if the map $\psi\colon E \times G \to E \times E$ defined by $\psi(x, g) = (\theta_g(x), x)$ is proper: for any compact $K \subset E \times E$, $\psi^{-1}(E)$ is compact [Lee13, Ch. 21].

Then the orbit space E/G is a topological manifold of dimension $\dim E - \dim G$, and has a unique smooth structure with the property that $\pi\colon E \to E/G$ is a smooth submersion [Lee13, Thm 21.10]. Furthermore, $(E, E/G, G, \pi)$ is a smooth principal G-bundle as in Definition C.13. Conversely, every principal bundle arises in this way, and therefore every principal bundle is equipped with a natural G action on its total space.

Definition C.15 (Ehresmann and principal connections)**.** Let (E, B, F, π) be a smooth fiber bundle. An *Ehresmann connection* is a vector subbundle H of $\mathsf{T}E$ with the property that

$$\mathsf{T}E = H \oplus \ker \mathsf{D}\pi. \tag{C.1}$$

Let (E, B, G, π) be a principal G-bundle. A *principal connection* is an Ehresmann connection

H with the additional property that

$$\forall g \in G : \forall x \in E : \mathsf{D}\theta_g H_x = H_{\theta_g(x)}, \tag{C.2}$$

where θ is the associated left G action on the principal bundle (see Example C.14). A principal connection is equivalently specified by a $\mathfrak{g}$-valued 1-form Γ on E satisfying the following two properties.

1. $\forall v \in \mathsf{T}E \colon \Gamma(\mathsf{D}\theta_g v) = \mathrm{Ad}_g \Gamma(v)$.

2. $\forall \xi \in \mathfrak{g} \colon \Gamma(\xi_E) = E$,

where $\xi_E(q) := \frac{\partial}{\partial_t}\theta_{\exp(t\xi)}(q)$ is the *infinitesimal generator* of $\xi \in \mathfrak{g}$ at q. The Ehresmann connection satisfying (C.2) is recovered as $H = \ker \Gamma$.

Note that the infinitesimal generator ξ_E as defined above is a smooth vector field on E.

Definition C.16 (Horizontal paths and lifts)**.** Let (E, B, F, π) be a fiber bundle equipped with an Ehresmann connection H. Let $I \subset \mathbb{R}$ be any interval. A smooth path $\sigma \colon I \to E$ is *horizontal* if

$$\forall t \in I : \dot{\sigma}(t) \in H_{\sigma(t)}. \tag{C.3}$$

Let $\gamma \colon I \to B$ be a smooth path in B, where $I \subset \mathbb{R}$ is an interval. A horizontal path $\tilde{\gamma} \colon I \to E$ is a *horizontal lift* of γ if $\pi \circ \tilde{\gamma} = \gamma$.

Remark C.17. The existence and uniqueness theorem from ordinary differential equations shows that, for any smooth path $\gamma \colon I \to B$ and any $x \in \pi^{-1}(\gamma(0))$, there exists a unique smooth horizontal lift $\tilde{\gamma} \colon I \to E$ with $\gamma(0) = x$.

Remark C.18. Let (E, B, G, π) be a principal G-bundle equipped with a principal connection H, and consider a smooth loop $\gamma \colon I \to B$ with $b = \gamma(0) = \gamma(1)$. Then there exists $g \in G$

such that for all horizontal lifts $\tilde{\gamma}\colon I \to E$ of γ, $\tilde{\gamma}(1) = \theta_g(\tilde{\gamma}(0))$, where θ is the natural left G action on E. The set of all such g for all such loops γ is called the *holonomy group* of H based at b.

APPENDIX D

Chebyshev's weak law of large numbers

In this appendix, we collect some results on the weak law of large numbers (LLN) which are essentially due to Chebyshev [Che99]. These results apply to sequences of random variables which are independent, but *not* necessarily identically distributed.

In what follows, we consider a probability space $(\Omega, \Sigma, \mathbb{P})$ comprising a set Ω, sigma algebra $\Sigma \subset 2^{\Omega}$, and probability measure $\mathbb{P}\colon \Sigma \to \mathbb{R}$. A *random variable* is a Borel-measurable function $X\colon \Omega \to \mathbb{R}^n$. The *expected value* or *mean* $\mathbb{E}[X]$ of of X is given by the Lebesgue integral

$$\mathbb{E}[X] \coloneqq \int_{\Omega} X \, d\mathbb{P}.$$

If $\mu = \mathbb{E}[X]$, then the *covariance* $\operatorname{cov}(X)$ of X is defined via

$$\operatorname{cov}(X) \coloneqq \mathbb{E}[(X-\mu)(X-\mu)^T]$$

and the *variance* $\operatorname{var}(X)$ of X is given by

$$\operatorname{var}(X) \coloneqq \operatorname{trace}(\operatorname{cov}(X)) = \mathbb{E}[\|X-\mu\|^2].$$

Given a sequence $X_1, X_2, \ldots$ of random variables, we define

$$\overline{X}_N \coloneqq \frac{1}{N} \sum_{i=1}^{N} X_i.$$

We now state the main theorem, which is a slight generalization of Chebyshev's weak law of large numbers. We then state some corollaries including Chebyshev's weak LLN, deferring the proof of Theorem D.1 to the end of the Appendix.

Theorem D.1. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables. Assume that*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \operatorname{var}(X_i) = 0.$$

Then

$$\forall \epsilon > 0\colon \lim_{N\to\infty} \mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) = 0.$$

Furthermore, we have the explicit rate of convergence

$$\mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) \leq \frac{1}{\epsilon^2 N^2} \sum_{i=1}^{N} \operatorname{var}(X_i).$$

Corollary D.2 (Chebyshev's weak LLN)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent random variables such that there exists $\sigma^2 > 0$ such that $\forall i\colon \operatorname{var}(X_i) \leq \sigma^2$. Then*

$$\forall \epsilon > 0\colon \lim_{N\to\infty} \mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) = 0.$$

Furthermore, we have the explicit rate of convergence

$$\mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2 N}.$$

Corollary D.3 (Approximate weak LLN)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent* *random variables. Assume that*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{var}(X_i) = 0,$$

and that there exists $\mu \in \mathbb{R}^n$ *and* $\delta > 0$ *such that*

$$\forall i : \|\mathbb{E}[X_i] - \mu\| < \delta.$$

Then

$$\forall \epsilon > 0\colon \lim_{N\to\infty} \mathbb{P}(\|\overline{X}_N - \mu\| > \epsilon + \delta) = 0.$$

Furthermore, we have the explicit rate of convergence

$$\mathbb{P}(\|\overline{X}_N - \mu]\| > \epsilon) \leq \frac{1}{\epsilon^2 N^2} \sum_{i=1}^{N} \mathrm{var}(X_i).$$

Proof of Corollary D.3. The triangle inequality yields for all N

$$\|\overline{X}_N - \mu\| - \delta \leq \|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| + \|\mathbb{E}[\overline{X}_N] - \mu\| - \delta < \|\overline{X}_N - \mathbb{E}[\overline{X}_N]\|.$$

Hence

$$\mathbb{P}(\|\overline{X}_N - \mu\| > \epsilon + \delta) < \mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) \leq \frac{1}{\epsilon^2 N^2} \sum_{i=1}^{N} \mathrm{var}(X_i),$$

with the second inequality following from Theorem D.1. For fixed $\epsilon > 0$, the right hand side tends to zero as $N \to \infty$ by assumption. □

We will use Chebyshev's inequality to prove Theorem D.1. For completeness, we state this result and give a proof.

Lemma D.4 (Extended Markov's inequality)**.** *Let* X *be a random variable with mean* μ*.*

Then for any $k, p > 0$,

$$\mathbb{P}(\|X - \mu\| > k) \leq \frac{\mathbb{E}[\|X - \mu\|^p]}{k^p}.$$

In particular, if $\sigma^2 = \operatorname{var}(X)$ then by taking $p = 2$ we obtain Chebyshev's inequality

$$\mathbb{P}(\|X - \mu\| > k) \leq \frac{\sigma^2}{k^2}.$$

Proof of Lemma D.4.

$$\mathbb{P}(\|X - \mu\| > k) = \int_{\{\|X-\mu\|^p/k^p > 1\}} 1\, d\mathbb{P} \leq \frac{1}{k^p} \int_\Omega \{\|X - \mu\|^p \, d\mathbb{P} = \frac{\mathbb{E}[\|X - \mu\|^p}{k^p}.$$

□

Proof of Theorem D.1. By Lemma D.4, we have

$$\mathbb{P}(\|\overline{X}_N - \mathbb{E}[\overline{X}_N]\| > \epsilon) \leq \frac{\operatorname{var}(\overline{X}_N)}{\epsilon^2}.$$

But $\operatorname{var}(\overline{X}_N) = \frac{1}{N^2} \sum_{i=1}^N \operatorname{var}(X_i)$ by independence of the X_i, and this quantity tends to zero by assumption. □

APPENDIX E

Grönwall's inequality

In this appendix, we present a reasonably general version of results due to Grönwall [Gro19] and Bellman [Bel43]. Theorem E.1 is similar to [Kha02], though we do not require continuity of the function α and we provide a different proof.

Theorem E.1. *Let J denote an interval of the real line of the form $[a,\infty)$ or $[a,b]$ or $[a,b)$, with $a<b$. Let u and β be continuous real-valued functions on J, and let α be a real-valued function which is integrable on every compact subinterval of J. Assume also that β is nonnegative, and that*

$$\forall t \in J : u(t) \leq \alpha(t) + \int_a^t \beta(s)u(s)\,ds, \tag{E.1}$$

Then

$$\forall t \in J : u(t) \leq \alpha(t) + \int_a^t \alpha(s)\beta(s)e^{\int_s^t \beta(r)\,dr}\,ds. \tag{E.2}$$

Remark E.2. No assumptions are needed on the signs of the functions α or u.

Corollary E.3. *Assume the hypotheses of Theorem E.1. Additionally assume that α is nondecreasing. Then*

$$u(t) \leq \alpha(t)e^{\int_a^t \beta(r)\,dr}. \tag{E.3}$$

Proof of Corollary E.3.

$$u(t) \leq \alpha(t) \left(1 + \int_a^t \beta(s) e^{\int_s^t \beta(r)\,dr}\,ds\right) = \alpha(t) \left(1 - \int_a^t \frac{d}{ds} e^{\int_s^t \beta(r)\,dr}\,ds\right) = \alpha(t) e^{\int_a^t \beta(r)\,dr}.$$

□

Proof of Theorem E.1. Define $I(t) \coloneqq \int_a^t \beta(s)u(s)\,ds$. Then $\dot{I} \coloneqq \frac{d}{dt}I = \beta u \leq \beta(\alpha + I)$, with the last inequality following since $\beta \geq 0$. (Here we used the continuity of u, β to invoke part 1 of the fundamental theorem of calculus, FTC-1 hereafter.) Hence

$$\dot{I} - \beta I \leq \alpha\beta.$$

Define the antiderivative $B(t) \coloneqq \int_a^t \beta(s)\,ds$. (Here continuity of β is again required, in anticipation of another invocation of FTC-1.) Multiplying both sides of the above inequality by the integrating factor $e^{-B(t)}$, we obtain

$$\frac{d}{dt}\left(e^{-B(t)} I(t)\right) \leq e^{-B(t)} \alpha(t) \beta(t).$$

Integrating both sides from a to t (this is where we require integrability of α) and using the fact that $B(a) = 0$, we obtain

$$I(t) \leq \int_a^t e^{B(t)-B(s)} \alpha(s)\beta(s)\,ds.$$

Since $u(t) \leq \alpha(t) + I(t)$, we obtain

$$u(t) \leq \alpha(t) + \int_a^t e^{B(t)-B(s)} \alpha(s)\beta(s)\,ds = \alpha(t) + \int_a^t \alpha(s)\beta(s) e^{\int_s^t \beta(r)\,dr}\,ds.$$

□

BIBLIOGRAPHY

BIBLIOGRAPHY

[AC14] M M Ankarali and N J Cowan, *System identification of rhythmic hybrid dynamical systems via discrete time harmonic transfer functions*, 53rd IEEE Conference on Decision and Control, IEEE, 2014, pp. 1017–1022.

[Arn74] L Arnold, *Stochastic differential equations: theory and applications*, 1 ed., John Wiley and Sons, 1974.

[Bel43] R Bellman, *The stability of solutions of linear differential equations*, Duke Mathematical Journal **10** (1943), no. 4, 643–647.

[BHR18] B Bittner, R L Hatton, and S Revzen, *Geometrically optimal gaits: a data-driven approach*, Nonlinear Dynamics (2018), 1–16.

[BK94] I U Bronstein and A Y Kopanskiĭ, *Smooth invariant manifolds and normal forms*, World Scientific Series on Nonlinear Science. Series A: Monographs and Treatises, vol. 7, World Scientific Publishing Co., Inc., River Edge, NJ, 1994. MR 1337026

[BKMM96] A M Bloch, P S Krishnaprasad, J E Marsden, and R M Murray, *Nonholonomic mechanical systems with symmetry*, Archive for Rational Mechanics and Analysis **136** (1996), no. 1, 21–99.

[Blo15] A M Bloch, *Nonholonomic mechanics and control*, 2 ed., vol. 24, Springer-Verlag, 2015.

[BLZ98] P W Bates, K Lu, and C Zeng, *Existence and persistence of invariant manifolds for semiflows in banach space*, vol. 645, American Mathematical Soc., 1998.

[Bre81] V N Brendelev, *On the realization of constraints in nonholonomic mechanics*, Journal of Applied Mathematics and Mechanics **45** (1981), no. 3, 351–355.

[Bru96] P Brunovskỳ, *Tracking invariant manifolds without differential forms*, Acta Math. Univ. Comenianae **65** (1996), no. 1, 23–32.

[Car82] J Carr, *Applications of centre manifold theory*, vol. 35, Springer Science & Business Media, 1982.

[Che99] P L Chebyshev, *Oeuvres de pl tchebychef*, vol. 1, Commissionaires de l'Académie impériale des sciences, 1899.

[Cox70] R G Cox, *The motion of long slender bodies in a viscous fluid part 1. general theory*, Journal of Fluid mechanics **44** (1970), no. 4, 791–810.

[CS88] H Cruse and W Schwarze, *Mechanisms of coupling between the ipsilateral legs of a walking insect (carausius morosus)*, J Exp Biol **138** (1988), 455–469.

[dC92] M P do Carmo, *Riemannian geometry*, 2 ed., Birkhäuser, 1992.

[DG17] X Da and J Grizzle, *Combining trajectory optimization, supervised machine learning, and model structure for mitigating the curse of dimensionality in the control of bipedal robots*, arXiv preprint arXiv:1711.02223 (2017).

[DHF+15] S M Danner, U S Hofstoetter, B Freundl, H Binder, W Mayr, F Rattay, and K Minassian, *Human spinal locomotor control is based on flexibly organized burst generators*, Brain **138** (2015), no. 3, 577–588.

[dR84] G de Rham, *Differentiable manifolds*, Springer-Verlag, 1984.

[Duf18] G Duffing, *Erzwungene schwingungen bei veränderlicher eigenfrequenz und ihre technische bedeutung*, no. 41-42, R, Vieweg & Sohn, 1918.

[EJ16] J Eldering and H O Jacobs, *The role of symmetry and dissipation in biolocomotion*, SIAM Journal on Applied Dynamical Systems **15** (2016), no. 1, 24–59.

[EKR18] J Eldering, M Kvalheim, and S Revzen, *Global linearization and fiber bundle structure of invariant manifolds*, Nonlinearity **31** (2018), no. 9, 4202–4245.

[Eld13] J Eldering, *Normally hyperbolic invariant manifolds: the noncompact case*, Atlantis Press, 2013.

[Eld16] ______, *Realizing nonholonomic dynamics as limit of friction forces*, Regular and Chaotic Dynamics **21** (2016), no. 4, 390–409.

[EP98] I R Epstein and J A Pojman, *An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos*, Oxford University Press, 1998.

[Erm96] B Ermentrout, *Type i membranes, phase resetting curves, and synchrony*, Neural computation **8** (1996), no. 5, 979–1001.

[Eva13] L C Evans, *An introduction to stochastic differential equations*, 1 ed., American Mathematical Society, 2013.

[Fed68] H Federer, *Geometric measure theory*, Springer-Verlag, 1968.

[Fen71] N Fenichel, *Persistence and smoothness of invariant manifolds for flows*, Indiana Univ. Math. J. **21** (1971), 193–226. MR 0287106

[Fen74] ———, *Asymptotic stability with rate conditions*, Indiana University Mathematics Journal **23** (1974), no. 12, 1109–1137.

[Fen77] ———, *Asymptotic stability with rate conditions, ii*, Indiana University Mathematics Journal **26** (1977), no. 1, 81–93.

[Fen79] ———, *Geometric singular perturbation theory for ordinary differential equations*, Journal of Differential Equations **31** (1979), no. 1, 53–98.

[Fri75] A Friedman, *Stochastic differential equations and applications*, 1 ed., vol. 1, Academic Press, 1975.

[Gar04] C W Gardiner, *Handbook of stochastic methods*, 3 ed., Springer-Verlag, 2004.

[GFCE92] J F Gibson, J D Farmer, M Casdagli, and S Eubank, *An analytic approach to practical state space reconstruction*, Physica D: Nonlinear Phenomena **57** (1992), no. 1–2, 1 – 30.

[GH83] J Guckenheimer and P Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Applied Mathematical Sciences, vol. 42, Springer-Verlag, New York, 1983. MR 709768

[GP10] V Guillemin and A Pollack, *Differential topology*, AMS Chelsea Publishing, Providence, RI, 2010, Reprint of the 1974 original. MR 2680546

[Gro19] T H Gronwall, *Note on the derivatives with respect to a parameter of the solutions of a system of differential equations*, Annals of Mathematics (1919), 292–296.

[Guc75] J M Guckenheimer, *Isochrons and phaseless sets*, Journal of Mathematical Biology **1** (1975), 259–273.

[GW02] S D Głazek and K G Wilson, *Limit cycles in quantum theories*, Physical review letters **89** (2002), no. 23, 230401.

[Hae61] A Haefliger, *Differentiable imbeddings*, Bulletin of the American Mathematical Society **67** (1961), no. 1, 109–112.

[Hal80] J K Hale, *Ordinary differential equations*, 2 ed., RE Krieger Publishing, New York, New York, 1980.

[HC11] R L Hatton and H Choset, *Geometric motion planning: The local connection, stokes theorem, and the importance of coordinate choice*, The International Journal of Robotics Research **30** (2011), no. 8, 988–1014.

[HC13] ———, *Geometric swimming at low and high reynolds numbers*, IEEE Transactions on Robotics **29** (2013), no. 3, 615–624.

[HdlL12] G Huguet and R de-la Llave, *Computation of limit cycles and their isochrones: fast algorithms and their convergence*, Tech. Report 2415, Institute for mathematics and its applications, University of Minnesota, December 2012.

[HG17] K A Hamed and J W Grizzle, *Reduced-order framework for exponential stabilization of periodic orbits on parameterized hybrid zero dynamics manifolds: Application to bipedal locomotion*, Nonlinear Analysis: Hybrid Systems **25** (2017), 227–245.

[HI97] F C Hoppensteadt and E M Izhikevich, *Weakly connected neural networks*, vol. 126, Springer-Verlag, 1997.

[Hir94] M W Hirsch, *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer-Verlag, New York, 1994, Corrected reprint of the 1976 original. MR 1336822

[HPS77] M W Hirsch, C C Pugh, and M Shub, *Invariant manifolds*, Lecture Notes in Mathematics, Vol. 583, Springer-Verlag, Berlin-New York, 1977. MR 0501173

[Hus66] D Husemoller, *Fibre bundles*, vol. 5, Springer, 1966.

[IM79] N Ikeda and S Manabe, *Integral of differential forms along the path of diffusion processes*, Publications of the Research Institute for Mathematical Sciences **15** (1979), no. 3, 827–852.

[JF02] D L Jindrich and R J Full, *Dynamic stabilization of rapid hexapedal locomotion*, J Exp Biol **205** (2002), no. 18, 2803–2823 (English).

[JK94] C K R T Jones and N Kopell, *Tracking invariant manifolds with differential forms in singularly perturbed systems*, Journal of Differential Equations **108** (1994), no. 1, 64–88.

[JKK96] C K R T Jones, T J Kaper, and N Kopell, *Tracking invariant manifolds up to exponentially small errors*, SIAM Journal on Mathematical Analysis **27** (1996), no. 2, 558–577.

[Jon95] C K R T Jones, *Geometric singular perturbation theory*, Dynamical systems, Springer, 1995, pp. 44–118.

[Jos00] K Josić, *Synchronization of chaotic systems and invariant manifolds*, Nonlinearity **13** (2000), no. 4, 1321.

[JT09] C K R T Jones and S-K Tin, *Generalized exchange lemmas and orbits heteroclinic to invariant manifolds*, Discrete and Continuous Dynamical Systems-Series S **2** (2009), no. 4, 967–1023.

[Kap99] T J Kaper, *An introduction to geometric methods and dynamical systems theory for singular perturbation problems*, Analyzing Multiscale Phenomena Using Singular Perturbation Methods: American Mathematical Society Short Course, January 5-6, 1998, Baltimore, Maryland **56** (1999), no. 56, 85.

[Kar81] A V Karapetian, *On realizing nonholonomic constraints by viscous friction forces and celtic stones stability*, Journal of Applied Mathematics and Mechanics **45** (1981), no. 1, 30–36.

[KBS14] K U Kristiansen, M Brøns, and J Starke, *An iterative method for the approximation of fibers in slow-fast systems*, SIAM Journal on Applied Dynamical Systems **13** (2014), no. 2, 861–900.

[Kha02] H K Khalil, *Nonlinear systems*, 3 ed., Prentice Hall, 2002.

[KJ01] T J Kaper and C K R T Jones, *A primer on the exchange lemma for fast-slow systems*, Multiple-time-scale dynamical systems, Springer, 2001, pp. 65–87.

[KM95] S D Kelly and R M Murray, *Geometric phases and robotic locomotion*, Journal of Field Robotics **12** (1995), no. 6, 417–431.

[KM96] ———, *The geometry and control of dissipative systems*, Decision and Control, 1996., Proceedings of the 35th IEEE Conference on, vol. 1, IEEE, 1996, pp. 981–986.

[KN63] S Kobayashi and K Nomizu, *Foundations of differential geometry. Vol I*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963. MR 0152974

[KOS76] P V Kokotovic, R E O'malley, and P Sannuti, *Singular perturbations and order reduction in control theory–an overview*, Automatica **12** (1976), no. 2, 123–132.

[Kue15] C Kuehn, *Multiple time scale dynamics*, Springer, 2015.

[KY90] E J Kostelich and J A Yorke, *Noise reduction: Finding the simplest dynamical system consistent with the data*, Phys D: Nonlinear Phenom **41** (1990), 183–196.

[Lam93] S H Lam, *Using csp to understand complex chemical kinetics*, Combustion Science and Technology **89** (1993), no. 5-6, 375–404.

[Lee13] J M Lee, *Introduction to smooth manifolds*, 2 ed., Springer-Verlag, 2013.

[LG89] S H Lam and D A Goussis, *Understanding complex chemical kinetics with computational singular perturbation*, Symposium (International) on Combustion, vol. 22, Elsevier, 1989, pp. 931–941.

[LG94] ______, *The csp method for simplifying kinetics*, International Journal of Chemical Kinetics **26** (1994), no. 4, 461–486.

[Liu06] W Liu, *Geometric singular perturbations for multiple turning points: invariant manifolds and exchange lemmas*, Journal of Dynamics and Differential Equations **18** (2006), no. 3, 667–691.

[LKO14] P Langfield, B Krauskopf, and H M Osinga, *Solving winfree's puzzle: The isochrons in the fitzhugh-nagumo model*, Chaos: An Interdisciplinary Journal of Nonlinear Science **24** (2014), no. 1, 013131.

[LKO15] ______, *Forward-time and backward-time isochrons and their interactions*, SIAM Journal on Applied Dynamical Systems **14** (2015), no. 3, 1418–1453.

[LM13] Y Lan and I Mezić, *Linearization in the large of nonlinear systems and Koopman operator spectrum*, Phys. D **242** (2013), 42–53. MR 3001394

[Mañ78] R Mañé, *Persistent manifolds are normally hyperbolic*, Transactions of the American Mathematical Society **246** (1978), 261–283.

[Mar93] J E Marsden, *Lectures on mechanics*, 1 ed., Cambridge University Press, 1993.

[May72] R M May, *Limit cycles in predator-prey communities*, Science **177** (1972), no. 4052, 900–902.

[Mei02] G Meigniez, *Submersions, fibrations and bundles*, Transactions of the American Mathematical Society **354** (2002), no. 9, 3771–3787.

[Mez17] I Mezić, *Spectral koopman operator methods in dynamical systems*, Unpublished textbook, 2017.

[Mil69] J Milnor, *Morse theory*, Annals of Mathematics Studies, vol. 51, Princeton University Press, 1969.

[MM12] A Mauroy and I Mezić, *On the use of fourier averages to compute the global isochrons of (quasi) periodic dynamics*, Chaos: An Interdisciplinary Journal of Nonlinear Science **22** (2012), no. 3, 033112.

[MOWZ91] J E Marsden, OM O'Reilly, F J Wicklin, and B W Zombros, *Symmetry, stability, geometric phases, and mechanical integrators*, Nonlinear Science Today **1** (1991), no. 1, 4–11.

[MR94] J E Marsden and T S Ratiu, *Introduction to mechanics and symmetry*, 1 ed., Springer-Verlag, 1994.

[MRMM14] A Mauroy, B Rhoads, J Moehlis, and I Mezic, *Global isochrons and phase sensitivity of bursting neurons*, SIAM Journal on Applied Dynamical Systems **13** (2014), no. 1, 306–338.

[Nee10] K-H Neeb, *Differential topology of fiber bundles*, 2010.

[Nes03] J Nestruev, *Smooth manifolds and observables*, 1 ed., Springer-Verlag, 2003.

[Øks13] B Øksendal, *Stochastic differential equations*, 6 ed., Springer-Verlag, 2013.

[OM10] H M Osinga and J Moehlis, *Continuation-based computation of global isochrons*, SIAM Journal on Applied Dynamical Systems **9** (2010), 1201.

[PS70] C C Pugh and M Shub, *Linearization of normally hyperbolic diffeomorphisms and flows*, Invent. Math. **10** (1970), 187–198. MR 0283825

[PS08] J Peters and S Schaal, *Natural actor-critic*, Neurocomputing **71** (2008), no. 7-9, 1180 – 1190.

[PSW97] C C Pugh, M Shub, and A Wilkinson, *Hölder foliations*, Duke Math. J. **86** (1997), no. 3, 517–546. MR MR1432307 (97m:58155)

[PT77] J Palis and F Takens, *Topological equivalence of normally hyperbolic dynamical systems*, Topology **16** (1977), no. 4, 335–345. MR 0474409

[Pur77] E M Purcell, *Life at low reynolds number*, American Journal of Physics **45** (1977), no. 1, 3–11.

[Rev09] S Revzen, *Neuromechanical control architectures in arthropod locomotion*, Ph.D. thesis, Univeristy of California, Berkeley, Dec 2009, Department of Integrative Biology.

[RG08] S Revzen and J M Guckenheimer, *Estimating the phase of synchronized oscillators*, Phys Rev E **78** (2008), no. 5, 051907.

[RG12] S Revzen and J M Guckenheimer, *Finding the dimension of slow dynamics in a rhythmic system*, J R Soc Lond Interface **9** (2012), no. 70, 957–971.

[RGF11] S Revzen, J M Guckenheimer, and R J Full, *Subtle differences in gaits: the perspective of data driven floquet analysis*, Yearly meeting of the Society for Integrative and Comparative Biology, Society for Integrative and Comparative Biology, Jan 2011.

[RK15] S Revzen and M Kvalheim, *Data driven models of legged locomotion*, SPIE Defense+ Security, International Society for Optics and Photonics, 2015, pp. 94671V–94671V.

[RKF08] S Revzen, D E Koditschek, and R J Full, *Progress in motor control - a multidisciplinary perspective*, ch. Towards testable neuromechanical control architectures for running, pp. 25–56, Springer Science+Business Media, LLC - NY, 2008.

[Rob71] C Robinson, *Differentiable conjugacy near compact invariant manifolds*, Bol. Soc. Brasil. Mat **2** (1971), no. 1, 33–44.

[Rob89] A J Roberts, *Appropriate initial conditions for asymptotic descriptions of the long term evolution of dynamical systems*, The ANZIAM Journal **31** (1989), no. 1, 48–75.

[Rob00] ———, *Computer algebra derives correct initial conditions for low-dimensional dynamical models*, Computer physics communications **126** (2000), no. 3, 187–206.

[Rud76] W Rudin, *Principles of mathematical analysis*, 3 ed., McGraw-Hill, 1976.

[SA10] S. Schaal and C. Atkeson, *Learning control in robotics*, Robotics Automation Magazine, IEEE **17** (2010), no. 2, 20 –29.

[Sak90] K Sakamoto, *Invariant manifolds in singular perturbation problems for ordinary differential equations*, Proceedings of the Royal Society of Edinburgh Section A: Mathematics **116** (1990), no. 1-2, 45–78.

[Sak94] ———, *Smooth linearization of vector fields near invariant manifolds*, Hiroshima Mathematical Journal **24** (1994), no. 2, 331–355.

[Sch08] S Schecter, *Exchange lemmas 2: General exchange lemma*, Journal of Differential Equations **245** (2008), no. 2, 411–441.

[Sel83] G R Sell, *Vector fields in the vicinity of a compact invariant manifold*, Equadiff 82, Springer, 1983, pp. 568–574.

[Sel84] ———, *Linearization and global dynamics*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Warsaw, 1983), PWN, Warsaw, 1984, pp. 1283–1296. MR 804778

[SKK87] M Spong, K Khorasani, and P Kokotovic, *An integral manifold approach to the feedback control of flexible joint robots*, IEEE Journal on Robotics and Automation **3** (1987), no. 4, 291–300.

[SP10] J T C Schwabedal and A Pikovsky, *Effective phase dynamics of noise-induced oscillations in excitable systems*, Physical Review E **81** (2010), no. 4, 046218.

[SP13] ———, *Phase description of stochastic oscillations*, Physical review letters **110** (2013), no. 20, 204102.

[SPNI05] S Schaal, J Peters, J Nakanishi, and A Ijspeert, *Learning movement primitives*, Robotics Research, Springer, 2005, pp. 561–572.

[Ste51] N E Steenrod, *The topology of fibre bundles*, vol. 14, Princeton University Press, 1951.

[Sto50] J J Stoker, *Nonlinear vibrations in mechanical and electrical systems*, vol. 2, Interscience Publishers New York, 1950.

[SW99] H L Smith and P Waltman, *Perturbation of a globally stable steady state*, Proceedings of the American Mathematical Society **127** (1999), no. 2, 447–453.

[SWT06] B Schelter, M Winterhalder, and J Timmer (eds.), *Handbook of time series analysis: recent theoretical developments and applications*, Wiley, 2006, ISBN: 9783527406234.

[Tak71] F Takens, *Partially hyperbolic fixed points*, Topology **10** (1971), 133–147. MR 0307279

[TBF94] L H Ting, R Blickhan, and R J Full, *Dynamic and static stability in hexapedal runners*, J Exp Biol **197** (1994), 251–269 (English).

[TL14] P J Thomas and B Lindner, *Asymptotic phase for stochastic oscillators*, Physical review letters **113** (2014), no. 25, 254101.

[Tyt13] E D Tytell, *The intrinsic dynamical properties of muscle are self-stabilizing for rhythmic movements*, Soc Integ Compar Biol, Jan 2013, p. 40.1.

[VdP34] B Van der Pol, *The nonlinear theory of electric oscillations*, Proceedings of the Institute of Radio Engineers **22** (1934), no. 9, 1051–1086.

[vS79] S J van Strien, *Center manifolds are not C^∞*, Mathematische Zeitschrift **166** (1979), no. 2, 143–145.

[WGK03] E R Westervelt, J W Grizzle, and DE Koditschek, *Hybrid zero dynamics of planar biped walkers*, Automatic Control, IEEE Transactions on **48** (2003), no. 1, 42–56.

[Wig94] S Wiggins, *Normally hyperbolic invariant manifolds in dynamical systems*, Applied Mathematical Sciences, vol. 105, Springer-Verlag, New York, 1994, With the assistance of György Haller and Igor Mezić. MR 1278264

[Wil67] F W Wilson, Jr., *The structure of the level surfaces of a Lyapunov function*, J. Differential Equations **3** (1967), 323–329. MR 0231409

[Wil69] ———, *Smoothing derivatives of functions and applications*, Trans. Amer. Math. Soc. **139** (1969), 413–428. MR 0251747

[Win80] A T Winfree, *The geometry of biological time*, Springer-Verlag, New York, 1980.

[WS12] Y Wang and M Srinivasan, *System identification and stability analysis of steady human walking and the swing leg dynamics*, ASME 5th Annual Dynamic Systems and Control Conference, vol. 2, ASME, Oct 2012, pp. 19–23.

[Yi93a] Y F Yi, *A generalized integral manifold theorem*, Journal of differential equations **102** (1993), no. 1, 153–187.

[Yi93b] ———, *Stability of integral manifold and orbital attraction of quasi-periodic motion*, Journal of differential equations **103** (1993), no. 2, 278–322.

[ZKK04] A Zagaris, H G Kaper, and T J Kaper, *Fast and slow dynamics for the computational singular perturbation method*, Multiscale Model. Simul. **2** (2004), no. 4, 613–638. MR 2113172