Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Wissenschaftliches Rechnen

Validated computation of connecting orbits in ordinary differential equations

Christian P. Reinhardt

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ. Prof. Dr. Daniel Matthes
Prüfer der Dissertation:	1.	Univ. Prof. Dr. Oliver Junge
	2.	Prof. Jean-Philippe Lessard, PhD., Univ. Laval, Québec/Kanada
	3.	Prof. dr. Jan Bouwe van den Berg, Vrije Univ. Amsterdam/Niederlande (nur schriftliche Beurteilung)

Die Dissertation wurde am 10.04.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 17.06.2013 angenommen.

Contents

Abstract

This thesis is concerned with the validated computation of connecting orbits in continuous dynamical systems. Our approach provides approximations to connecting orbits with exact error bounds and leads to a mathematically rigorous existence proof based on numerical calculations. We first formulate equivalent zero finding problems on appropriate Banach spaces and validate approximate solutions via fixed point arguments. As applications we consider the Lorenz system and the Gray-Scott equations.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der validierten Berechnung verbindender Orbits in kontinuierlichen dynamischen Systemen. Unsere Methode liefert Näherungen an verbindende Orbits, exakte Fehlerschranken und nutzt numerische Berechnungen zu einem mathematisch rigorosen Existenzbeweis. Wir formulieren äquivalente Nullstellengleichungen auf geeigneten Banachräumen und validieren Näherungslösungen über Fixpunktargumente. Als Anwendungen betrachten wir das Lorenz-System und die Gray-Scott Gleichung.

Acknowledgments

First I would like to thank my supervisor Oliver Junge for the advice and support over the years. His inspiration and guidance enabled me to become an active member of the mathematical research community.

Second I owe a warm thanks to my colleagues and friends Jason Mireles-James and Jean-Philippe Lessard for the great collaboration whose outgrowth is this thesis. Their hospitality made my research visits to Rutgers University in New Brunswick and Université Laval in Québec City not only very productive but also great fun. In particular I also thank Konstantin Mischaikow for his support and warm welcome to his group at Rutgers.

I also thank my colleagues at M3 for creating a great working atmosphere and for their indulgence with my cake debts.

Additionally I would like to thank TopMath for creating a very good environment to support young mathematicians.

Last but not least I am grateful to my family and friends for their enduring support in my decisions.

T.T.T.

Chapter 1

Introduction

Connecting orbits are one of the fundamental building blocks for the understanding of the global dynamics of a given dynamical system. Thinking for instance of a dynamical model for a physical system the first step in obtaining an understanding of the model is to understand its equilibrium states and investigate the existence of periodic or oscillatory motion. The next step is to ask what happens if we perturb a stationary or periodic solution. Here connecting orbits come into play. If there exists for example a connecting orbit between one stationary state and an oscillatory solution, this means that starting from a suitable perturbation of the equilibrium we asymptotically converge to periodic motion. In this sense connecting orbits help to build a dynamical scaffold organizing the possible dynamics of a given equation. But connecting orbits not only assist in establishing structure, they also lead to the creation of complexity. A well-known example is the Shilnikov Theorem [43, 36]. In this context the existence of a connecting orbit from an equilibrium to itself implies the existence of infinitely many periodic as well as infinitely many nonperiodic bounded solutions.

The goal of this work is to provide numerical verification methods for the existence of connecting orbits in dynamical systems induced by an ordinary differential equation. A fundamental objective of our rigorous numerical algorithms is to bridge the gap between numerical simulation and rigorous mathematical analysis. More concretely, in accordance with the general idea of the field of validated numerics as for example advocated in [47], we use a combination of analysis and numerics to derive rigorous a-posteriori error bounds for an approximation to a connecting orbit within which we can guarantee, in the mathematically exact sense, a real connection to be found.

Given the importance of connecting orbits it comes as no surprise that there are several algorithms for their numerical approximation available. More precisely for dynamical systems induced either by the iteration of an invertible map or by an ordinary differential equation (ODE) there are a number of classical algorithms for approximating equilibria, periodic orbits and connections between them. However according to Palmer in [33, p.428] theorems rigorously assuring the existence for transversal connecting orbits in dynamical systems induced by flows are scarce. Therefore we consider our results on the rigorous computation of certain homoclinic and transversal heteroclinic orbits in dynamical systems induced by ODEs as a valuable contribution in this context. We hasten to add that this is by no means the first time a rigorous numerical approach is taken to prove existence of connecting orbits and we will give a review of the existing approaches in the sequel. Before we describe the details of our method we give an overview of the foundational algorithms we build on.

We first discuss the basic idea of methods due to Beyn and Doedel [19, 23, 1] for the case of heteroclinic orbits between hyperbolic fixed points. We emphasize that a similar idea applies to the computation of homoclinic orbits and connections between more general invariant sets like periodic orbits. Suppose we are given two hyperbolic fixed points $p_{1,2} \in \mathbb{R}^d$ of the nonlinear ODE

$$\dot{u} = g(u) \quad u \in \mathbb{R}^d. \tag{1.1}$$

A connecting orbit from p_1 to p_2 corresponds to a solution $u : (-\infty, \infty) \to \mathbb{R}^d$ of (1.1) such that

$$\lim_{t \to -\infty} u(t) = p_1 \qquad \lim_{t \to \infty} u(t) = p_2. \tag{1.2}$$

The first fundamental challenge from a numerical perspective is that (1.2) is an inherently asymptotic statement. As a consequence we need to truncate the time interval $(-\infty, \infty)$ to a finite interval $[t_0, t_1]$. An immediate problem entailed by this truncation is the control of the boundary values. More concretely, we have to make sure that the values $u(t_0)$ and $u(t_1)$ are chosen appropriately to ensure condition (1.2) to be satisfied.

One way to deal with this problem is to use the concept of invariant manifolds. More precisely, we know that associated to every hyperbolic fixed point p of (1.1) for sufficiently smooth g there is a stable and an unstable manifold $W^{s,u}(p)$ characterized by the fact that elements of $W^s(p)$ are

asymptotic to p in forward time and elements of $W^u(p)$ are asymptotic to p in backward time. Therefore in order to assure that the solution u on $[t_0, t_1]$ corresponds to a connecting orbit we need to demand that $u(t_0) \in W^u(p_1)$ and $u(t_1) \in W^s(p_2)$. To treat this requirement numerically, we consequently need to approximate the nonlinear objects $W^{s,u}(p_{1,2})$.
The classical approach followed by Doedel et. al in [19, 23] consists in imposing $u(t_0)$ and $u(t_1)$ to lie in linear approximations of the manifolds and solve (1.1) on the finite interval augmented by appropriate additional conditions. This method is also used for continuation of connecting orbits [22] and is implemented in the software package AUTO [49]. A similar approach is followed in the method of projected boundary conditions of Beyn et al. [1]. This boundary value approach is also foundational to Lin's method utilized by Krauskopf et al. for finding connecting orbits [35, 32]. A complimentary software package to AUTO specialized on homoclinic orbits is given by HomCont (see [49]). In this context we also mention the powerful continuation software MatCont [18].

At this point it becomes evident that following this strategy there are two sources of numerical errors to control. First the numerical integration of (1.1) induces errors and second the defect introduced by the truncation to the finite interval has to be controlled. It is shown in [1] using the theory of exponential dichotomies that for the method of projected boundary conditions the latter error decays exponentially with the length of the integration interval. This in turn clarifies a fundamental trade-off: the longer the integration time the better we can control the error introduced by the approximation of the manifold but the more errors are potentially accumulated during integration. The shorter the integration the fewer errors are introduced but the harder it gets to control the approximation of the manifolds.

Another approach to compute approximations to connecting orbits is given by a set-oriented method applied in [29]. This procedure is based on the subdivision technique to approximate invariant manifolds of invariant objects [16, 17]. The idea is to encode the continuous dynamics as a combinatorial graph and compute box coverings of the involved stable and unstable manifolds. Searching for intersections between them leads to a box covering of connecting orbits. These set-oriented techniques are implemented in the software package GAIO [30].

We now turn to the description of the approach for the rigorous computation of connecting orbits followed in this work. The general philosophy, inspired by the one employed in [60], consists in encoding the property of being a connecting orbit in an appropriate operator equation of the form

$$F(x) = 0$$

on a Banach space X. We will take two different paths to achieve this.

Our first approach tailored to the computation of transversal heteroclinic orbits between hyperbolic fixed points of (1.1) is based on the parametrization method developed in [7, 8, 9]. Note that we will not use the dynamics of the flow induced by (1.1) explicitly. Hence in the numerical treatment we do not need to discretize it and therefore we term this method the **discretization free approach**.
Assume two hyperbolic fixed points $p_{1,2} \in \mathbb{R}^d$ of (1.1) with unstable dimension n_u and stable dimension n_s such that $n_s + n_u - 1 = d$ to be given. We will henceforth refer to hyperbolic fixed points fulfilling these dimensional requirements as generic hyperbolic fixed points. The parametrization method provides us with power series expansions of maps Q and P together with their domains of definition $V_{\nu_{u,s}} \subset \mathbb{R}^{n_{u,s}}$ such that $Q(V_{\nu_u}) \subset W^u(p_1)$ and $P(V_{\nu_s}) \subset W^s(p_2)$. The main tool in their computation are certain invariance equations fulfilled by Q and P, lending themselves also to powerful methods for flow computations on the manifolds (see 4.2.1). Taking into account the fact that connecting orbits lie in the intersection of the stable and unstable manifolds of the corresponding fixed points, finding values $\tilde{\varphi} \in \mathbb{R}^{n_u}$ and $\tilde{\phi} \in \mathbb{R}^{n_s}$ such that

$$Q(\tilde{\varphi}) = P(\tilde{\phi}) \Leftrightarrow Q(\tilde{\varphi}) - P(\tilde{\phi}) = 0$$

implies that there exists a connecting orbit between p_1 and p_2. In this sense the discretization free approach is a local approach, as it builds on the intersection of local unstable and stable manifolds of p_1 and p_2. In order to obtain a well-posed equation with isolated zeros we need to impose an additional phase condition. We realize this by locking one parameter variable in the unstable parameter space. Formally we achieve this by composing Q with a parametrization $\Theta : \mathbb{R}^{n_u - 1} \to \mathbb{R}^{n_u}$ of a co-dimension 1 submanifold in parameter space. This leads us to define $F : \mathbb{R}^d \to \mathbb{R}^d$ by

$$F(\alpha, \phi) = Q(\Theta(\alpha)) - P(\phi). \tag{1.3}$$

Now finding $\tilde{x} = (\tilde{\alpha}, \tilde{\phi}) \in \mathbb{R}^d$ such that $F(\tilde{x}) = 0$ is equivalent with finding a connecting orbit between p_1 and p_2. The rigorous numerical evaluation of this map involves interval arithmetical evaluations of truncations Q_M and P_N of the power series for Q and P. We use the MATLAB interval library INTLAB [48] to carry out these computations. The truncation errors $Q - Q_M$ and $P - P_N$ together with their derivatives are controlled by the theory introduced in [60], where Cauchy estimates are used to control the derivative. Assuming an approximate solution $\bar{x} = (\bar{\alpha}, \bar{\phi}) \in \mathbb{R}^d$ of (1.3) to be given, we conclude the existence of a unique zero $(\tilde{\alpha}, \tilde{\phi})$ by calling upon the Newton Kantorovich Theorem [45]. More precisely under certain assumptions on the quality of the approximation $\bar{x}$, that we obtain by a classical Newton iteration, and the invertibility of the derivative $DF(\bar{x})$ we are able to compute a radius $\bar{r} > 0$ such that we can guarantee a unique solution $\tilde{x}$ to exist in a ball $B_{\bar{x}}(\bar{r}) \subset \mathbb{R}^d$ around the approximate solution $\bar{x}$. Taking the structure of F into account we are furthermore able to guarantee that the heteroclinic connection we compute corresponds to a transversal intersection of the corresponding stable and unstable manifolds. More precisely we show that invertibility of $DF(\tilde{x})$ implies transversality of the connection. In particular we can assure the invertibility of $DF(\tilde{x})$ despite the fact that we only have bounds on the location of $\tilde{x}$ but do not know it exactly.

For our second approach we first rewrite (1.1) as integral equation

$$F_1(u) \overset{\text{def}}{=} p_0 + \int_{t_0}^{t} g(u(s))ds - u(t) = 0 \tag{1.4}$$

for a given initial time t_0, where $u(t_0) = p_0$. Next we follow an idea similar to the classical approach of projected boundary conditions. Therefore we term this method **boundary value approach**. The main difference to the classical algorithm, beside the quest for a-posteriori error bounds, is that we use a higher order approximation of the invariant manifolds. Let us describe the idea for heteroclinic connections between generic hyperbolic fixed points $p_{1,2} \in \mathbb{R}^d$ of (1.1). Assume parametrizations Q and P of the unstable and stable manifold $W^{u,s}(p_{1,2})$ to be given and an interval $[t_0, t_1]$ to be chosen. As above we require $u(t_0) \in W^u(p_1)$ and $u(t_1) \in W^s(p_2)$ which in terms of (1.4) becomes

$$Q(\varphi) + \int_{t_0}^{t_1} g(u(s))ds = P(\phi). \tag{1.5}$$

Appending (1.5) to (1.4), resorting again to a suitable phase condition encoded by a map $\Theta : \mathbb{R}^{n_u-1} \to \mathbb{R}^{n_u}$ and setting $\theta = (\alpha, \phi) \in \mathbb{R}^d$ we aim to find a zero of the operator

$$\begin{aligned} F(\theta, u) &= \begin{pmatrix} Q(\Theta(\alpha)) + \int_{t_0}^{t_1} g(u(s))ds - P(\phi) \\ Q(\Theta(\alpha)) + \int_{t_0}^{t} g(u(s))ds - u(t) \end{pmatrix} \\ &\stackrel{\text{def}}{=} \begin{pmatrix} Q(\Theta(\alpha)) + \int_{t_0}^{t_1} g(u(s))ds - P(\phi) \\ F_1(\theta, u) \end{pmatrix} \end{aligned} \tag{1.6}$$

defined on $X = \mathbb{R}^d \times B$, where B is a suitable function space and F_1 was redefined to account for the phase condition. We will give more details on the choice of B as we proceed. Note that zeros of F correspond to connecting orbits between $p_{1,2}$. In a more general context we assume F to be of the form

$$F(\theta, u) = \begin{pmatrix} \mathcal{G}(\mathrm{p}_1(\theta), \mathrm{p}_2(\theta)) \\ F_{1,2}(\theta, u) \end{pmatrix} \tag{1.7}$$

where $\theta \in \mathbb{R}^p$ and $\mathcal{G} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^p$ encodes the boundary conditions and where p is the dimension of parameter space. More precisely the maps $\mathrm{p}_1(\theta)$ and $\mathrm{p}_2(\theta)$ represent the boundary conditions at time $t = t_0$ and $t = t_1$ respectively. As a difference to the discretization free approach in the form presented above, these can be chosen for example in a way accounting for symmetries in the system or also more general forms of boundary conditions. F_2 is obtained in a similar way as F_1 by integrating backwards in time. The benefit of this general notation is that we can take symmetries into account and adapt to the computation of more general connecting orbits. We elaborate on this in more detail in Section 3.2.

Concerning the rigorous numerical treatment we follow two different approaches. The common feature is that in both cases we use the method of radii polynomials (e.g. see [15]), providing an effective approach for the verification of the applicability of the Banach Fixed Point Theorem (BFP). More precisely, the idea of this method is to prove the existence of a zero to a map defined on a possibly infinite dimensional Banach space via proving the existence of a solution to an equivalent fixed point problem for an associated Newton-like fixed point operator T constructed from an approximate solution. To achieve this, given the approximation the goal is to apply the BFP on a ball of a-priori unknown radius r around this approximation. As an analogue to the discretization-free approach, a central issue in the construction of T is to obtain a good approximate inverse of the Fréchet derivative Df at the approximate solution.

To obtain numerically amenable conditions for the application of the BFP,

i.e. T should be a contractive self-mapping on the ball around the approximate solution, the requirements are encoded in a set of finitely many strict polynomial inequalities, where the unknown radius r is the variable. More precisely, if one finds a radius $\bar{r}$ such that the polynomial inequalities are simultaneously fulfilled, the operator T is a contraction on the ball of radius $\bar{r}$ around the approximation. As a result we can guarantee a unique fixed point of T to exist in this ball and hence we also obtain the existence of a solution to our original zero finding problem in the possibly infinite dimensional space.
We point out that in the case of an infinite dimensional problem a so-called tail radii polynomial is constructed via analytic estimates controlling the truncation error introduced during the Galerkin projection. The exact derivation depends on the choice of method that shall be our next concern.

1. First we discretize F by using linear splines. This is much in the spirit of [60] with the difference that we adapt the approach to generic first order ODEs and obtain results on transversality. More precisely we represent u by its linear spline approximation and obtain a splitting of F of the form $F = F_m \oplus F_\infty$, where F_m is defined on a finite dimensional space. Thus the approximate zero $(\bar{\theta}, \bar{u})$ can be found by applying a classical non-rigorous Newton approach to F_m. In addition F_m can be evaluated rigorously together with its derivative again by a combination of interval arithmetic and rigorous error analysis based on [60]. The fixed point operator T is of the form $T = T_m \oplus T_\infty$, where T_m is Newton-like. It is constructed from the splitting $F_m \oplus F_\infty$. By applying the method of radii polynomials we get the existence of a zero $(\tilde{\theta}, \tilde{u})$ of F within a ball of radius $\bar{r}$ around the approximation $(\bar{\theta}, \bar{u})$. The tail errors involved in the construction of the tail radii polynomial are controlled via classical a-priori spline estimates [50]. In addition we obtain the invertibility of $DF(\tilde{\theta}, \tilde{u})$. In the case of heteroclinic connections between generic hyperbolic fixed points this invertibility again implies transversality.

2. Second we use a spectral discretization based on the Chebyshev expansion of the unknown function u [56, 57]. A central feature is that Chebyshev series are Fourier series in disguise [57] which paves the way to the direct transfer of the mathematical machinery developed in the last ten years [24, 67, 13, 14] to prove existence of solutions to ODEs and PDEs with periodic profiles to the realm of existence

proofs for non-periodic solutions like connecting orbits.

To the best of our knowledge an approach based on Chebyshev spectral discretization has not been used for the rigorous solution of nonlinear differential equations before. We mention the work [6] where the authors develop Chebyshev interpolation polynomial-based tools for rigorous computing. However, it seems that they have not yet applied their methods to rigorously solve nonlinear differential equations.

More precisely the idea of our approach is to use the smoothness of the vector field g and construct a map $f = f(\theta, x)$ defined on the Banach space Ω^s of rapidly decaying sequences, augmented by a parameter variable $\theta \in \mathbb{R}^p$, such that its zeros are in one-to-one correspondence with the zeros of F. In this context the formulation of (1.1) as integral equation is crucial as we can exploit elegant relations between the Chebyshev polynomials and its antiderivatives. To obtain a numerically tractable expression we compute with a Galerkin projection $f^{(m)}$. This enables us to compute an approximate zero $(\bar{\theta}, \bar{x})$ again by a classical Newton method. In addition $f^{(m)}$ can be evaluated rigorously together with its derivative again by a combination of interval arithmetic and rigorous error analysis based on [60]. Given the approximation $(\bar{\theta}, \bar{x})$ we apply the method of radii polynomials by constructing a Newton-like fixed point operator T associated to an approximate zero of f. We again get the existence of a unique solution $(\tilde{\theta}, \tilde{x})$ in a ball of radius $\bar{r}$ around $(\bar{\theta}, \bar{x})$. In order to construct the radii polynomials and in particular to control the tail term we use the connection of Chebyshev series to Fourier series [57] in order to be able to use the analytic convolution estimates introduced in [24, 67, 13, 14].

In addition to the existence of a zero $(\tilde{\theta}, \tilde{x})$ we also assure the invertibility of $Df(\tilde{\theta}, \tilde{x})$.

Before giving an outline of the rest of the work we aim for an overview of alternative rigorous approaches to the computation of connecting orbits for ODEs. For the context of maps we refer to [3] and the introduction of [34] and references therein.

First we mention the work of Palmer based on shadowing techniques [33]. The idea is to establish the existence of a true connecting orbit near a numerically computed one via checking the invertibility of a certain operator. As in our case numerically verifiable conditions for this invertibility are derived.

Furthermore we refer to the work of the CAPD group [40]. Based on the rigorous integration of the flow and the variational equation various existence proofs for connecting orbits in the Circular restricted 3-body problem [64, 65], the Rössler system [63] and the Michelson system [62] were obtained. The direct integration of the flow marks a difference to our approach based on a reformulation of the problem as an operator equation. In addition we mention [44, 66, 42] that are also based on fixed point type arguments. Next we point out topological approaches [37, 39] based on index theoretic arguments using the Conley index [11] and the connection matrix [21]. Applications include the computation of connecting orbits between equilibria of the Swift-Hohenberg PDE [54, 39].

The thesis is structured as follows. In Chapter 2 we compile some background material. In Section 2.1 we collect the necessary facts from dynamical systems theory with a special focus on the parametrization method for the computation of invariant manifolds [7, 8, 9] together with the a-posteriori error analysis [60]. In Section 2.2 we introduce some classical concepts from approximation theory important for the discretization in the boundary value approach. In Chapter 3 we develop the main theoretical results. In Section 3.1 we introduce our discretization-free approach based solely on the parametrization method and an Newton-Kantorovich like argument for the verification. In Section 3.2 we establish the boundary value approach, where we dedicate Section 3.2.1 to the discretization and validation using the linear spline approximation and Section 3.2.2 to the discretization and validation using the spectral approximation. In both cases we use the method of radii polynomials in the validation. In Section 3.3 we establish results guaranteeing in the discretization-free case and in the boundary value approach using linear spline case for heteroclinic orbits between generic hyperbolic equilibria that the intersection of the stable and unstable manifolds implied by our verified connecting orbit is transverse. In both cases this is achieved by showing that we not only get a verified solution of the respective operator equations but also invertibility of the derivative of the corresponding operator at this solution. This fact is then tied to the transversality of the manifolds. For the spectral approximation we obtain an analogue result for the invertibility of the derivative of the operator on the sequence space. In Chapter 4 we present some applications in the context of the Lorenz equations in Section 4.2 and the Gray-Scott equations in Section 4.1. In particular we compare the spline and spectral discretization approach to some extent and extend results obtained in [60] concerning symmetric homoclinics in the Gray-Scott equations using a spline based method by applying our spectral approach. Finally we finish by giving an outlook on some possible future applications.

Chapter 2

Background

This chapter serves as a review of basic notions important for the rest of the thesis. As we will consider validated numerical methods for dynamical systems we concentrate on introducing the relevant facts from the respective fields. We will start with the background material from dynamical systems theory necessary in order to understand the sequel. First we offer a more detailed description of what is mathematically understood by a dynamical system and continue by defining invariant manifolds of hyperbolic fixed points together with the concept of connecting orbits between them. These will be the main characters in the following. For more details we refer to [10, 46]. We dedicate a section to the parametrization method developed in [7, 8, 9] together with its numerical validation theory from [60] as it will be a crucial tool in order to get rigorous approximations to the aforementioned invariant manifolds.

Next we will focus on approximation theory and some numerical aspects of this classical subject. We will by no means aim to give a complete review of the field but restrict ourselves to the topics of direct relevance to this thesis. More concretely our goal is to recall the approach of local spline approximation and the global spectral approximation approach for real-valued functions of one variable.

2.1 Dynamical systems

2.1.1 Basic definitions from dynamical systems theory

Dynamical systems model processes evolving with time, ranging from relatively simple ones like the motion of the harmonic oscillator to much more complex ones like the behavior of chemical reactants. More specifically one considers the state of a system and asks how, depending on the current state, the evolution of the system will carry on in the future. Thinking for example of the oscillator, given the initial position and speed, how can we determine its future fate? The mathematical theory of dynamical systems seeks for an abstract formulation of this heuristic idea. There are two immediate central questions to ask: what do we understand by the state of the system and how do we determine the evolution in the future?

The state of the system is modeled as an element of the so called *phase space* X. Mathematically this is, in our case, a finite or infinite dimensional Banach space. Considering for example the phase space of the harmonic oscillator we obtain $X = \mathbb{R}^2$ as the oscillator models the motion of an elastic spring which is characterized by the two scalar quantities of the displacement from the equilibrium and speed. For spatially inhomogeneous problems in continuous time like a chemical reaction the phase space will typically be an infinite dimensional function space, where at each point in time the function represents the spatial distribution of the reactants. We will elaborate on these types of dynamical systems further below.

The evolution rule is given by the (semi-)flow map which is induced for example by iteration of an invertible map, by an ordinary or a partial differential equation. More formally we have the following definition.

Definition 2.1.1 *Let $\mathbb{T} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}_+\}$ and $\Phi : \mathbb{T} \times X \to X$ be a mapping with the following properties:*

1. $\Phi(x, 0) = x \quad \forall x \in X$
2. $\Phi(\Phi(x,t), s) = \Phi(x, s+t) \qquad \forall s, t \in \mathbb{T}$

If $\mathbb{T} = \mathbb{N}, \mathbb{R}_+$ we call Φ a semi-flow and for $\mathbb{T} = \mathbb{Z}, \mathbb{R}$ we speak of a flow. Depending on whether the state space is finite or infinite dimensional and the time $\mathbb{T}$ is $\mathbb{R}$ or $\mathbb{Z}$ we speak of a (in)finite continuous or discrete dynamical system.

To sum up a dynamical system is a pair (X, Φ) of a phase space X and a (semi)flow Φ on X.

Concretely we will consider two classes of dynamical systems in this work.

1. Smooth dynamical systems on $\mathbb{R}^d$: the flow map is induced by the solution of a nonlinear ODE

$$\dot{u} = g(u), \tag{2.1}$$

where $g : \mathbb{R}^d \to \mathbb{R}^d$ is a nonlinear vectorfield. This will be our primary object of study. The above mentionned oscillator is a very basic example.

2. Nonlinear evolution equations: the semiflow is induced by a semilinear parabolic PDE of the form

$$u_t = g(u) \tag{2.2}$$

where $g(u) = Au + n(u)$ with a linear partial differential operator A and a nonlinearity $n(u)$. We will study equilibrium solutions to these equations that reduce the consideration of the PDE to an ODE problem. An example application can be found in the dynamics of chemical reactions we previously alluded to.

We henceforth assume that we are given a dynamical system $(\mathbb{R}^d, \Phi)$ induced by a nonlinear ODE (2.1).

The central question is, given an initial point $x \in \mathbb{R}^d$ what is the asymptotic result of the action of the flow Φ? This is analyzed by considering the *orbit* of x, defined as follows.

Definition 2.1.2 *Let $\Phi : \mathbb{R}^d \times \mathbb{T} \to \mathbb{R}^d$ be a flow, the orbit of a point $x \in \mathbb{R}^d$ is given by*

$$O(x) = \{\Phi(x,t) : t \in \mathbb{T}\}.$$

Generally it is hard, if not impossible, to compute the flow map analytically. Thus $O(x)$ cannot be computed exactly but is only amenable to numerical investigation. One of the goals of the field of rigorous numerical methods is to bridge the gap between numerical simulation on the one hand and analytical investigation in the mathematically strict sense on the other. The algorithms presented in this thesis aim to contribute to this idea.

As it is not possible to analyze all orbits individually a common strategy to analyze a given dynamical system is to look for points $x \in \mathbb{R}^d$ with particularly simple orbits and use these as building blocks for more complicated dynamical behavior. In this context the starting point is to seek *fixed points* and *periodic points*.

Definition 2.1.3 *Let a flow $\Phi : \mathbb{R}^d \times \mathbb{T} \to \mathbb{R}^d$ be given. $p \in \mathbb{R}^d$ is called fixed point of Φ if $\Phi(p,t) = p$ for all $t \in \mathbb{T}$ or equivalently if $g(p) = 0$.*
$p \in \mathbb{R}^d$ is called periodic orbit if there exists a $T > 0$ such that $\Phi(x,T) = x$.

A third important class is given by *connecting orbits* which help organizing dynamics between fixed and periodic points. For the sake of notational simplicity we only define connections between fixed points. We emphasize that our methods are potentially extendable to connecting orbits between periodic orbits.

Definition 2.1.4 *Let $p_{1,2}$ be fixed points of* (2.1) *with corresponding flow Φ. The orbit $O(x)$ of $x \in \mathbb{R}^d$ is called a connecting orbit from p_1 to p_2 if*

$$\lim_{t\to-\infty} \Phi(x,t) = p_1 \qquad \textit{and} \qquad \lim_{t\to\infty} \Phi(x,t) = p_2.$$

If $p_1 \neq p_2$ then $O(x)$ is called a heteroclinic orbit. For $p_1 = p_2$ one refers to $O(x)$ as a homoclinic orbit.

Together one obtains a dynamical scaffold structuring phase space. We restrict our attention to the computation of connecting orbits between a special class of fixed points, namely *hyperbolic fixed points*.

Definition 2.1.5 *A fixed point $p \in \mathbb{R}^d$ of* (2.1) *is called hyperbolic if*

$$Re(\lambda) \neq 0 \quad \forall \lambda \ \textit{eigenvalue of Dg(p).}$$

Our method to compute connections between hyperbolic fixed points $p_{1,2}$ builds on the fact that associated to a hyperbolic fixed point there are invariant manifolds characterized by the dynamical behavior of their elements, namely the stable and unstable manifold. As these objects are of some importance in our context let us give a precise definition. We restrict our attention to the stable manifold, the unstable manifold can be obtained via time reversal. First we define the local stable manifold $W^s_{loc}(p)$ with respect to a neighborhood U of p by

$$W^s_{loc}(p) = \{x \in U : \lim_{t\to\infty} \Phi(x,t) = p \text{ and } \Phi(x,t) \in U \ \forall t \geq 0\}.$$

The Stable Manifold Theorem [46, 10] states that if the flow Φ is C^k then $W^s_{loc}(p)$ can be represented as a graph of a C^k function over E^s, and hence it is justified to speak of a manifold. Once we have the local stable manifold the global stable manifold of p is defined by

$$W^s(p) = \bigcup_{t\geq 0} \Phi(W^s_{loc}(p), t).$$

In other words the stable manifold of $p \in \mathbb{R}^d$ is characterized by

$$W^s(p) = \{x \in \mathbb{R}^d : \lim_{t\to\infty} \Phi(x,t) = p\}.$$

The classical proofs of the (Un)stable Manifold Theorem either are based on Ljapunov Perron Method [10] or the so called graph transform [46]. The recently developed parametrization method by de la Llave, Fontich et al. [7, 8, 9] provides an alternative approach with far reaching generalizations. The central property of this approach with respect to this work is that it lends itself to efficient numerical implementations including error bounds. We will elaborate on this in more detail in the next section.

Hence for the case of hyperbolic fixed points the following equivalence is valid:

$$O(x) \text{ is a connecting orbit between } p_1 \text{ and } p_2 \Leftrightarrow x \in W^u(p_1) \cap W^s(p_2). \tag{2.3}$$

The relation (2.3) will be central when we formulate the problem of finding connecting orbits in a twofold way as an equivalent zero finding problem on an appropriate Banach space.

Verbalized (2.3) means that an orbit is a connecting orbit between the hyperbolic fixed points $p_{1,2}$ if it lies in the intersection of the unstable manifold of p_1 with the stable manifold of p_2. This in turn implies that this intersection, if non-empty, is always at least one dimensional. Consequently we expect connecting orbits between hyperbolic fixed points generically to occur if the dimensions $n_{u,s}$ of $W^{u,s}(p_{1,2})$ fulfill the relation

$$n_u + n_s \geq d + 1. \tag{2.4}$$

If a given connecting orbit connects two hyperbolic equilibria $p_{1,2}$ for which (2.4) is met, we call it a generic connecting orbit.

We remark that requirement (2.4) is not satisifed for homoclinic orbits to hyperbolic fixed points, as hyperbolicity implies that $n_u + n_s = d$. This makes a homoclinic orbit a codimension one phenomena, that is we expect them to generically occur in one-parameter families of ODEs thus foreshadowing the relation of the occurrence of connecting orbits to global bifurcations. See for example [2] for a more thorough treatment of the role of parameters in connecting orbit problems.

We now go on to give a detailed discussion of the parametrization method including the strategy employed in numerical calculations.

2.1.2 The parametrization method

The present discussion is aimed at reviewing some elements of the parameterization method for stable and unstable manifolds of hyperbolic equilibria of vector fields. We concentrate on those aspects relevant for this thesis. For the full development in all generality see [7, 8, 9]. The fundamental point of the following discussion is that the parametrization P that we aim to compute has to fulfill a *functional equation* that, under some regularity assumptions, we can use in order to derive a power series representation of P. We also review the a-posteriori error analysis from [60].

Derivation and solution of the functional equation

Let us restrict our attention to the computation of the n-dimensional stable manifold of a hyperbolic fixed point p of (2.1) with flow Φ. The unstable manifold can be computed by time reversal. Before we go to the details we state that the overall goal of the parametrization method is to find a neighborhood $\mathbb{R}^n \supset V_\nu$ and a map $P : V_\nu \to \mathbb{R}^d$ such that

$$P(V_\nu) \subset W^s(p).$$

Assume that $Dg(p)$ is diagonalizable over $\mathbb{C}$. Note that this is not a major restriction as diagonalizable matrices form a dense subset of $\mathbb{C}^{n,n}$. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues with negative real part of $Dg(p)$ and $\Lambda \in \mathbb{C}^{n,n}$ the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Assume without loss of generality that there are m real eigenvalues and l pairs of complex conjugate eigenvalues. Let $\lambda_1, \ldots, \lambda_m$ be the real eigenvalues and $\lambda_{m+2j-1}, \lambda_{m+2j}$ for $j = 1, \ldots, l$ be the pairs of complex conjugate eigenvalues. Note that $n = m + 2l$. In order for the parametrization method to succeed we need to impose that $\lambda_1, \ldots, \lambda_n$ are non-resonant.

Definition 2.1.6 *A set of eigenvalues is called non-resonant if for all $k_1, \ldots, k_n \in \mathbb{N}$ with $k_1 + \cdots + k_n \geq 2$ we have that*

$$k_1\lambda_1 + \ldots + k_n\lambda_n \neq \lambda_i$$

for all $i = 1, \ldots, n$.

Henceforth assume that $\lambda_1, \ldots, \lambda_n$ are non-resonant. We will see that the non-resonance condition is crucial for the success of the herein described implementation. In addition assume that the eigenvalues $\lambda_1, \dots, \lambda_n$ are ordered such that for $m \geq 0$ the eigenvalues $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ are real and for $j = 1, \dots, \frac{n-m}{2}$ the eigenvalues λ_{m+2j-1} and λ_{m+2j} are complex conjugate. We set $l = \frac{n-m}{2}$.

For later use let us define for $x \in \mathbb{R}^n$ the norm

$$\|x\|_{(m,l)} = \max\left(\max_{1\leq i\leq m} |x_i|, \max_{1\leq j\leq l} \sqrt{x_{m+2j-1}^2 + x_{m+2j}^2} \right). \tag{2.5}$$

More precisely the domain V_ν will naturally turn out to be a ball with respect to this norm. Further let $\mathcal{A}$ be the matrix with columns $\xi_1, \dots, \xi_n$, where $\xi_1, \dots, \xi_n$ are the possibly complex eigenvectors of $Dg(p)$. In addition let A be the real matrix with columns $v_1, \dots, v_n$ constituting the basis of the corresponding real invariant subspace obtained by defining

$$\begin{aligned} v_i &= \xi_i & i &= 1, \dots, m \\ v_{m+2j-1} &= 2Re(\xi_{m+2j-1}) & j &= 1, \dots, l \\ v_{m+2j} &= -2Im(\xi_{m+2j-1}) & j &= 1, \dots, l. \end{aligned} \tag{2.6}$$

The choice involving the factor 2 will become clear as we proceed. Finally set $J = A^{-1}Dg(p)A$, which explicitly gives

$$J = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & 0 & \\ & & \lambda_m & & & \\ & & & J_1 & & \\ & 0 & & & \ddots & \\ & & & & & J_l \end{pmatrix} \tag{2.7}$$

where

$$J_j = \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix}$$

with $\lambda_{m+2j-1} = a_j + ib_j$ for $j = 1, \dots, l$ and $b_j < 0$.

The aim of the parametrization method is to find an open set $V_\nu \subset \mathbb{R}^n$ together with a parametrization $P : \mathbb{R}^n \supset V_\nu \to \mathbb{R}^d$ such that

$$P(0) = p \qquad DP(0) = A, \tag{2.8}$$

that in addition conjugates the linear flow of $J = A^{-1}Dg(p)A$ in the parameter space with the nonlinear flow on the stable manifold. More precisely we require

$$\Phi(P(\phi), t) = P(e^{Jt}\phi) \qquad \forall \phi \in V_\nu \tag{2.9}$$

for all t where both sides of the equation are defined. Remark that (2.9) then ensures that

$$P(V_\nu) \subset W^s(p).$$

This can be seen by considering for an arbitrary $\phi \in V_\nu$

$$\lim_{t\to\infty} \Phi(P(\phi), t) = \lim_{t\to\infty} P(e^{Jt}\phi) = P(\underbrace{\lim_{t\to\infty} e^{Jt}\phi}_{=0}) = p,$$

where $\lim\limits_{t\to\infty} e^{Jt}\phi = 0$ stems from the fact that the J is a block diagonal matrix corresponding to the eigenvalues of $Dg(p)$ with negative real parts.

Remark 2.1.1 *Knowing the parametrization P, using* (2.9) *enables to compute the dynamics on the stable manifold by computing the linear flow in parameter space and lifting to phase space by applying the parametrization. This is a feature that is useful when the dynamics is sensitive to leave the stable manifold. We will consider a concrete application of this notion in Section 4.2.1. Also in a more general context the parametrization method potentially offers the possibility to compute the flow on more general invariant manifolds using the above described technique.*

By differentiating (2.9) on both sides and evaluating at zero yields the *functional equation*

$$g(P(\phi)) = DP(\phi)J\phi \qquad \forall\phi \in V_\nu, \tag{2.10}$$

which, together with the initial constraints (2.8) is equivalent to (2.9).

For technical reasons, becoming evident momentarily, the strategy to solve (2.10) consists in considering a complex valued extension

$$f : \mathbb{C}^n \supset U \to \mathbb{C}^d$$

satisfying

$$g(f(z)) = Df(z)\Lambda z \qquad \forall z \in \mathbb{B}_\nu, \tag{2.11}$$

where $\mathbb{B}_\nu$ is the ball with respect to the complex sup-norm $\|(z_1, \dots, z_n)\|_\infty = \max\limits_{i=1,\dots,n} |z_i|$ and Λ defined above . Additionally we demand the initial constraints

$$f(0) = p \qquad Df(0) = \mathcal{A} \tag{2.12}$$

with $\mathcal{A} \in \mathbb{C}^{d,n}$ defined earlier. Using the complex extension f, we define the real valued parametrization via a complex change of coordinates by

$$\begin{aligned}
&P(\phi_1, \ldots, \phi_m, \phi_{m+1}, \phi_{m+2}, \ldots, \phi_{m+2l-1}, \phi_{m+2l}) = \\
&f\,(\phi_1, \ldots, \phi_m, \phi_{m+1} + i\phi_{m+2}, \phi_{m+1} - i\phi_{m+2}, \ldots, \\
&\phi_{m+2l-1} + i\phi_{m+2l}, \phi_{m+2l-1} - i\phi_{m+2l})\,.
\end{aligned} \tag{2.13}$$

In other words we define

$$P(\phi) = f(T\phi)$$

with the matrix T given by

$$T = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & & B & & \\ & 0 & & & \ddots & \\ & & & & & B \end{pmatrix}$$

where

$$B = \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}.$$

The real domain of definition of P induced by the complex sup-norm $\|.\|_\infty$ is $V_\nu = \{\phi \in \mathbb{R}^n : \|\phi\|_{m,l} \leq \nu\}$. This can be seen as follows. Assume for

$$\begin{aligned}
z = (&\phi_1, \ldots, \phi_m, \phi_{m+1} + i\phi_{m+2}, \phi_{m+1} - i\phi_{m+2}, \ldots, \\
&\phi_{m+2l-1} + i\phi_{m+2l}, \phi_{m+2l-1} - i\phi_{m+2l}) \in \mathbb{C}^n
\end{aligned}$$

that $\|z\|_\infty < \nu$. This means in particular that $\max_{i=1,\ldots,n} |z_i| < \nu$. Using the standard definition of the complex absolute value we translate this to a requirement for $\phi = (\phi_1, \ldots, \phi_n)$. More precisely this yields

$$\max(\max_{i=1,\ldots,m} |\phi_i|, \max_{j=1,\ldots,l} \|(\phi_{m+2j-1}, \phi_{m+2j})\|_2) < \nu$$

which is exactly described by $\|\phi\|_{(m,l)} < \nu$ as defined in (2.5). We note that we have the estimate $\|x\|_{(m,l)} \leq \sqrt{2}\|x\|_\infty$.

The fact that this also leads to a real valued map with image in $W^s(p)$ and satisfying (2.10) with constraints (2.8) will be clarified after we describe the strategy to compute f numerically.

In order to solve (2.11) we plug in the power series ansatz

$$f(z) = \sum_{|k|\geq 0} a_k z^k \tag{2.14}$$

where $k = (k_1, \ldots, k_n)$ is a multi-index with $|k| = k_1 + \ldots + k_n$, $a_k \in \mathbb{C}^d$ and $z^k = z_1^{k_1} \cdots z_n^{k_n}$. Assuming a Taylor expansion

$$g(z) = g(p) + Dg(p)z + \sum_{|k|\geq 2} b_k z^k = Dg(p)z + \sum_{|k|\geq 2} b_k z^k$$

of g around its hyperbolic fixed point p, we can formulate the following lemma to solve (2.11).

Lemma 2.1.1 *Let $f(z) = \sum_{|k|\geq 0} a_k z^k$ the power series ansatz for the complex parametrization f fulfilling* (2.12) *under the constraints* (2.11)*. In addition let the stable eigenvalues $\lambda_1, \ldots, \lambda_n$ of Dg(p) be non-resonant.*

The coefficients $(a_k)_{|k|\geq 0}$ can be computed recursively setting

$$a_0 = p \qquad\qquad a_{e_i} = \xi_i \quad (i = 1, \ldots, n)$$

where e_i is the i-th standard basis vector. For $|k| \geq 2$ we need to solve the linear systems

$$(Dg(p) - (k_1\lambda_1 + \cdots + k_n\lambda_n)\mathbf{1_{d,d}})a_k = -c_k \tag{2.15}$$

where $c_k = c_k(a_{\bar{k}}, |\bar{k}| < |k|) \in \mathbb{R}^d$ is a function of multi-indices $\bar{k}$ of absolute value strictly less than k determined by the nonlinearity g.

Proof 2.1.1 *First the initial constraints* (2.12) *yield*

$$f(0) = a_0 = p \qquad \textit{and} \qquad Df(0) = [a_{e_1}, \ldots, a_{e_n}] = \mathcal{A}.$$

Next we expand both sides of (2.11) *and match like powers. Let us start with the left hand side.*

$$\begin{aligned} g(f(z)) &= Dg(p)f(z) + \sum_{|k|\geq 2} b_k (\sum_{|k|\geq 0} a_k z^k)^k \\ &= \sum_{|k|\geq 0} Dg(p)a_k z^k + \sum_{|k|\geq 2} c_k z^k \end{aligned}$$

where $c_k = c_k(a_{\bar{k}}, \bar{k} < k)$. For the right hand side we first realize that

$$\frac{\partial z^k}{\partial z_i} = k_i z^{k-e_i}$$

and hence

$$\left(\frac{\partial z^k}{\partial z_1}, \dots, \frac{\partial z^k}{\partial z_n}\right) \Lambda z = (\lambda_1 k_1 + \dots + \lambda_n k_n) z^k$$

where we recall that Λ *is the diagonal matrix with diagonal entries* $\lambda_1, \dots, \lambda_n$. *Thus we get*

$$Df(z)\Lambda z = \sum_{|k|\geq 0} a_k (\lambda_1 k_1 + \dots + \lambda_n k_n) z^k.$$

Therefore (2.11) *is equivalent to*

$$\sum_{|k|\geq 0} Dg(p) a_k z^k + \sum_{|k|\geq 2} c_k z^k = \sum_{|k|\geq 0} a_k (\lambda_1 k_1 + \dots + \lambda_n k_n) z^k$$

or

$$\sum_{|k|\geq 0} (Dg(p) - (\lambda_1 k_1 + \dots + \lambda_n k_n) \mathbf{1_{d,d}}) a_k z^k = - \sum_{|k|\geq 2} c_k z^k.$$

Matching like powers yields (2.15). □

Remark 2.1.2 1. *Note that the non-degeneracy assumption on the eigenvalue is crucial. In other words if there would be* $k_1, \dots, k_n \in \mathbb{N}$ *such that*

$$k_1 \lambda_1 + \dots k_n \lambda_n = \lambda_j$$

for some $j = 1, \dots, n$ *then the system* (2.15) *may not be solvable!*

2. (2.15) *is often referred to as homological equation.*

3. *We will see concrete formulas for* c_k *in the sequel when we consider flows induced by polynomial vector fields.*

4. *Realize that the fact that* Λ *is a diagonal matrix, and not a nondiagonal matrix like* J, *is crucial for simplifying the algebra!*

5. *By solving* (2.15) *to arbitrary order* N *we get high order polynomial approximation*

$$f_N(z) = \sum_{|k|=0}^{N} a_k z^k \tag{2.16}$$

for which we will be able to state error estimates. We will return to this point later.

We finish this discussion by assuring that (2.13) indeed induces a real valued map satisfying (2.8) and (2.10).

Lemma 2.1.2 *Let P be the map defined on V_ν by* (2.13). *Then $P(V_\nu) \subset \mathbb{R}^n$ and*

$$P(0) = p \qquad DP(0) = A.$$

In addition the functional equation

$$g(P(\phi)) = DP(\phi)J\phi$$

is fulfilled for all $\phi \in V_\nu$.

Proof 2.1.2 *To show that P is real valued we show that $conj(f(T\phi)) = f(T\phi)$ where $conj$ denotes complex conjugation. We also use the bar notation when it is convenient. Choose a multi-index $(k_1, \ldots, k_n)$ and compute*

$$conj(a_{(k_1,\ldots,k_n)} \prod_{i=1}^{m} \phi_i^{k_i} \prod_{j=1}^{l} z_j^{k_{m+2j-1}} \bar{z}_j^{k_{m+2j}}) =$$
$$a_{\Pi((k_1,\ldots,k_n))} \prod_{i=1}^{m} \phi_i^{k_i} \prod_{j=1}^{l} \bar{z}_j^{k_{m+2j-1}} z_j^{k_{m+2j}})$$

where $\Pi((k_1, \ldots, k_n))$ denotes the permutation that exchanges k_{m+2j-1} and k_{m+2j} for $j = 1, \ldots, l$. This stems form the fact that complex conjugation of (2.15) *permutes exactly these indices. Hence $conj(f(T\phi))$ corresponds to a permutation of the summands and hence*

$$conj(f(T\phi)) = f(T(\phi))$$

and $P(\phi) = f(T\phi)$ is real for all $\phi \in V_\nu$.

Next we have by definition

$$P(0) = f(\underbrace{T0}_{=0}) = p.$$

Furthermore

$$DP(0) = Df(T0)T = \mathcal{A}T.$$

Denote the columns of $\mathcal{A}T$ by $\pi_1, \ldots, \pi_n$ and recall that the columns of $\mathcal{A}$ are $\xi_1, \ldots, \xi_n$. By definition we have for $i = 1, \ldots, m$

$$\pi_i = \xi_i = v_i.$$

For $j = 1, \ldots, l$ we have that

$$\pi_{m+2j-1} = \xi_{m+2j-1} + \xi_{2j} = 2Re(\xi_{m+2j-1}) = v_{m+2j-1}$$
$$\pi_{m+2j} = i(\xi_{m+2j-1} - \xi_{2j}) = -2Im(\xi_{m+2j-1}) = v_{m+2j},$$

where we recall that ξ_{m+2j-1} and ξ_{m+2j} are complex conjugate eigenvectors for $j = 1, \dots, l$. Hence remembering (2.6) *yields $DP(0) = A$.*

To show the functional equation (2.10) *we realize that*

$$g(P(\phi)) = g(f(T\phi)) \underset{(2.11)}{=} Df(T\phi)\Lambda T\phi = DP(\phi)T^{-1}\Lambda T\phi, \tag{2.17}$$

where we use that

$$DP(\phi) = Df(T\phi)T.$$

From the fact that for $a, b \in \mathbb{R}$

$$\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}^{-1} = \begin{pmatrix} a+ib & 0 \\ 0 & a-ib \end{pmatrix},$$

we get for $j = 1, \dots, l$ that

$$BJ_jB^{-1} = \begin{pmatrix} \lambda_{m+2j-1} & 0 \\ 0 & \bar{\lambda}_{m+2j-1} \end{pmatrix}$$

and hence $\Lambda = TJT^{-1}$. Plugging this into (2.17) *yields*

$$g(P(\phi)) = DP(\phi)J\phi.$$

□

Let us next consider the a-posteriori error bounds. In preparation for the error analysis of the parametrization computation we topologize the space of analytic functions $f : \mathbb{B}_\nu \subset \mathbb{C}^n \to \mathbb{C}^d$ with the norm

$$\|f\|_\nu = \sup_{z \in \mathbb{B}_\nu} \|f(z)\|_\infty.$$

In addition, considering a power series expansion of f on $\mathbb{B}_\nu$ given by

$$f(z) = \sum_{|k|=0}^{\infty} b_k z^k, \quad (b_k \in \mathbb{C}^d),$$

where $k = (k_1, \dots, k_n)$ is a multi-index, $|k| = \sum_{i=1}^n k_i$ and $z^k = \prod_{i=1}^n z_i^{k_i}$, define the norm

$$\|f\|_{\Sigma,\nu} = \sum_{|k|\geq 0}^{\infty} \|b_k\|_\infty \nu^{|k|}.$$

Note that $\|.\|_{\Sigma,\nu}$ is efficiently computable if f is a polynomial and that $\|f\|_\nu \leq \|f\|_{\Sigma,\nu}$. These facts will be exploited in our computations. In the a-posteriori analysis for both our approaches to connecting orbit computation we will also have to control derivatives of the parametrizations. Therefore for matrix valued analytic functions $A : \mathbb{B}_\nu \subset \mathbb{C}^n \to \mathbb{C}^{d,d}$ we set

$$\begin{aligned} \|A\|_{M,\nu} &= \sup_{z\in\mathbb{B}_\nu} \sup_{\|w\|_\infty=1} \|A(z)w\|_\infty \\ &= \sup_{z\in\mathbb{B}_\nu} \|A(z)\|_\infty \end{aligned}$$

with the usual matrix ∞-norm given by the maximal absolute value row sum.

A-posteriori analysis

Let f_N be the N-th order polynomial approximation of f obtained by solving the homological equations (2.15) up to order N and let P_N be the N-th order polynomial defined by the same complex conjugate change of variables as in Equation (2.13). We now want to ascertain the quality of the approximation. The philosophy of the a-posteriori analysis is the following: given the approximate parametrization P_N, prove that there is an exact parametrization P nearby. More precisely we will try to find a parameter disk $V_\nu \subset \mathbb{R}^n$ and a $\delta > 0$ such that

$$\|P(\phi) - P_N(\phi)\|_\infty < \delta, \quad \text{for all } \phi \in V_\nu.$$

Theorem 4.2 in [60] provides numerically verifiable sufficient conditions under which this is possible. In the following we describe the necessary ingredients. Given an approximate solution $f_N : \mathbb{B}_\nu \subset \mathbb{C}^n \to \mathbb{C}^d$ to (2.11) fulfilling the constraints (2.12) with

$$f_N(z) = \sum_{|k|=0}^{N} b_k z^k, \tag{2.18}$$

we derive error bounds which by construction carry over to the real valued restriction given by (2.13). We define the following validation values.

Definition 2.1.7 (Validation Values) *The collection of positive constants ν, ϵ_{tol}, C_1, C_2, ρ', ρ and μ are called validation values if they possess the following properties.*

1. $\|g \circ f_N - Df_N\Lambda\|_{\Sigma,\nu} < \epsilon_{tol}$.
2. $\|f_N\|_\nu \leq \rho' < \rho$.

3. $\|Dg(f_N)\|_{M,\nu} \leq C_1$.

4. $\max\limits_{|\alpha|=2} \max\limits_{1\leq j\leq n} \sup\limits_{|z-p_0|\leq\rho} |\partial^\alpha g_j(p+z)| \leq C_2$.

5. $\max\limits_{1\leq i\leq n} Re(\lambda_i) < -\mu$.

It is important that all these values can be computed rigorously using interval arithmetic. The following theorem is the basis of the a-posteriori analysis. We use as shorthand notation

$$N_g = \max_{j=1,\dots,d} \#\{(k,l)|1 \leq k,l \leq d \text{ such that } \partial^k\partial^l g_j \not\equiv 0\},$$

and suppose we are given constants $N_{1,2}$ such that

$$\begin{aligned} N_1 &\geq N_g, \\ N_2 &\geq \frac{d(d+2)^{d+2}}{(d+1)^{d+1}}. \end{aligned}$$

Theorem 2.1.1 *[Theorem 4.2 in [60]] Suppose that for an approximation f_N in the sense of* (2.18) *we are given validation values as in Definition 2.1.7. Assume that N and δ fulfill*

$$\begin{aligned} (N+1)\mu - C_1 > 0&, \\ \delta &> \frac{2\epsilon_{tol}}{(N+1)\mu - C_1}, \\ \delta &< \min\left\{\frac{(N+1)\mu - C_1}{C_2N_1N_2}, \frac{\rho-\rho'}{d+2}\right\}. \end{aligned}$$

Then there exists a unique solution $f\colon \mathbb{B}_\nu \to \mathbb{C}^d$ to (2.11) *fulfilling the initial value constraints* (2.12) *such that*

$$\|f - f_N\|_\nu \leq \delta. \tag{2.19}$$

Furthermore the series coefficients for $|k| > N$ satisfy the growth bounds $\|a_k\|_\infty \leq \frac{\delta}{\nu^{|k|}}$.

In particular it follows from (2.19) that

$$\|P(\phi) - P_N(\phi)\|_\infty < \delta, \quad \text{for all } \phi \in V_\nu,$$

as we wished.

We remark that the proof in [60] actually shows that the truncation error $E(z) = f(z) - f_N(z)$ is itself an analytic function on $\mathbb{B}_\nu$ with $\|E\|_\nu \leq$

δ. In particular we know that $e(\phi) = P(\phi) - P_N(\phi)$ is a real analytic function with

$$\|e(\phi)\|_\infty \leq \delta \quad \text{for all } \phi \in V_\nu.$$

Furthermore we can use classical results from complex analysis in order to obtain bounds on the derivatives of the truncation error on smaller disks. As this observation is essential when formulating a-posteriori validation theorems for connecting orbits we give some further details.

Cauchy bounds The following result, whose proof can be found in [38], allows estimating the derivatives of an analytic function of several complex variables given a bound on the supremum of the function itself, but on strictly smaller domain disks. The size of the bounds depends on how much domain we are *willing to give up*. The proof relies on the multivariable Cauchy integral formula for derivatives of analytic functions of several complex variables.

Lemma 2.1.3 (Cauchy Bounds) *Suppose that $f : \mathbb{B}_\nu \subset \mathbb{C}^n \to \mathbb{C}^d$ is bounded and analytic. Then for any $0 < \sigma \leq 1$ we have for $i = 1, \ldots, n$ that*

$$\|\partial_i f\|_{\nu \exp(-\sigma)} \leq \frac{2\pi}{\nu\sigma}\|f\|_\nu \quad \textit{so that} \quad \|Df\|_{M,\nu \exp(-\sigma)} \leq \frac{2\pi n}{\nu\sigma}\|f\|_\nu,$$

as well as for $i, j = 1, \ldots, n$

$$\|\partial_i \partial_j f\|_{\nu \exp(-\sigma)} \leq \frac{4\pi^2}{\nu^2\sigma^2}\|f\|_\nu \quad \textit{so that} \quad \|D^2 f\|_{\nu \exp(-\sigma)} \leq \frac{4\pi^2 n^2}{\nu^2\sigma^2}\|f\|_\nu.$$

When some of the eigenvalues occur in complex conjugate pairs we require the following adaption of the Cauchy bounds.

Remark 2.1.3 ***Bounds on Derivatives When There Are Complex Conjugate Eigenvalue Pairs*** Recall that in the case of complex conjugate pairs of eigenvalues $\lambda_{j+1} = \overline{\lambda}_j$ of $Dg(p)$, the parameterization of the real stable manifold is given by the complex conjugate change of variable in Equation (2.13). Assume that we have m real eigenvalues and l complex conjugate pairs of eigenvalues with $n = m + 2l$. Next as the complex change of variables indicated by (2.13) can in particular be viewed as a composition of functions we need to apply the chain rule to adapt the Cauchy estimates. As a starting point let us treat the case where we only have one pair of complex conjugate eigenvalues, meaning $m = 0$, $l = 1$ and hence $n = 2$.

Then the parametrization $P : \mathbb{R}^2 \supset V_\nu \to \mathbb{R}^d$ is for $\phi = (\phi_1, \phi_2) \in \mathbb{R}^2$ given by $P(\phi) = P_N(\phi) + e(\phi) = f_N(z, \bar{z}) + E(z, \bar{z})$. In particular $E : \mathbb{C}^2 \supset \mathbb{B}_\nu \to \mathbb{C}^d$ has d component functions E_k $k = 1, \ldots, d$ inducing the d components of e.
In particular suppose for a component function $E_k \colon \mathbb{B}_\nu \subset \mathbb{C}^2 \to \mathbb{C}$, that $E_k(z, \bar{z}) \subset \mathbb{R}$ and $\|E_k(z, \bar{z})\|_\infty \leq \delta$ for each $z \in B_\nu$. Let $e_k \colon V_\nu \subset \mathbb{R}^2 \to \mathbb{R}$ be defined by $e_k(\phi_1, \phi_2) = E_k(\phi_1 + i\phi_2, \phi_1 - i\phi_2)$. For $j = 1, 2$ we have

$$\frac{\partial}{\partial \phi_j} e_k(\phi_1, \phi_2) = -i^{j+1} \left(\frac{\partial}{\partial z} E_k(z, \bar{z}) + (-1)^{j+1} \frac{\partial}{\partial \bar{z}} E_k(z, \bar{z}) \right). \tag{2.20}$$

Then for any $0 < \sigma \leq 1$ and $k = 1, \ldots, d$, applying Lemma 2.1.3 gives the bound

$$\left| \frac{\partial}{\partial \phi_j} e_k(\phi) \right| \leq \|\partial_1 E\|_{\nu \exp(-\sigma)} + \|\partial_2 E\|_{\nu \exp(-\sigma)} \leq \frac{4\pi}{\nu\sigma} \delta \tag{2.21}$$

for all $\phi \in V_\nu$. Taking another derivative requires applying the chain rule to both terms in Equation (2.20), leading to four terms which must be bounded, so that

$$\left| \frac{\partial^2}{\partial \phi_j \partial \phi_l} e_k(\phi) \right| \leq \frac{16\pi^2}{\nu^2 \sigma^2} \delta, \qquad \text{for} \qquad j, l = 1, 2 \tag{2.22}$$

and for all $\phi \in V_\nu$. Next we consider the case of general $n = m + 2l$. Setting for $A : \mathbb{R}^n \to \mathbb{R}^{n,d}$

$$\begin{aligned} |A|_{M,\nu} &= \sup_{\|\phi\|_{(m,l)} \leq \nu} \sup_{\|x\|_{(m,l)} = 1} \|A(\phi)x\|_\infty \\ &= \sup_{\|\phi\|_{(m,l)} \leq \nu} \|A(\phi)\|_\infty \end{aligned} \tag{2.23}$$

with the usual maximal row sum matrix norm $\| \cdot \|_\infty$ and applying the component-wise estimates given by Equations (2.21) and (2.22) gives

$$|DP|_{M, \nu \exp(-\sigma)} \leq |DP_N|_{M, \nu \exp(-\sigma)} + \frac{(2m + 4l) n \pi}{\nu \sigma} \delta. \tag{2.24}$$

Concerning the second derivative we get for every component function P_k $(k = 1, \ldots, d)$

$$|D^2 P_k|_{\nu \exp(-\sigma)} \leq |D(P_N)_k|_{\nu \exp(-\sigma)} + \frac{(2m + 4l)^2 n^2 \pi}{\nu^2 \sigma^2} \delta \tag{2.25}$$

for any loss of domain parameter $0 < \sigma \leq 1$. We note that for the applications to two dimensional stable and unstable manifolds associated with a single complex conjugate pair of eigenvalues we have that $m = 0$ and $l = 1$. The example we consider using the Lorenz equation will be of this type.

2.2 Spline and spectral approximation

In this section we review two different numerical approximation techniques for real-valued functions $u : I \to \mathbb{R}$ defined on an interval $I \subset \mathbb{R}$. That is we look for a function u_h that can be described by a finite amount of variables such that the approximation error

$$\|u - u_h\|$$

is controllable in a mathematically concise way. We consider two philosophies to obtain the approximation u_h:

- Representation by many piecewise polynomials of low degree with local support. This strategy is classically referred to as **spline approximation**.
- Representation by one polynomial of high degree defined on the whole interval obtained by truncation of a Chebyshev series. This strategy is classically referred to as **spectral approximation**.

We remark that there exist different other approaches to this approximation problem, like for example meshfree methods [20] and wavelets [12, 61] which we will nevertheless not consider in this thesis.

We will restrict our attention to linear splines for the spline approach, where our main source is [50]. The spectral approach is based on the truncation of a Chebyshev series expansion. As references see for example [57] and [56]. While we refer to [50] for the details on the spline approximation, we give some more details about the spectral approximation with an emphasis on the application of discrete convolution techniques. In particular we give an overview of the convolution estimates from [5] and [25].

2.2.1 Spectral approximation

Let $I = [-1, 1]$. The central result in this context, which can be found together with its proof in [57], provides an analogue of the Fourier expansion for non-periodic functions on an interval.

Theorem 2.2.1 *Every Lipschitz continuous function $u : [-1, 1] \to \mathbb{R}$ has a unique representation as an absolutely convergent series of the form*

$$u(t) = \sum_{k=0}^{\infty} a_k T_k(t) \tag{2.26}$$

where the polynomial basis T_k $(k \geq 0)$ is given by the two term recursion

$$T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t) \tag{2.27}$$

with $T_0(t) = 1$ and $T_1(t) = t$.

The coefficients $(a_k)_{k \in \mathbb{N}}$ can be explicitly computed by

$$a_k = \frac{2}{\pi} \int_{-1}^{1} \frac{u(t)T_k(t)}{\sqrt{1-t^2}} dt \tag{2.28}$$

for $k \geq 1$ and for $k = 0$ the coefficient $\frac{2}{\pi}$ has to be changed to $\frac{1}{\pi}$. The degree k polynomials $T_k(t)$ are called Chebyshev polynomials (of 1st kind).

Before proceeding further we recall some basic properties of the Chebyshev polynomials that we wish to refer to later.

Lemma 2.2.1 *Let $T_k(t) : [-1, 1] \to \mathbb{R}$ denote the kth Chebyshev polynomials of first kind. Then the following statements are true:*

1.

$$T_k(-1) = (-1)^k \qquad\qquad T_k(1) = 1 \tag{2.29}$$

2.

$$\begin{aligned} &\int T_0(s)ds = T_1(s), \quad \int T_1(s)ds = (T_2(s) + T_0(s))/4 \\ &\int T_k(s)ds = \frac{1}{2}\left(\frac{T_{k+1}(s)}{k+1} - \frac{T_{k-1}(s)}{k-1}\right) \ \textit{for} \ k \geq 2 \end{aligned} \tag{2.30}$$

For a proof see [4].

By a rescaling of the coefficients a_k for $k \geq 1$ we assume for technical reasons a Chebyshev series expansion of the form

$$u(t) = a_0 + 2\sum_{k=1}^{\infty} a_k T_k(t). \tag{2.31}$$

To obtain the numerical approximation we use a Galerkin projection. That is, given a dimension m we define

$$u_h(t) = a_0 + 2\sum_{k=1}^{m-1} a_k T_k(t).$$

We identify the function u_h with the m-dimensional vector $(a_0, \ldots, a_{m-1})$ of its Chebyshev coefficients and obtain a finite dimensional object that we can manipulate numerically.

The numerical usefulness of this expansion becomes more evident if we consider the following equivalent characterization of the Chebyshev polynomials. For $t \in [-1, 1]$ we have

$$T_k(t) = \cos(k\theta) \quad \text{where} \quad t = \cos(\theta) \text{ for a } \theta \in \mathbb{R}. \tag{2.32}$$

In this form it is less obvious that $T_k(t)$ is a degree k polynomial. Without giving the details of the derivation, which can be found in [56], we indicate the foundation for this fact. Given a $t \in [-1, 1]$ we can find a complex number $z \in \mathbb{S}^1 \subset \mathbb{C}$ in the unit circle such that $t = \frac{1}{2}(z + z^{-1})$ and hereby obtain

$$T_k(t) = \cos(k\theta) = \frac{1}{2}(z^k + z^{-k}).$$

Then we directly check that

$$\begin{aligned} 2tT_k - T_{k-1}(t) &= 2\cos(\theta)\cos(k\theta) - \cos((k-1)\theta) \\ &= 2\frac{1}{2}(z + z^{-1})\frac{1}{2}(z^k + z^{-k}) - \frac{1}{2}(z^{k-1} + z^{1-k}) \\ &= \frac{1}{2}(z^{k+1} + z^{-k-1}) = \cos((k+1)\theta) = T_{k+1}(t). \end{aligned} \tag{2.33}$$

(2.32) obviously yields $T_0(t) = 1$ and $T_1(t) = t$ and together with (2.33) we have that $T_k(t) = \cos(k\theta)$ with $t = \cos(\theta)$ defines a degree k polynomial in t.

The benefit of (2.32) is that the series expansion (2.26) can be identified as a Fourier series in disguise. This central fact heralds the application of the machinery of analytic estimates introduced in [24, 67, 13, 14] and the Banach space of rapidly decaying coefficients used in [58, 24] in our non-periodic problem setting. Thus let us be more precise. If we define

$$\tilde{a}_k = \begin{cases} a_k & k \geq 0 \\ a_{-k} & k < 0 \end{cases}$$

we obtain

$$u(t) = a_0 + 2\sum_{k=1}^{\infty} a_k T_k(t) = a_0 + 2\sum_{k=1}^{\infty} a_k \frac{1}{2}(e^{ik\theta} + e^{-ik\theta}) = \sum_{k=-\infty}^{\infty} \tilde{a}_k e^{ik\theta}. \tag{2.34}$$

By (2.34) we can derive the following central lemma.

Lemma 2.2.2 *Let $u : [-1,1] \to \mathbb{R}$ and $v : [-1,1] \to \mathbb{R}$ be two Lipschitz continuous functions with corresponding Chebyshev expansions*

$$u(t) = a_0 + 2\sum_{k=1}^{\infty} a_k T_k(t) \qquad v(t) = b_0 + 2\sum_{k=1}^{\infty} b_k T_k(t)$$

such that their product $uv : [-1,1] \to \mathbb{R}$ is Lipschitz continuous. Then

$$uv = (a*b)_0 + 2\sum_{k=1}^{\infty}(a*b)_k T_k(t)$$

where $(a*b)_k$ is given for $k \geq 0$ by the discrete convolution sum

$$(a*b)_k = \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} a_{|k_1|} b_{|k_2|}. \tag{2.35}$$

Proof 2.2.1 *Following* (2.34) *we compute*

$$uv(t) = \left(\sum_{k=-\infty}^{\infty} \tilde{a}_k e^{ik\theta}\right)\left(\sum_{k=-\infty}^{\infty} \tilde{b}_k e^{ik\theta}\right) = \sum_{k=-\infty}^{\infty} (\tilde{a}*\tilde{b})_k e^{ik\theta}$$

where

$$(\tilde{a}*\tilde{b})_k = \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} \tilde{a}_{k_1}\tilde{b}_{k_2}$$

which is the classical convolution sum for Fourier series that can be found for example in [53]. By definition of $\tilde{a}_k$ and $\tilde{b}_k$ we obtain for $k \geq 0$

$$\sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} \tilde{a}_{k_1}\tilde{b}_{k_2} = \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} a_{|k_1|} b_{|k_2|}. \tag{2.36}$$

Using the fact that by construction $\tilde{a}_k = \tilde{a}_{-k}$ and $\tilde{b}_k = \tilde{b}_{-k}$ we get that $(\tilde{a}*\tilde{b})_k = (\tilde{a}*\tilde{b})_{-k}$ and thus using (2.36)

$$\begin{aligned}
uv(t) &= \sum_{k=-\infty}^{\infty} (\tilde{a}*\tilde{b})_k e^{ik\theta} = (\tilde{a}*\tilde{b})_0 + \sum_{k=1}^{\infty}(\tilde{a}*\tilde{b})_k (e^{ik\theta} + e^{-ik\theta}) \\
&= (a*b)_0 + \sum_{k=1}^{\infty}(a*b)_k (e^{ik\theta} + e^{-ik\theta}) = \\
&= (a*b)_0 + 2\sum_{k=1}^{\infty}(a*b)_k \frac{1}{2}(e^{ik}\theta + e^{-ik\theta}) \\
&= (a*b)_0 + 2\sum_{k=1}^{\infty}(a*b)_k \cos(k\theta) = (a*b)_0 + 2\sum_{k=1}^{\infty}(a*b)_k T_k(t).
\end{aligned}$$

□

Thus by representing two sufficiently smooth functions in Chebyshev basis T_k their multiplication can be viewed as convolution of their Chebyshev coefficient sequences $(a_k)_{k\in\mathbb{N}}$.

As a further analogue to Fourier series the question about the decay rates of these coefficients a_k is of crucial importance. Like in the case of Fourier series an overall rule of thumb is: the smoother a function the faster the coefficients decay. In our context the functions u we wish to expand are components of solutions to differential equations of the form (2.1) where the vector field $g : \mathbb{R}^d \to \mathbb{R}^d$ is real analytic. This in particular implies that u is a real analytic function which in turn entails geometric decay of the Chebyshev coefficients. As this is of central importance let us state the main theorem in this context that again can be found in [57].

Theorem 2.2.2 *Let a function $u : [-1, 1] \to \mathbb{R}$ be real analytic and analytically continuable onto the ρ-ellipse E_ρ for some $\rho > 1$ and let it be bounded on E_ρ by some $R > 0$ that is $|u(z)| \leq R$ for all $z \in E_\rho$ then*

$$|a_k| \leq 2R\rho^{-k} \tag{2.37}$$

with $|a_0| \leq R$. This decay behaviour $O(\rho^{-k})$ for $k \to \infty$ is referred to as geometric decay.

Proof 2.2.2 *See [57].*

Remark 2.2.1

1. *The ρ ellipse is a classical object in approximation theory and is defined as follows: fix a $\rho > 1$ and consider the image of the circle with radius ρ in the complex plane $\mathbb{C}$ under the Joukowski map $w = \frac{1}{2}(z + z^{-1})$. This is an ellipse with foci at ± 1 which is particularly suitable for our situation as we consider analytic continuations of functions defined on $[-1, 1]$.*
2. *If two functions u and v are analytic, then so is their product. Hence we are obviously in the position to use Lemma 2.2.2.*

Furthermore we can derive the following elementary bound on the approximation error in the infinity norm $\|u - u_h\|_\infty \overset{\text{def}}{=} \sup_{t\in[-1,1]} |u - u_h(t)|$.

Lemma 2.2.3 *Let the conditions of Theorem 2.2.2 be fulfilled then we can estimate for any $m > 1$*

$$\|u - u_h\|_\infty \leq \frac{4R}{\rho - 1}\rho^{-m+1}$$

Proof 2.2.3 *We directly compute for $t \in [-1,1]$ and $m > 1$ arbitrary but fixed that*

$$|u(t) - u_h(t)| \leq 2 \sum_{k=m}^{\infty} \underbrace{|a_k|}_{\leq 2R\rho^{-k}} \underbrace{|T_k(t)|}_{\leq 1} \leq 4R \sum_{k=m}^{\infty} \rho^{-k}$$
$$= 4R \left(\frac{1}{1-\frac{1}{\rho}} - \frac{1-\left(\frac{1}{\rho}\right)^m}{1-\frac{1}{\rho}} \right) = \frac{4R\rho^{-m+1}}{\rho - 1}.$$

Taking the supremum over all $t \in [-1,1]$ yields the result. □

It is a well-known fact that geometric decay of a sequence $(a_k)_{k\in\mathbb{N}}$ implies that it is also algebraically decaying. More precisely, defining the weights

$$\omega_k = \begin{cases} |k| & k \neq 0 \\ 1 & k = 0 \end{cases} \tag{2.38}$$

we have the following implication: given a $\rho > 1$

$$\sup_{k\geq 0} |a_k|\rho^k < \infty \Rightarrow \sup_{k\geq 0} |a_k|\omega_k^s < \infty$$

for all algebraic decay rates $s > 1$. Thus we know that the Chebyshev coefficient sequence of an analytic function is algebraically decaying for all decay rates $s > 1$. This motivates the definition of the space of algebraically decaying sequences

$$\Omega^s = \{(a_k)_{k\in\mathbb{N}} : \sup_{k\geq 0} |a_k|\omega_k^s < \infty\}. \tag{2.39}$$

Moreover if we define for $a \in \Omega^s$

$$\|a\|_{\Omega^s} = \sup_{k\geq 0} |a_k|\omega_k^s$$

the pair $(\Omega^s, \| \|_{\Omega^s})$ is a Banach space. Inspired by Lemma 2.2.2 we add the additional operation of discrete convolution of two sequences $a_1, a_2 \in \Omega^s$ denoted by $*$ and defined in (2.35). Then $(\Omega^s, \| \|_{\Omega^s}, *)$ becomes an algebra. More concretely this means that given two sequences $a, b \in \Omega^s$ there is a constant $C = C(s)$ such that

$$\|a * b\|_{\Omega^s} \leq C\|a_1\|_{\Omega^s}\|a_2\|_{\Omega^s}.$$

Constructive proofs of this fact can be found for the case $1 < s < 2$ in [5] and for the case $s \geq 2$ in [25]. In particular concrete values of the constants $C(s)$ can be derived from these results. It will turn out in the sequel that it is indispensable in our approach that these estimates are as sharp as possible. We will therefore give some details on the strategy of their derivation used in [25].

Fixing a decay rate $s \geq 2$ we have to bound

$$\sup_{k \geq 0} |(a_1 * a_2)_k| \omega_k^s,$$

where $(a * b)_k$ is given by (2.35). The following Lemma, corresponding to Lemma **A.2** in [25] leads us the way to achieving this task. Following [25] we first define

$$\gamma_M(s) = 2\left[\frac{M}{M-1}\right]^s + \left[\frac{4\ln(M-2)}{M} + \frac{\pi^2-6}{3}\right]\left[\frac{2}{M}+\frac{1}{2}\right]^{s-2}.$$

Lemma 2.2.4 *Assume two sequences $a_{1,2} \in \Omega^s$ to be given and set $A_{1,2} = \|a_{1,2}\|_s$. Let $M \in \mathbb{N}$ with $M \geq 6$. Define $\alpha_0^2, \ldots, \alpha_M^2$ by*

$$\alpha_k^2 \overset{\text{def}}{=} \begin{cases} 1 + 2\sum_{k_1=1}^M \frac{1}{\omega_{k_1}^{2s}} + \frac{2}{M^{2s-1}(s-1)} & k = 0 \\ \sum_{k_1=1}^M \frac{2\omega_k^s}{\omega_{k_1}\omega_{k+k_1}} + \frac{2\omega_k^s}{(k+M+1)^s M^{s-1}(s-1)} + 2 + \sum_{k_1=1}^{k-1} \frac{\omega_k^s}{\omega_{k_1}^s \omega_{k-k_1}^s} & 1 \leq k \leq M-1\,. \\ 2 + 2\sum_{k_1=1}^M \frac{1}{\omega_{k_1}^s} + \frac{2}{M^{s-1}(s-1)} + \gamma_M(2) & k \geq M \end{cases} \tag{2.40}$$

Then we have that for $k \geq 0$

$$\sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} \frac{1}{\omega_{k_1}^s \omega_{k_2}^s} \leq \frac{\alpha_k^2}{\omega_k^s}$$

and hence

$$|(a_1 * a_2)_k| \leq A_1 \cdot A_2 \frac{\alpha_k^2}{\omega_k^s} \tag{2.41}$$

Proof 2.2.4 *[25]*

Using Lemma 2.2.4 we can set

$$C = \max_{k=0,\ldots,M} \alpha_k^2$$

and obtain that $(\Omega^s, \|\|_{\Omega^s}, *)$ is an algebra.

Remark 2.2.2 *If one sets*

$$\|\cdot\|_s = C\|\cdot\|_{\Omega^s},$$

then $(\Omega^s), \|\cdot\|_s$ *becomes a Banach algebra. This can be seen by realizing that for* $a_{1,2} \in \Omega^s$ *we have*

$$\begin{aligned}\|a_1 * a_2\|_s &= C\|a_1 * a_2\|_{\Omega^s} \leq C^2\|a_1\|_{\Omega^s}\|a_2\|_{\Omega^s}\\ &= \big(C\|a_1\|_{\Omega^s}\big)\big(C\|a_2\|_{\Omega^s}\big) = \|a_1\|_s\|a_2\|_s.\end{aligned}$$

Taking the next step from Lemma 2.2.2 leads to the consideration of products of n functions $u_1, \ldots, u_n$, induced for example by nth order polynomial nonlinearities in the vector field g in (2.1). We therefore extend the above reasoning to the estimation of the norm of convolution terms of the form $a_1 * \ldots * a_n$ for $n \geq 2$.
As this is a central aspect we wish to refer to later we explain the strategy used in [25] to obtain these estimates in more detail. Choosing $n \geq 3$ and $M \geq 6$ the main ingredient is to inductively construct $\alpha_0^n, \ldots, \alpha_M^n$ fulfilling

$$\sum_{\substack{k_1+\ldots+k_n=k\\ k_i\in\mathbb{Z}}} \frac{1}{\omega_{k_1}^s \cdots \omega_{k_n}^s} \leq \frac{\alpha_k^n}{\omega_k^s}$$

for $0 \leq k \leq M-1$ and

$$\sum_{\substack{k_1+\ldots+k_n=k\\ k_i\in\mathbb{Z}}} \frac{1}{\omega_{k_1}^s \cdots \omega_{k_n}^s} \leq \frac{\alpha_M^n}{\omega_k^s}$$

for $k \geq M$. The following Lemma corresponding to Lemma **A.3** in [25] effectuates this strategy.

Lemma 2.2.5 *Assume sequences* $a_1, \ldots, a_n \in \Omega^s$ *for a given* $n \geq 3$ *to be given and set* $A_i = \|a_i\|_{\Omega^s}$ *for* $i = 1, \ldots, n$. *Let* $M \geq 6$ *and define* $\alpha_0^n, \ldots, \alpha_M^n$ *by*

$$\alpha_k^n = \begin{cases} \alpha_0^{n-1} + 2\sum_{k_1=1}^{M-1} \frac{\alpha_{k_1}}{\omega_{k_1}^{2s}} + \frac{2\alpha_M^{n-1}}{(M-1)^{2s-1}(2s-1)} & k=0\\ \sum_{k_1=1}^{M-k} \frac{\alpha_{k+k_1}^{n-1}\omega_k^s}{\omega_{k_1}^s\omega_{k+k_1}^s} + \frac{\alpha_M^{n-1}\omega_k^s}{(M+1)^s(M-k)^{s-1}(s-1)} + \sum_{k_1=1}^{k-1} \frac{\alpha_{k_1}^{n-1}\omega_k^s}{\omega_{k_1}^s\omega_{k-k_1}^s} + & \\ \sum_{k_1=1}^{M} \frac{\alpha_{k_1}^{n-1}\omega_k^s}{\omega_{k_1}^s\omega_{k+k_1}^s} + \frac{\alpha_M^{n-1}\omega_k^s}{(M+k+1)^s(M)^{s-1}(s-1)} + \alpha_k^{n-1} + \alpha_0^{n-1} & \\ & 1 \leq k \leq M-1\\ \alpha_M^{(n-1)}\sum_{k_1=1}^{M} \frac{1}{\omega_{k_1}^s} + \frac{2\alpha_M^{n-1}}{M^{s-1}(s-1)} + \Sigma^* + \sum_{k_1=1}^{M} \frac{\alpha_{k_1}^{(n-1)}}{\omega_{k_1}^s} + & \\ \alpha_M^{n-1} + \alpha_0^{n-1} & k \geq M \end{cases} \tag{2.42}$$

where Σ^* *is defined by* $\Sigma^* \overset{\text{def}}{=} \min(\Sigma^a, \Sigma^b)$ *with*

$$\begin{aligned}\Sigma^a &= \sum_{k_1=1}^{M-1} \frac{\alpha_{k_1}^{n-1} M^s}{\omega_{k_1}^s (M-k_1)^s} + \alpha_M^{n-1}(\gamma_M - \sum_{k_1=1}^{M-1} \frac{1}{\omega_{k_1}^s}) \\ \Sigma^b &= \gamma_M \max_{k=0,\ldots,M} \alpha_k^{n-1}\end{aligned}$$

Then

$$\sum_{\substack{k_1+\ldots+k_n=k \\ k_i \in \mathbb{Z}}} \frac{1}{\omega_{k_1}^s \cdots \omega_{k_n}^s} \leq \frac{\alpha_k^n}{\omega_k^s}$$

and hence

$$|(a_1 * \ldots * a_n)_k| \leq A_1 \cdots A_n \frac{\alpha_k^n}{\omega_k^s} \tag{2.43}$$

for all $k \geq 0$.

Proof 2.2.5 *See [25]*

Chapter 3

Rigorous numerics for connecting orbits

Our strategy for the verification of connecting orbits between hyperbolic fixed points of (2.1) is to consider an equivalent nonlinear operator equation

$$F(x) = 0 \qquad x \in X, \tag{3.1}$$

where X is a Banach space and F is a Fréchet differentiable operator and develop methods to validate solutions of these equations. The basis for this approach is contained in equation (2.3). The common goal of all methods is, given an approximate solution $\bar{x}$ of the corresponding equation (3.1), to find a ball $B_r(\bar{x})$ in the respective Banach space around the approximate solution in which a genuine solution $\tilde{x}$ is guaranteed to exist.

We start with an algorithm that we term discretization free approach as it is based on the intersection of the local stable and unstable manifold with respect to some particularly chosen neighborhoods and lends itself to an equivalent finite dimensional operator equation without discretization of the flow induced by (2.1). We validate solutions to this equation using the Newton-Kantorovich Theorem.

In the case when these local manifolds do not intersect the corresponding equation (3.1) is infinite dimensional and necessitates the discretization of the flow induced by (2.1). While the Newton-Kantorovich Theorem would be applicable in principle in this situation the infinite dimensionality induces substantial technical hurdles that we are able to circumvent by the method of radii polynomials originally designed to validate equilibria of PDEs (e.g. see [15]). In particular it provides an efficient means of controlling the error induced by infinite dimensionality. Concerning the

discretization we follow two different paths. First we utilize linear splines, in a similar spirit as in [60], and secondly we use a spectral method based on the truncation of Chebyshev series.
We finish this section by discussing results related to transversality of the connecting orbits we validate.

3.1 Discretization free approach

3.1.1 General validation method using Newton-Kantorovich

The Newton-Kantorovich Theorem is a classical result in nonlinear analysis giving information about the convergence behavior of the Newton iteration. Stated in the following way we can use it to validate solutions to (3.1) in the sense that we find a ball around an approximate solution $\bar{x}$ in which we can guarantee a unique genuine solution $\tilde{x}$ to exist.

Theorem 3.1.1 (Newton-Kantorovich Theorem) *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces and $F : X \to Y$ be a Fréchet differentiable mapping. Consider $\bar{x} \in X$, $\bar{r} > 0$ and $B_{\bar{x}}(\bar{r}) \subset X$ the closed ball of radius $\bar{r}$ centered at $\bar{x}$. Let $B(X,Y)$ be the space of bounded linear operators on X with the operator norm $\|\cdot\|_{B(X)}$ and similarly $B(Y,X)$. Assume that*

(i) $DF(\bar{x})$ has a bounded inverse, and

(ii) $\|DF(x) - DF(y)\|_{B(X,Y)} \leq \kappa\|x - y\|_X$ for all $x, y \in B_{\bar{x}}(\bar{r})$,

for $\kappa \geq 0$. If

(I)

$$\epsilon_{NK} \geq \|DF(\bar{x})^{-1} F(\bar{x})\|_X,$$

(II)

$$\epsilon_{NK} \leq \frac{\bar{r}}{2},$$

and

(III)

$$4\epsilon_{NK}\,\kappa\,\|DF(\bar{x})^{-1}\|_{B(Y,X)} \leq 1,$$

then there is a unique $\tilde{x} \in B_{\bar{x}}(\bar{r})$ so that $F(\tilde{x}) = 0$.

A proof of the Newton-Kantorovich Theorem can be found in [45]. Note that the connection to the Newton iteration is suggested in the following way. Thinking of the classical Newton operator

$$T(x) = x - DF^{-1}(x)F(x)$$

to find a zero of a differentiable map f and assuming that we start the iteration with $x^0 = \bar{x}$, then we realize that the difference between the first iterate $T(\bar{x})$ and the initial point $\bar{x}$ is exactly

$$\bar{x} - DF(\bar{x})^{-1}F(\bar{x}) - \bar{x} = DF(\bar{x})^{-1}F(\bar{x})$$

and thus ϵ_{NK} measures this initial defect. Instead of giving more details on the proof we make the following remark to clarify the strategy to apply Theorem 3.1.1 to validation purposes.

Remark 3.1.1 *In order to use Theorem 3.1.1 to validate a numerical approximation to a solution of* (3.1) *defined on a finite dimensional spaces X and Y we take the following steps:*

1. *Check if $\|DF^{-1}(\bar{x})\|_{B(Y,X)}$ is bounded.*

2. *Compute ϵ_{NK} such that*
$$\|DF(\bar{x})^{-1}F(\bar{x})\|_X \leq \epsilon_{NK} \tag{3.2}$$

3. *Set*
$$\bar{r} = 2\epsilon_{NK}. \tag{3.3}$$

4. *Compute κ for this $\bar{r}$.*

5. *Check if*
$$4\epsilon_{NK}\kappa\|DF^{-1}(\bar{x})\|_{B(Y,X)} \leq 1. \tag{3.4}$$

By using interval arithmetic we can check the strict inequality in (3.4). The benefit is that the strict inequality implies invertibility of $DF(\bar{x})$. (see Lemma 3.3.1). In the sequel this fact will be connected to the transversality of the intersection of the stable and unstable manifold.

In order to illustrate the basic mechanics involved in applying Theorem 3.1.1 for validation purposes we consider the following elementary example. We stir the reader's attention to the following two facts, clearly visible in this easy example:

- The quality of the numerical approximation, i.e. the magnitude of ϵ_{NK} is crucial for the success of the validation

- The central technical tool to compute κ is the Mean Value Theorem

Example 3.1.1 *We consider an iterative approximation of $\sqrt[3]{3}$. Realize that the boundedness requirements condense to scalar nonzero conditions in this context. A Matlab implementation can be found in the software accompanying this thesis. In this (trivial) illustration we obtain*

$$F(x) = x^3 - 3 \tag{3.5}$$

resulting in the iteration

$$x_{k+1} = x_k - \frac{x_k^3 - 3}{3x_k^2} \tag{3.6}$$

with an appropriate initial condition $x_0 \in \mathbb{R}$. Assuming an approximate zero $\bar{x} = x_{k_{end}}$ of (3.5) for some iteration number k_{end} let us check the assumptions of Theorem 3.1.1.

- $$DF(\bar{x})^{-1} \text{ bounded } \Leftrightarrow \frac{1}{3\bar{x}^2} < \infty. \tag{3.7}$$

- *Assuming*

  $$\bar{r} \geq 2\epsilon_{NK} \tag{3.8}$$

  where

  $$|\frac{\bar{x}^3 - 3}{3\bar{x}^2}| \leq \epsilon_{NK} \tag{3.9}$$

  we demand from $\kappa > 0$ that

  $$|3x^2 - 3y^2| \leq \kappa|x - y| \tag{3.10}$$

  for x, y such that $|x - \bar{x}|, |y - \bar{x}| < \bar{r}$.

By applying the mean value theorem we can compute κ symbolically in this case. We readily estimate for all $x < y$ with $|x - \bar{x}|, |y - \bar{x}| < \bar{r}$

$$|3x^2 - 3y^2| \leq 6\xi|x - y| \leq \kappa|x - y|, \tag{3.11}$$

where $\xi \in (x, y)$ and $\kappa = 6(\bar{x} + \bar{r})$. We explicitely see that the bigger ϵ_{NK} the bigger r and the bigger κ will be. To complete the proof we thus have to check for the above constructed constant ϵ_{NK} and κ that we have

$$4\epsilon_{NK}\kappa|\frac{1}{3\bar{x}^2}| \leq 1. \tag{3.12}$$

The point distinguishing the finite from the infinite dimensional setting is that in the finite dimensional case we can estimate ϵ_{NK} and κ directly by evaluating F and DF using interval arithmetic. When working on infinite dimensional spaces the additional error produced by discretizing X has to be taken into account. Hence, it is more difficult to get estimates for ϵ_{NK} and κ, as we can neither evaluate the operator F nor its derivative DF numerically. One strategy to cope with this problem raised by infinite dimensionality is to use the method of radii polynomials that we shall describe in Section 3.2. Now we shall go on to describe how we use Theorem 3.1.1 to validate connecting orbits.

3.1.2 Validation of connecting orbits

We would like to use the above presented procedure to validate connecting orbits between two hyperbolic equilibria $p_{1,2}$ of (2.1). The idea is to look for intersections of the local stable and unstable manifolds which can be encoded as a zero of an operator F defined on a finite dimensional space. Let us start with the derivation of F before we turn to the validation technique.

Definition of F

Let p_2 have stable dimension n_s and denote by $P : \mathbb{R}^{n_s} \supset V_{\nu_s} \to \mathbb{R}^d$ a parametrization of the local stable manifold of p_2. Similarly for p_1 we assume the unstable dimension to be n_u and set $Q : \mathbb{R}^{n_u} \supset V_{\nu_u} \to \mathbb{R}^d$ to be a parametrization of the local unstable manifold of p_1. Additionally we take the non-degeneracy condition $n_u + n_s = d + 1$ as given. Denote the parameter in V_{ν_u} as φ and in V_{ν_s} as ϕ. As described in (2.3) an orbit is a connecting orbit if it lies in the intersection of $W^u(p_1)$ and $W^s(p_2)$. Thus if we find $\tilde{\varphi} \in V_{\nu_u}$ and $\tilde{\phi} \in V_{n_s}$ such that $Q(\tilde{\varphi}) = P(\tilde{\phi}) \overset{\text{def}}{=} q$ then

$$q \in W^s(p_2) \cap W^u(p_1)$$

and hence $O(q)$ is a connecting orbit from p_1 to p_2. As a consequence setting

$$F(\varphi, \phi) = Q(\varphi) - P(\phi) \tag{3.13}$$

we obtain that the existence of $\tilde{x} = (\tilde{\varphi}, \tilde{\phi}) \in \mathbb{R}^{d+1}$ such that $F(\tilde{x}) = 0$ is equivalent to the existence of a connecting orbit from p_1 to p_2. However $F : \mathbb{R}^{d+1} \to \mathbb{R}^d$ and we can not expect F to have isolated zeros. As the Newton-Kantorovich Theorem by construction detects isolated zeros, we need to impose a phase condition in order to ensure isolation of the zeros we seek for.

Phase condition The idea of the phase condition is to fix one of the parameters φ or ϕ to lie on a prescribed co-dimension one submanifold of the parameter space and hereby reduce the number of independent variables by one. We choose this to be done in the unstable parameter space emphazising that this is a choice that can just as well be made in favor of the stable parameter space. Let

$$\Theta_\nu \colon B_1 \subset \mathbb{R}^{n_u - 1} \to \operatorname{int}(V_{\nu_u}) \subset \mathbb{R}^{n_u} \tag{3.14}$$

be an immersion of the $(n_u - 1)$-sphere of radius ν, where we let B_1 be the unit euclidean ball of radius 1. Moreover we require that image(Θ_ν) is transverse to the linear vector field Λ_u in unstable parameter space. This transversality condition insures that for any $\alpha \in B_1$ the columns of $D\Theta_\nu(\alpha)$ and the single vector $\Lambda_u \Theta_\nu(\alpha)$ are a linearly independent set of vectors which span $\mathbb{R}^{n_u}$. Then the columns of $DP[\Theta_\nu(\alpha)]\Theta_\nu(\alpha)$ and the vector $DP[\Theta_\nu(\alpha)]\Lambda_u\Theta_\nu(\alpha)$ span $T_{P[\Theta_\nu(\alpha)]}W^u(p_1)$. Note that since the dynamics on the manifold are conjugate to the linear dynamics in parameter space we know that $DP[\Theta_\nu(\alpha)]\Lambda_u\Theta_\nu(\alpha)$ is the tangent vector to the orbit through $P[\Theta_\nu(\alpha)]$. Then the columns of $DP[\Theta_\nu(\alpha)]D\Theta_\nu(\alpha)$ span the subspace of $T_{P[\Theta_\nu(\alpha)]}W^u(p_1)$ transverse to the orbit.
We thus redefine F by

$$F \colon B_1 \times V_{\nu_s} \subset \mathbb{R}^{n_u - 1} \times \mathbb{R}^{n_s} = \mathbb{R}^d \to \mathbb{R}^d \tag{3.15}$$

given by

$$F(\alpha, \phi) = Q[\Theta_\nu(\alpha)] - P(\phi). \tag{3.16}$$

with the effect that F maps $\mathbb{R}^d$ into itself and the Newton-Kantorovich technique is applicable.

In order to make F amenable to a numerical treatment we choose orders N, M for the finite sum approximations P_N and Q_M to the infinite sum expressions of the parametrizations P and Q. Recall that P_N is derived from a complex valued extension $f_N : \mathbb{C}^{n_s} \to \mathbb{C}^d$ as defined in (2.16) and further explained in Section 2.1.2. In a similar way we assume Q_M to be derived from a complex valued map $h_M : \mathbb{C}^{n_u} \to \mathbb{C}^d$. Next we assume $V_{\nu_{u,s}}$ to be chosen as in Theorem 2.1.1 which in particular entails that

$$P = P_N + e_s \qquad Q = Q_M + e_u$$

with analytic tail errors $e_{s,u} : \mathbb{R}^{n_{s,u}} \to \mathbb{R}^d$, where $\|e_u(\varphi)\|_\infty \leq \delta_u$ for all $\varphi \in V_{\nu_u}$ and $\|e_s(\phi)\|_\infty \leq \delta_s$ for all $\phi \in V_{\nu_s}$. This enables us to express F as

$$\begin{aligned} F(\alpha, \phi) &= Q_M(\Theta(\alpha)) + e_u(\Theta(\alpha)) - P_N(\phi) - e_s(\phi) \\ &= Q_M(\Theta(\alpha)) - P_N(\phi) + e_u(\Theta(\alpha)) - e_s(\phi) \\ &\overset{\text{def}}{=} F_{N,M}(\alpha, \phi) + E(\alpha, \phi). \end{aligned} \tag{3.17}$$

We point out that in order to use the Newton-Kantorovich Theorem 3.1.1 it is indispensable to have control over the derivatives of F up to second order. This clarifies the importances of the Cauchy estimates introduced in Lemma 2.1.3 as these can be used to estimate the derivatives of E for which we only ever have an upper bound on $\|E(\alpha, \phi)\|_\infty$. We recall in this context that Cauchy bounds enable to estimate derivatives of an analytic function f on a certain domain only using knowledge about the norm of f.

Validation of solution to $F(\alpha, \phi) = 0$

The idea for the Newton-Kantorovich based validation is described in Remark 3.1.1. We now aim for an implementation of this technique to the validation of approximate solutions to $F(x) = 0$ with F given in (3.16). The following remark describes an important tool in this process.

Remark 3.1.2 (Neumann series) *The Neumann series makes a statement about the invertibility of a certain linear operator on a Banach space X. More specifically let T be a linear operator on X and let $\|\cdot\|$ be a norm on the space of linear operators on X such that it becomes a Banach space. Then if $\|T\| < 1$ then $\mathrm{I} - T$ is invertible and*

$$\|(\mathrm{I} - T)^{-1}\| \leq \frac{1}{1 - \|T\|}.$$

This is a direct analogue of the classical geometric series.

Next assume an approximate zero $\bar{x} = (\bar{\alpha}, \bar{\phi})$ to be given. We will take steps 1 to 5 from Remark 3.1.1 to build a ball $B_{\bar{x}}(\bar{r})$ of radius $\bar{r}$ around $\bar{x}$ where we can guarantee a genuine zero $\tilde{x}$ to exist. First we need to be exact with the norms we use in the respective spaces. Recall that we suppose that p_1 has n_u unstable eigenvalues where m_u are real and l_u are complex conjugate pairs and similarly p_2 has n_s stable eigenvalues with m_s stable eigenvalues and l_s complex conjugate pairs. Recall further that we have $n_{u,s} = m_{u,s} + 2l_{u,s}$. Taking Definition 3.15 into account we explicitly state that F maps the Banach space $X = (B_1 \times V_{\nu_s}, \|\cdot\|)$ to $Y = (\mathbb{R}^d, \|\cdot\|_\infty)$.

Before we give more details on the choice of norms in the following remark we emphasize that it is important for us to demand that the immersion Θ_ν maps B_1 into V_{ν_u} so that we can use the Cauchy estimates naturally given for the operator norms defined by (2.23). We now return to the question of norms and elaborate in addition on the induced matrix norms as these will be of some importance in the sequel. The goal of the following remark is to show that it is sufficient to bound the canonical matrix infinity norm, as it serves as an upper bound for the various induced norms.

Remark 3.1.3 *Consider first $A \in \mathbb{R}^{d,d}$ mapping from $X = (B_1 \times V_{\nu_s}, \|\cdot\|)$ to $Y = (\mathbb{R}^d, \|\cdot\|_\infty)$, where we set*

$$\|x\| = \max\left(\|(x_1, \dots, x_{n_u - 1})\|_2, \|(x_1, \dots, x_{n_s})\|_{m_s, l_s}\right). \tag{3.18}$$

By definition the matrix norm is given by

$$\sup_{\|x\|=1} \|Ax\|_\infty = \sup_{\|x\|=1} \max_{i=1,\dots,d} \left| \sum_{j=1}^{d} a_{ij} x_j \right|.$$

Let us estimate

$$\left| \sum_{j=1}^{d} a_{ij} x_j \right| \leq \sum_{j=1}^{d} |a_{ij}| \underbrace{|x_j|}_{\leq 1 \text{ if } \|x\| \leq 1}.$$

Thus we obtain $\sup_{\|x\|=1} \|Ax\|_\infty \leq \max_{i=1,\dots,d} \sum_{j=1}^{d} |a_{ij}|$. As usual by setting $x = e_i$ we note $\sup_{\|x\|=1} \|Ax\|_\infty = \max_{i=1,\dots,d} \sum_{j=1}^{d} |a_{ij}| = \|A\|_\infty$, the usual matrix ∞-norm.

Next we let $A \in \mathbb{R}^{d,d}$ map Y to X. We denote by $\|A\|$ the induced matrix norm and get by definition

$$\|A\| = \sup_{\|x\|_\infty = 1} \|Ax\| \leq \sup_{\|x\|_\infty} \sqrt{2}\|Ax\|_\infty = \sqrt{2}\|A\|_\infty. \tag{3.19}$$

Thus for computational convenience to bound $\|A\|$ we bound $\|A\|_\infty$ and use (3.19).

Last but not least a similar reasoning can be applied to bound the operator norm $\|\cdot\|_P$ for matrices A mapping $(B_1, \|\cdot\|_2)$ to $(V_{\nu_u}, \|\cdot\|_{(m_u, l_u)})$ to obtain $\|A\|_P \leq \sqrt{2}\|A\|_\infty$. Note that the subscript P refers to phase condition.

Before we state the validation values explicitly we introduce the following notation.

Definition 3.1.1 *We say $x \preceq y$ for $x, y \in \mathbb{R}^d$ iff $x_i \leq y_i$ for all $i = 1, \dots, d$.*

Assume the following validation values to be given. We point out that they all can be computed rigorously using interval arithmetic. In the sequel we motivate in detail how these quantities enter in the validation algorithm.

Definition 3.1.2 (Discretization-free validation values) We call $\bar{\alpha} \in B_1 \subset \mathbb{R}^{n_u-1}$, $\bar{\phi} \in B_{\nu_s} \subset \mathbb{R}^{n_s}$ and the positive constants ϵ, σ_u, σ_s, C_1, C_2, $\tilde{C}_1^i \in \mathbb{R}^{n_u}$,$i = 1, \dots, d$, $\tilde{C}_2^j \in \mathbb{R}^{n_u}$,$j = 1, \dots, n_u - 1$, $\tilde{C}_3^i \in \mathbb{R}^{d,n_u}$,$i = 1, \dots, d$, $\tilde{C}_4^{j,k} \in \mathbb{R}^{n_u}$,$j,k = 1, \dots, n_u - 1$, $\tilde{C}_5^{j,k}\mathbb{R}^d$, $j,k = 1, \dots, n_s$, A and $\bar{r}$ *validation values* if

(i)
$$\|\Theta_\nu(\bar{\alpha})\|_{(m_u,l_u)} < \nu_u \quad \text{and} \quad \|\bar{\phi}\|_{(m_s,l_s)} < \nu_s,$$

(ii)
$$\sigma_u \leq -\ln\left(\frac{\|\Theta_\nu(\bar{\alpha})\|_{(m_u,l_u)}}{\nu_u}\right) \quad \text{and} \quad \sigma_s \leq -\ln\left(\frac{\|\bar{\phi}\|_{(m_s,l_s)}}{\nu_s}\right),$$

(iii)
$$\|F_{(M,N)}(\bar{\alpha},\bar{\phi})\|_\infty \leq \epsilon_{num},$$

(iv)
$$\|[DF_{(M,N)}(\bar{\alpha},\bar{\phi})]^{-1}\| \leq C_1,$$

(v)
$$\|D\Theta_\nu(\bar{\alpha})\|_P \leq C_2,$$

(vi)
$$\pi C_1\left(\frac{(2l_u+4m_u)C_2}{\nu_u\sigma_u}\delta_u + \frac{2l_s+4m_s}{\nu_s\sigma_s}\delta_s\right) \leq A,$$

(vii)
$$\frac{2C_1}{1-A}(\epsilon_{num}+\delta_u+\delta_s) \leq \bar{r},$$

(viii)
$$\sup_{\|\varphi-\bar{\varphi}\|_{(m_u,l_u)}<r} |D(Q_M)_i(\varphi)| \preceq \tilde{C}_1^i \in \mathbb{R}^{n_u} \;\; (i = 1, \dots, d)$$
$$\sup_{\|\alpha-\bar{\alpha}\|_2<r} \left|\frac{\partial}{\partial\alpha_j}\Theta(\alpha)\right| \preceq \tilde{C}_2^j \in \mathbb{R}^{n_u} \;\; (j = 1, \dots, n_u - 1)$$
$$\sup_{\|\varphi-\bar{\varphi}\|_{(m_u,l_u)}<r} \left|D^2(Q_M)_i(\varphi)\right| \preceq \tilde{C}_3^i \in \mathbb{R}^{d,n_u} \;\; (i = 1, \dots, d)$$
$$\sup_{\|\alpha-\bar{\alpha}\|_2<r} \left|\frac{\partial}{\partial\alpha_k\partial\alpha_j}\Theta(\alpha)\right| \preceq \tilde{C}_4^{j,k} \in \mathbb{R}^d \;\; (j,k = 1, \dots, n_u - 1)$$

(ix)

$$\sup_{\|\phi-\bar{\phi}\|_{(m_s,l_s)}<r} \left| \frac{\partial}{\partial\phi_j\partial\phi_k} P_N(\phi) \right| \preceq \tilde{C}_5^{j,k} \;\; (j,k=1,\ldots,n_s).$$

Given these validation values we now describe a procedure how to verify the existence of a zero to the map $F : \mathbb{R}^d \to \mathbb{R}^d$ given by (3.16) which will correspond to a connecting orbit from p_1 to p_2. We term this the discretization-free procedure.

Discretization free procedure

Step 1 We need to show that that $DF(\bar{x})^{-1}$ has bounded norm. To this end we note that under the assumption that $DF_{N,M}(\bar{x})^{-1}$ exists, we have

$$\begin{aligned} DF(\bar{x})^{-1} &= [DF_{N,M}(\bar{x}) + DE(\bar{x})]^{-1} \\ &= \left[DF_{N,M}(\bar{x}) \left(I + DF_{N,M}(\bar{x})^{-1} DE(\bar{x}) \right) \right]^{-1} \\ &= \left(I + DF_{N,M}(\bar{x})^{-1} DE(\bar{x}) \right)^{-1} DF_{N,M}(\bar{x})^{-1}. \end{aligned} \tag{3.20}$$

As we aim to use a Neumann series argument as described in Remark 3.1.2 to show bounded invertibility of $\left(I + DF_{N,M}(\bar{x})^{-1} DE(\bar{x})\right)$, this poses us the problem of estimating $\|DF_{N,M}(\bar{x})^{-1} DE(\bar{x})\|_{B(X,X)}$ where $E(\alpha,\phi) = e_u(\Theta(\alpha)) - e_s(\phi)$. Recalling $\bar{x} = (\bar{\alpha},\bar{\phi})$ we compute, using submultiplicativity, that

$$\begin{aligned} &\|DF_{N,M}^{-1}(\bar{x}) DE(\bar{x})\|_{B(X,X)} \le \|DF_{N,M}(\bar{x})^{-1}\| \|De_u(\Theta(\alpha)) D\Theta(\alpha) - De_s(\phi)\|_\infty \\ &\le \|DF_{N,M}(\bar{x})^{-1}\| \left(|De_u(\Theta(\alpha))|_{\nu_u \exp(-\sigma_u)} \|D\Theta(\alpha)\|_P + |De_s(\phi)|_{\nu_s \exp(-\sigma_s)} \right). \end{aligned}$$

The next step is to use the Cauchy estimates specified in Lemma 2.1.3 and further adapted to the case of complex conjugate eigenvalues in Remark 2.1.3 as well as the bounds $C_{1,2}$ specified in 3.1.2(iv) and 3.1.2(v). Note that 3.1.2(ii) is equivalent to

$$\nu_u \exp(-\sigma_u) \le \|\Theta(\bar{\alpha})\|_{(m_u,l_u)} \text{ and } \nu_s \exp(-\sigma_s) \le \|\bar{\phi}\|_{(m_s,l_s)}$$

which together with 3.1.2(i) assures applicability of the Cauchy estimates. We obtain

$$\|DF_{N,M}^{-1}(\bar{x}) DE(\bar{x})\| \le C_1 \left(\frac{(2m_u + 4l_u)\pi}{\nu_u \sigma_u} \delta_u C_2 + \frac{(2m_s + 4l_s)\pi}{\nu_s \sigma_s} \delta_s \right) = A. \tag{3.21}$$

If we demand $A < 1$, Remark 3.1.2 enables us to estimate

$$\|DF(\bar{x})^{-1}\| \le \frac{1}{1-A} C_1 \tag{3.22}$$

which completes Step 1.

Step 2 The next step consists in computing ϵ_{NK} defined in (3.2) which amounts to finding a normwise bound on $DF(\bar{x})^{-1}F(\bar{x})$. To this end we recall

$$F(\bar{x}) = F_{N,M}(\bar{x}) + E(\bar{x}) = F_{N,M}(\bar{x}) + e_u(\Theta(\alpha)) - e_s(\phi)$$

Building on 3.1.2(i) the use of the aposteriori bounds $\delta_{u,s}$ is justified and we get that

$$\|F(\bar{x})\|_\infty \leq \underbrace{\|F_{N,M}(\bar{x})\|_\infty}_{\epsilon_{num}} + \delta_u + \delta_s$$

where ϵ_{num} is amenable to interval computations. Using this combined with (3.22) we see

$$\|DF(\bar{x})^{-1}F(\bar{x})\|_\infty \leq \frac{C_1}{1-A}(\epsilon_{num} + \delta_u + \delta_s) \overset{\text{def}}{=} \epsilon_{NK}.$$

Step 3 Recalling (3.3) we set $\bar{r} = 2\epsilon_{NK}$. This identifies 3.1.2(vi) as the condition encoding this definition.

Before continuing with Step 4 we wish to emphasize that Step $1-3$ exploit the numerical results together with rigorous error estimates. The better the numerical results and the tighter the error bounds the more likely it is that Step 4 and 5 are successful.

Step 4 The next step we need to take consists in computing $\kappa > 0$ such that

$$\|DF(x) - DF(y)\|_\infty \leq \kappa \|x - y\|$$

for all x, y with $\|x - \bar{x}\|_\infty \leq \bar{r}$ and $\|y - \bar{x}\|_\infty \leq \bar{r}$. To achieve this we will apply the Mean Value Theorem component-wise. Therefore let suitable $x = (\alpha_x, \phi_x), y = (\alpha_y, \phi_y) \in \mathbb{R}^d$ be given. We compute with $\xi = (\alpha_\xi, \phi_\xi)$ where $\xi_i = t_i x_i + (1-t_i)y_i$ for $i = 1, \ldots, d$ and some $t_i \in (0,1)$

$$\max_{i=1,\ldots,d} \sum_{j=1}^{d} \left| \frac{\partial F_i}{\partial x_j}(x) - \frac{\partial F_i}{\partial x_j}(y) \right| = \max_{i=1,\ldots,d} \sum_{j=1}^{d} \left| \nabla \frac{\partial F_i}{\partial x_j}(\xi)(x-y) \right| =$$

$$\leq \max_{i=1,\ldots,d} \sum_{j=1}^{d} \sum_{k=1}^{d} \left| \frac{\partial F_i(\xi)}{\partial x_j \partial x_k} \right| \underbrace{|(x-y)_k|}_{\leq \|x-y\|} \leq \|x-y\| \max_{i=1,\ldots,d} \sum_{j=1}^{d} \sum_{k=1}^{d} \left| \frac{\partial F_i(\xi)}{\partial x_j \partial x_k} \right|.$$

This leaves us with the task of estimating $\sum_{j=1}^d \sum_{k=1}^d \left| \frac{\partial F_i(\xi)}{\partial x_j \partial x_k} \right|$. Therefore let us first consider for $i = 1, \ldots, d$

$$\frac{\partial F_i(x)}{\partial x_j \partial x_k} = \frac{\partial (Q_M(\Theta(\alpha_x)))_i - (P_N(\phi_x))_i}{\partial x_j \partial x_k} + \frac{\partial (e_u(\Theta(\alpha_x)))_i - (e_s(\phi_x))_i}{\partial x_j \partial x_k}$$

$$= \frac{\partial (Q_M(\Theta(\alpha_x)))_i}{\partial x_j \partial x_k} + \frac{\partial (e_u(\Theta(\alpha_x))_i}{\partial x_j \partial x_k} - \frac{(P_N(\phi))_i}{\partial x_j \partial x_k} - \frac{(e_s(\phi))_i}{\partial x_j \partial x_k}.$$

Recalling $x = (\alpha_1, \dots, \alpha_{n_s}, \phi_1, \dots, \phi_{n_u})$ we use this to obtain

$$\sum_{j=1}^{d}\sum_{k=1}^{d}\left|\frac{\partial F_i(x)}{\partial x_j \partial x_k}\right| \leq \sum_{j=1}^{n_u-1}\sum_{k=1}^{n_u-1}\left|\frac{\partial (Q_M(\Theta(\alpha_x)))_i}{\partial \alpha_j \partial \alpha_k}\right| + \left|\frac{\partial (e_u(\Theta(\alpha_x))_i}{\partial \alpha_j \partial \alpha_k}\right|$$
$$+ \sum_{j=1}^{n_s}\sum_{k=1}^{n_s}\left|\frac{\partial (P_N(\phi_x))_i}{\partial \phi_j \phi_k}\right| + \left|\frac{\partial (e_s(\phi_x))_i}{\partial \phi_j \partial \phi_k}\right|.$$

Now we are in the position to use Cauchy estimates to bound the error terms and complete the computation κ via interval arithmetic. More concretely using the chain rule and recalling 3.1.2(viii) and 3.1.2(ix) we first get for $j, k = 1, \dots, n_s - 1$

$$\left|\frac{\partial (Q_M(\Theta(\alpha_x)))_i}{\partial \alpha_j \partial \alpha_k}\right| + \left|\frac{\partial (e_u(\Theta(\alpha_x))_i}{\partial \alpha_j \partial \alpha_k}\right|$$
$$\leq \langle \tilde{C}_3^i \tilde{C}_2^j, \tilde{C}_2^k \rangle + \langle \tilde{C}_1^i, \tilde{C}_4^{j,k} \rangle + \frac{16\pi^2}{\nu_u^2 \sigma_u^2}\delta_u \sum_{r=1}^{n_u}\sum_{l=1}^{n_u}(\tilde{C}_2^j)_r(\tilde{C}_2^k)_l + \frac{4\pi}{\nu_u \sigma_u}\delta_u \sum_{r=1}^{n_u}(\tilde{C}_4^{j,k})_r \overset{\text{def}}{=} P_{jk}.$$

where $\langle \cdot, \cdot \rangle$ denotes the euclidian dot product. So finally we are able to define

$$\kappa = \sum_{j=1}^{n_u-1}\sum_{k=1}^{n_u-1} P_{jk} + \sum_{j=1}^{n_s}\sum_{k=1}^{n_s} \tilde{C}_5^{j,k} + n_s^2 \frac{16\pi^2}{\nu_s^2 \sigma_s^2}\delta_s.$$

Step 5 The last step consist in checking inequality (3.4). We recall that if this is a strict inequality then it will follow from Theorem 3.3.1 that the connecting orbit we validate corresponds to a transversal intersection of the involved unstable and stable manifolds.

Let us summarize the above procedure in the following theorem.

Theorem 3.1.2 *Assume validation values according to Definition 3.1.2 to be given. Then there exist a unique zero $\tilde{x} = (\tilde{\alpha}, \tilde{\phi})$ of F given by* (3.16) *in a ball of radius $\bar{r}$ with respect to $\|\cdot\|$ from* (3.18) *around the approximate solution $(\bar{\alpha}, \bar{\phi})$.*

The proof is given by the above reasoning.

3.2 Approach solving a boundary value problem

After considering the discretization free approach in Section 3.1 we now compute connecting orbits by solving a parametric boundary value problem for (2.1) that we rewrite as an integral equation augmented by some boundary conditions. Solutions hereof we interpret as zeros of an operator F defined on a Banach space X to be in the general setting of (3.1). Note that one main difference to the above approach is that we now have an operator which is defined on an infinite dimensional space as opposed to the finite dimensional setting of Section 3.1.

We aim to compute a solution of $F(x) = 0$ by applying the method of radii polynomials [15]. That is, as a first step we construct a fixed point operator $T : \mathcal{X} \to \mathcal{X}$ defined on an associated infinite dimensional Banach space $\mathcal{X}$ whose fixed points are in bijective correspondence to zeros of F. The construction of the fixed point operator T and the choice of the Banach space $\mathcal{X}$ will be different in the Spline and Chebyshev approach. More precisely, in the Spline approach we set $\mathcal{X} = X$ and construct a Newton-like fixed point operator T directly from F. In the Chebyshev approach we let $\mathcal{X}$ be the space of rapidly decaying Chebyshev coefficients and construct an equivalent problem of the form $f(x) = 0$ on $\mathcal{X}$. Then we derive T corresponding to f.

Once $T : \mathcal{X} \to \mathcal{X}$ is defined the idea is to find a fixed point of T in a ball $B_r(\bar{x})$ around an approximate solution $\bar{x}$ of $F(x) = 0$ or $f(x) = 0$ respectively. In order to do so we take the radius r of the ball as a variable and compute a finite number of polynomials $p_0(r), \ldots, p_M(r)$ such that the following implication is valid: if we find an $\bar{r} > 0$ such that $p_i(\bar{r})$ is negative for all $i = 0, \ldots, M$, then T is a contraction on the ball $B_{\bar{x}}(\bar{r}) \subset \mathcal{X}$. By using the Banach Fixed Point theorem we hereby obtain the existence of a unique genuine fixed point of T and hence a unique genuine zero of F.

Before we go on to construct the operator F we wish to emphasize the following point. We assume explicitly that we are given an approximation $\bar{x} \in \mathcal{X}$ in the infinite dimensional space $\mathcal{X}$. We will obtain this by numerical computations in a finite dimensional approximate space and embedding the result back again into the full space $\mathcal{X}$.

We start by constructing the general operator F and give the details on the respective fixed point operators in separate subsections.

Formulation of the parametrized BVP Let times $t_1 < t_2$ be given. Defining $\mathrm{p}_{1,2} = u(t_{1,2})$, solving (2.1) is equivalent to solving either the integral equation

$$\begin{aligned} u(t) = \mathrm{p}_1 + \int_{t_1}^{t} g(u(s))ds \quad &\forall t \in [t_1, t_2] \\ \Leftrightarrow \qquad\qquad & \qquad\qquad\qquad (3.23) \\ F_1(\mathrm{p}_1, u)(t) \overset{\text{def}}{=} \mathrm{p}_1 + \int_{t_1}^{t} g(u(s))ds - u(t) = 0 \quad &\forall t \in [t_1, t_2] \end{aligned}$$

or

$$\begin{aligned} u(t) = \mathrm{p}_2 - \int_{t}^{t_2} g(u(s))ds \quad &\forall t \in [t_1, t_2] \\ \Leftrightarrow \qquad\qquad & \qquad\qquad\qquad (3.24) \\ F_2(\mathrm{p}_2, u)(t) \overset{\text{def}}{=} u(t) + \int_{t}^{t_2} g(u(s))ds - \mathrm{p}_2 = 0 \quad &\forall t \in [t_1, t_2]. \end{aligned}$$

Our goal in defining F is, to use use either (3.23) or (3.24) and to additionally encode the property of being a connecting orbit in parametric boundary conditions. The idea is to extract the boundary conditions from the fact that a connecting orbit starts on a unstable manifold and ends on a stable manifold.
Mathematically this means we will set

$$X = \mathbb{R}^p \times B, \tag{3.25}$$

where B is an appropriate function space and consider an operator $F : X \to X$ given by

$$F(\theta, u)(t) \overset{\text{def}}{=} \begin{pmatrix} \mathcal{G}(\mathrm{p}_1(\theta), \mathrm{p}_2(\theta)) \\ F_{1,2}(\mathrm{p}_{1,2}(\theta), u)(t) \end{pmatrix}, \tag{3.26}$$

where $\mathcal{G} : \mathbb{R}^{2d} \to \mathbb{R}^p$ is an affine map and possibly p_1 and /or p_2 depend on the parameter $\theta \in \mathbb{R}^p$ in a way ensuring that the operator (3.26) has isolated zeros. We use the convention $p = 0$ for the absence of $\mathcal{G}$, as encountered in IVPs. To be clear we emphasize that the benefit of this general notation is that it allows us to unite both the solution of IVPs and the computation of generic connecting orbits while being able to take symmetries into account, as we will in the example of the Gray-Scott equation considered in Section 4.1. More precisely, the goal of this choice of notation is to make the methods described herein easily extendable to for example the context of generic homoclinic connecting orbits and systems possessing integrals of motions. This will necessitate the introduction to further constraints in order to assure isolation of the solutions to the operator equation that we wish to incorporate in the function $\mathcal{G}$.

For the sake of concreteness let us construct the operators to compute solutions of IVPs and of generic heteroclinic orbits between hyperbolic equilibria $p_{1,2}$.

Example 3.2.1 1. *Consider the initial value problem*

$$\dot{u} = g(u) \quad u(t_1) = \mathrm{p}_1 \tag{3.27}$$

which is equivalent to solving $F_1(p_1, u) = 0$ with F_1 defined in (3.23)*. Thus by setting $p = 0$ we obtain the operator $F : B \longrightarrow B$ with $F(u) = F_1(p_1, u)$ whose zeros correspond to solutions of* (3.27)*.*

2. *Let two hyperbolic equilibria $p_{1,2}$ fulfilling* (2.4) *be given together with parametrizations*

$$Q : \mathbb{R}^{n_u} \supset V_{\nu_u} \longrightarrow \mathbb{R}^d \qquad P : \mathbb{R}^{n_s} \supset V_{\nu_s} \longrightarrow \mathbb{R}^d$$

of the local unstable and stable manifold of $p_{1,2}$. Letting a phase condition $\Theta : \mathbb{R}^{n_u - 1} \longrightarrow \mathbb{R}^{n_u}$ explained in (3.14) *be given. Recalling that $n_u - 1 + n_s = d$, we define $p = d$ and set $\theta = (\alpha, \phi) \in \mathbb{R}^d$. Furthermore let*

$$\mathcal{G}(\mathrm{p}_1, \mathrm{p}_2) = \mathcal{G}(\mathrm{p}_1(\theta), \mathrm{p}_2(\theta)) = \underbrace{P(\phi)}_{\mathrm{p}_2(\theta)} - \underbrace{Q(\Theta(\alpha))}_{\mathrm{p}_1(\theta)} - \int_{t_1}^{t_2} g(u(s))ds \tag{3.28}$$

and set

$$F(\theta, u) = \begin{pmatrix} \mathcal{G}(\mathrm{p}_1(\theta), \mathrm{p}_2(\theta)) \\ F_1(\mathrm{p}_1(\theta), u)(t) \end{pmatrix}. \tag{3.29}$$

The choice of the function space B is a critical one and depends on several parameters, that we elaborate on in the following remark.

Remark 3.2.1 *First it depends on the equation and the regularity the equation prescribes for its solutions. This gives an upper bound on the regularity we can assume about our solution and hence to the regularity we prescribe by choosing the function space. Secondly it depends on the error norms that we have at our disposition. The natural norm to measure the error of a spline approximation is the sup-norm $\|\cdot\|_\infty$ and as we require B to be a Banach space we need to choose it accordingly. A generic choice is to pick*

$$B = C_0([0,1], \mathbb{R}^d),$$

the space of continuous functions from $[0,1] \to \mathbb{R}^d$*. In the Chebyshev approach it is convenient to assume that the Chebyshev coefficient decay algebraically to any order. Thus*

$$B = C^{\omega}([-1,1],\mathbb{R}^d)$$

denoting the space of real analytic functions is a good choice, as this ensures geometric decay which implies algebraic decay to any order (see Section 2.2.1). It is worth noting that we use less regularity than we could by stepping back to a space of algebraically decaying coefficients with some fixed decay rate s*.*

We now go on to construct the fixed point operator T encoding the zeros of F. We start by using the Spline approach before we go on with the Chebyshev approach. Before we go to the details of the derivation we wish to make the following remark.

Remark 3.2.2 *The goal in both the Spline and the Chebyshev approach is to divide the operator* T *into a finite dimensional part amenable to numerical investigation and an infinite dimensional tail part controlling the error introduced by the truncation. The Chebyshev algorithm will be more explicit in the sense that we start from a basis expansion of the unknown function that gives us an infinite set of concrete equations for every basis coefficient that we then truncate. In the spline approach we start from a finite approximation to the unknown function that gives us a finite dimensional equation. The infinite tail part has to be dealt with by a-priori estimates.*
This fact motivates that in the spline case we consider a splitting of $X = X_m \oplus X_\infty$ *in a finite dimensional space* X_m *and infinite dimensional* X_∞ *where we only have normwise bounds for the elements of* X_∞*.*
We point out that in this sense the Spline approach can be considered to be more rigid than the Chebyshev approach.

3.2.1 Validation of connecting orbits: Spline approach

Let F be given by (3.23). We consider F as a map from $X = \mathbb{R}^p \times C_0([0,1],\mathbb{R}^d)$ to itself. The goal is to derive a fixed point operator T defined on X whose fixed points correspond to zeros of F and show that T is a contraction in a ball around a numerical approximation $\bar{x} \in X$ to the solution of $F(x) = 0$. To obtain the approximation we first aim to discretize F using linear splines.

Discretization of F using linear splines to construct splitting $X = X_m \oplus X_\infty$

Set $t_{1,2}$ in (3.23) to $t_1 = 0$ and $t_2 = 1$. To account for the choice of this specific time interval we rescale (3.23) by a time factor L, set $B = C_0([0,1], \mathbb{R}^d)$ in (3.25) and consider $F : \mathbb{R}^p \times C_0([0,1], \mathbb{R}^d)$ to be given by

$$F(\theta, u) = \begin{pmatrix} \mathcal{G}(p_1(\theta), p_2(\theta)) \\ \mathrm{p}_1(\theta) + L \int_0^t g(u(s))ds - u(t) \end{pmatrix}. \tag{3.30}$$

We remark that we do not take L as a variable. We assume this to be part of the numerical approximation to our connecting orbit.

In addition let $\Delta : 0 = t_0 < t_1 < \cdots < t_m = 1$ be a grid on $[0,1]$. We aim to use the linear spline projection associated to Δ to construct a splitting $X = X_m \oplus X_\infty$, into a finite dimensional space X_m and an infinite dimensional error space X_∞.

Denote $\mathcal{S}_m$ the space of linear splines subordinate to the grid Δ and consider the linear spline projection $\Pi_m : C_0([0,1], \mathbb{R}) \to \mathcal{S}_m \cong \mathbb{R}^{m+1}$ consisting of computing the linear interpolation of u with respect to the mesh Δ. More concretely we have for $u \in C_0([0,1], \mathbb{R})$

$$\Pi_m(u) = (u(t_0), u(t_1), \ldots, u(t_m)) \in \mathbb{R}^{m+1}.$$

For $u \in C_0([0,1], \mathbb{R}^d)$ define

$$u_h = (\Pi_m)^d u = (\Pi_m u_1, \ldots, \Pi_m u_d) \in (\mathcal{S}_m)^d \cong \mathbb{R}^{d(m+1)}.$$

Depending on the context we interpret u_h as a continuous function, as a d times $m+1$ matrix, where we denote its columns by $(u_h)_l \in \mathbb{R}^d$ for $l = 1, \ldots, m+1$ or as a $d(m+1)$ column vector.

Define $X_m = \mathbb{R}^p \times (\mathcal{S}_m)^d$ and the finite dimensional projection $\Pi_m : X \to X_m : (\theta, u) \mapsto (\theta, (\Pi_m)^d u)$. By using the complementary projection $I - \Pi_m$ we get that $X_\infty = 0 \times (I - \Pi_m)^d C_0([0,1], \mathbb{R}^d)$, where $(I - \Pi_m)^d u = ((I - \Pi_m)u_1, \ldots, (I - \Pi_m)u_d)$. The associated projection $\Pi_\infty : X \to X_\infty$ is given by $(\theta, u) \mapsto (0, (I - \Pi_m)^d u)$. For $x \in X$ we write $x = (x_m, x_\infty)$, with $x_m \overset{\text{def}}{=} \Pi_m x$ and $x_\infty \overset{\text{def}}{=} \Pi_\infty x$.

Let us furthermore define the norms

$$\|\Pi_m(\theta, u)\|_{X_m} = \max\{\|\theta\|_\infty, \|\Pi_m u_1\|_\infty, \ldots, \|\Pi_m u_d\|_\infty\} \tag{3.31}$$

and

$$\|\Pi_\infty(\theta,u)\|_{X_\infty} = \max_{l=1,\dots,d} \sup_{t\in[0,1]} |(I-\Pi_m)u_l(t)|, \tag{3.32}$$

which qualify the pairs $(X_m, \|.\|_{X_m})$ and $(X_\infty, \|.\|_{X_\infty})$ as Banach spaces.

This splitting of the space X induces a splitting of the operator F that is defined on it. Furthermore we have the splitting of the operator

$$F = F_m \oplus F_\infty \overset{\text{def}}{=} \Pi_m F \oplus \Pi_\infty F. \tag{3.33}$$

We will derive explicit formulas for the discretized operator F_m in the application section. Before turning to the definition of the fixed point operator T we elaborate more on how we identify numerical outputs with elements in function space.
Set $M = p + d(m+1)$ and consider an isomorphism $i : \mathbb{R}^M \to X_m$ and define $F^m = i^{-1} \circ F_m \circ \tau$, where the embedding $\tau : \mathbb{R}^M \to X_m \oplus \{0_\infty\}$ is defined by $w \mapsto (i(w), 0_\infty)$. For sake of simplicity we identify X_m and $\mathbb{R}^M$, as well as $x_m \in X_m$ and $i^{-1}(x_m) \in \mathbb{R}^M$. In particular we write $x = (x_m, x_\infty) = ((x_m)_0 \dots, (x_m)_{d+1}, x_\infty)$, where $(x_m)_0 \in \mathbb{R}^p$ and $(x_m)_j \in \mathbb{R}^{m+1}$ for $j = 1, \dots, d$. Note that $F^m : \mathbb{R}^M \to \mathbb{R}^M$ and we can use standard numerical techniques (e.g. Newton's method, continuation techniques [31]) in order to compute an approximate solution $\bar{x}_m$ of

$$F^m(x_m) = 0. \tag{3.34}$$

From the above construction, one has that $\bar{x} = \tau(\bar{x}_m)$ is an approximate solution of (3.34).

Construction of the fixed point operator T and the radii polynomials

Assume a finite dimensional approximate solution $\bar{x} \in X$ such that $F_m(\bar{x}) \approx 0$. Let us define the set $B_{\bar{x}}(r) = \bar{x} + B(r,\omega)$, where

$$B(r,\omega) \overset{\text{def}}{=} \{x \in X : \|x_m\|_{X_m} \le r \text{ and } \|x_\infty\|_{X_\infty} \le \omega r\}, \tag{3.35}$$

with a fixed parameter ω that can be used to control the infinite dimensional error. Note that the dependence of the set $B_{\bar{x}}(r)$ on the variable radius r is a-priori unknown, and that the idea of the method of radii polynomials is to solve for a suitable r_* such that our corresponding fixed point operator T is a contraction on $B_{\bar{x}}(r)$.

In order to define T, assume first that the following assumptions (**RP**) are fulfilled.

- **RP1.** We have computed an approximate solution $\bar{x} = (\bar{x}_m, 0_\infty)$ for (3.34), that is there exists a *small* $\epsilon > 0$ such that $\|F^m(\bar{x}_m)\|_{X_m} \leq \epsilon$.
- **RP2.** We have computed the Fréchet derivative $DF^m(\bar{x}_m)$.
- **RP3.** We have computed an approximate inverse A_m for $DF^m(\bar{x}_m)$.
- **RP4.** A_m is injective.

Let us define the fixed point operator $T : X \to X$ to be

$$T(x) = (x_m - A_m F_m(x)) \oplus (F_\infty(x) + x_\infty). \tag{3.36}$$

Lemma 3.2.1 *Let $x \in X$. Then $F(x) = 0$ if and only if $T(x) = x$.*

Proof 3.2.1 *Assume $F(x) = 0$. By* (3.33)*, one has that $F_m(x) = 0_m$ and $F_\infty(x) = 0_\infty$. Therefore $T(x) = x_m \oplus x_\infty = x$. On the other hand if $T(x) = x$ it follows that $A_m F_m(x) = 0_m$ and by injectivity of A_m this amounts to $F_m(x) = 0_m$. Furthermore we have $F_\infty(x) = 0_\infty$ and by* (3.33) *it follows that $F(x) = 0 \in X$.*

We now describe the method of radii polynomials in more detail. The construction of the polynomials $p_0, \ldots, p_M(r)$ depends on some bounds encoding the requirements for the Banach Fixed Point Theorem to be applicable on the ball $B_{\bar{x}}(r)$. More precisely calling upon the contraction mapping theorem on $B_{\bar{x}}(r)$ necessitates to show that $T(B_{\bar{x}}(r)) \subset B_{\bar{x}}(r)$ and that T is a contraction on $B_{\bar{x}}(r)$. In order to do so, we consider the residual function

$$y \stackrel{\text{def}}{=} T(\bar{x}) - \bar{x} = -A_m F_m(\bar{x}) \oplus F_\infty(\bar{x}). \tag{3.37}$$

First let us assume that we are given positive constants $Y_0 \in \mathbb{R}^p$ and $Y_1, \ldots, Y_{m+1} \in \mathbb{R}^d$ and $Y_\infty \in \mathbb{R}$ such that

$$|\,(\Pi_m y)_i\,| \preceq Y_i, \quad \text{for all } i = 1, \ldots, m+1 \tag{3.38}$$

and

$$\|\Pi_\infty y\|_{X_\infty} \leq Y_\infty. \tag{3.39}$$

To show that T is a contraction, we introduce, for $\xi_1, \xi_2 \in B(r, \omega)$, the quantity

$$z(\xi_1, \xi_2) \stackrel{\text{def}}{=} DT(\bar{x} + \xi_1)\xi_2. \tag{3.40}$$

Realize that $z(\xi_1, \xi_2)$ is linear in ξ_2. Assume further that we are given polynomials bounds $Z_0(r) \in \mathbb{R}^p$, $Z_i(r) \in \mathbb{R}^d$ for $i = 1, \ldots, m+1$ and $Z_\infty(r) \in \mathbb{R}$ satisfying

$$\sup_{\xi_1, \xi_2 \in B(r,\omega)} |\left(\Pi_m(z(\xi_1, \xi_2))\right)_i| \preceq Z_i(r), \quad \text{for all } i = 1, \ldots, m+1 \tag{3.41}$$

and

$$\sup_{\xi_1, \xi_2 \in B(r,\omega)} \|\Pi_\infty(z(\xi_1, \xi_2))\|_{X_\infty} \leq Z_\infty(r). \tag{3.42}$$

Using the above ingredients we define the radii polynomials.

Definition 3.2.1 *Assume we are given bounds as in* (3.38) *and* (3.39) *and polynomial bounds as in* (3.41) *and* (3.42). *Then define for $i = 1, \ldots, m+1$ the finite radii polynomials*

$$p_i(r) = Y_i + Z_i(r) - r$$

and, given a number $\omega > 0$, define the tail radii polynomial

$$p_\infty(r) = Y_\infty + Z_\infty(r) - \omega r.$$

Let us now state the main result, whose proof can be found in [60].

Theorem 3.2.1 *[Theorem 2.6 in [60]] If there exists an $\bar{r} > 0$ such that $p_i(\bar{r}) \prec 0$ for all $i = 1, \ldots, m+1$ and $p_\infty(\bar{r}) < 0$ then T is a contraction on $B_{\bar{x}}(\bar{r})$ and hence there exists a unique zero $\tilde{x}$ of* (3.1) *in $B_{\bar{x}}(\bar{r})$.*

Remark 3.2.3 *The vector bounds $Y_1, \ldots, Y_{m+1}$ and $Z_1(r), \ldots, Z_{m+1}(r)$ defined in* (3.38) *and* (3.41) *are the direct analogues of ϵ_{NK} and κ that are also found numerically. The additional bounds Y_∞ and $Z_\infty(r)$ are introduced in order to control the truncation error introduced by discretization.*

3.2.2 Validation of connecting orbits: Chebyshev approach

In contrast to the Spline approach we do not directly discretize (3.26) but first compute a representation of F in terms of the Chebyshev basis.

Derivation of the Chebyshev representation We use the following corollary which is a direct consequence of Theorem 2.2.2.

Corollary 3.2.1 *Assume that $g : \mathbb{R}^d \to \mathbb{R}^d$ is real analytic and let $u : [-1, 1] \to \mathbb{R}^n$ be a solution of* (2.1). *Then each component u_j of u is real analytic and has a unique representation as an absolutely and uniformly convergent series of the* *form* $u_j(t) = \sum_{k=0}^{\infty}(a_j)_k T_k(t)$. *Also, for each* $j \in \{1, \dots, d\}$, *the sequence of Chebyshev coefficients* $\{(a_j)_k\}_{k\geq 0}$ *of* u_j *decreases to zero faster than any algebraic decay, that is, for any decay rate* $s > 1$, *there exists a constant* $A_j = A_j(s) < \infty$ *such that* $|(a_j)_k| \leq \frac{A_j}{k^s}$, *for* $k \geq 1$.

We will henceforth assume that the vector field $g : \mathbb{R}^d \to \mathbb{R}^d$ is real analytic and that $t_1 = -1$ and $t_2 = 1$ in (3.23) and (3.24) respectively. Note that this is no restriction as a rescaling of time can (and will) be considered in the autonomous vector field g of a particular application. Furthermore Corollary 3.2.1 assures that F is defined from $\mathbb{R}^p \times X^{\omega}([-1,1])^d$ to itself justifying the choice $B = X^{\omega}([-1,1])^d$ in (3.25). We will nevertheless not use this explicitly but now step to coefficient space.

By a rescaling of the Chebyshev coefficients $(a_j)_k$ in Corollary 3.2.1 we can furthermore assume a Chebyshev expansion of a solution u of (2.1) to be given by

$$u(t) = a_0 + 2\sum_{k\geq 1} a_k T_k(t), \tag{3.43}$$

where $a_k = \big((a_1)_k, (a_2)_k, \cdots, (a_d)_k\big)^T \in \mathbb{R}^d$. Letting

$$\|a_k\|_\infty = \max_{j=1,\dots,d}\{|(a_j)_k|\}$$

and using the weights defined in (2.38) one has by Corollary 3.2.1 that for any given $s > 1$

$$\|a\|_s \overset{\text{def}}{=} \sup_{k\geq 0}\{\|a_k\|_\infty \omega_k^s\} < \infty. \tag{3.44}$$

Defining X^s to be

$$\begin{aligned} X^s = \{x = (\theta, (a_k)_k) : a_k = \big((a_1)_k, (a_2)_k, \cdots, (a_d)_k\big)^T \in \mathbb{R}^d \\ \text{and } \|x\|_s = \sup_{k\geq k_0} \|x_k\|_\infty \omega_k^s < \infty\} \end{aligned}$$

the pair $(X^s, \|.\|_s)$ becomes a Banach space. Note that all d component sequences a_k are elements of Ω^s defined in (2.39). We use the notation $x = (x_k)_{k\geq k_0}$ with $k_0 = 0$ for $p = 0$ and $k_0 = -1$ for $p > 0$. Our goal is first to find a map f defined on X^s such that for $x = (\theta, a) \in X^s$

$$f(x) = 0 \Leftrightarrow F(\theta, u) = 0$$

where u corresponding to a is given by (3.43). In order to achieve this let us start by plugging in the expansion (3.43) into the vector field g to get

$$g(u(t)) = c_0 + 2\sum_{k\geq 1} c_k T_k(t), \tag{3.45}$$

where we explicitly assume that starting with $\|a\|_s < \infty$ we obtain $\|c\|_s < \infty$.

Remark 3.2.4 *The assumption that $\|c\|_s < \infty$ in* (3.45) *is fulfilled for all polynomial vector fields $g : \mathbb{R}^d \longrightarrow \mathbb{R}^d$. More precisely let g be an nth-order polynomial nonlinearity given by*

$$g(u_1, \ldots, u_d) = \sum_{|l|=0}^{n} d_l u_1^{l_1} \cdots u_d^{l_d},$$

where $d_l \in \mathbb{R}^d$ and $l = (l_1, \ldots, l_d)$ with $|l| = l_1 + \ldots + l_d$. Assuming the componentwise Chebyshev expansion (3.43)*, using the convolution Lemma 2.2.2 we obtain*

$$\begin{aligned} g(u_1, \ldots, u_d) &= \sum_{|l|=0}^{n} d_l \left((a_1^{l_1} * \ldots * a_d^{l_d})_0 + 2 \sum_{k=1}^{\infty} (a_1^{l_1} * \ldots * a_d^{l_d})_k T_k(t) \right) \\ &= d_0 + 2 \sum_{k=1}^{\infty} \left(\sum_{|l|=0}^{n} d_l (a_1^{l_1} * \ldots * a_d^{l_d})_k \right) T_k(t) \end{aligned} \tag{3.46}$$

and hence

$$c_0 = d_0 \qquad c_k = \sum_{|l|=0}^{n} d_l (a_1^{l_1} * \ldots * a_d^{l_d})_k \text{ for } k \geq 1. \tag{3.47}$$

where

$$a_k^{l_k} \overset{\text{def}}{=} \underbrace{a_k * \ldots * a_k}_{l_k \text{ times}}.$$

By the fact that $(\Omega^s, \|.\|_{\Omega^s}, *)$ with Ω^s defined in (2.39) *is an algebra we know that with $c = (c_k)_{k \in \mathbb{N}}$ given by* (3.47) *$\|c\|_s < \infty$. For non-polynomial real analytic vector fields a Taylor expansion can be considered.*

Next we plug (3.45) into the general expression for F given by (3.26). Let us furthermore assume that the second component of F is given by (3.23). The case where it is given by (3.24) is very similar and we will point out the differences that occur. Our goal is to obtain an expansion

$$F_1(\mathrm{p}_1, u) = \tilde{f}_0 + 2 \sum_{k=1}^{\infty} \tilde{f}_k T_k(t)$$

with $\tilde{f}_k \in \mathbb{R}^d$. Hence, using Lemma 2.2.1 we compute

$$
\begin{aligned}
&F_1(\mathrm{p}_1, u) = \mathrm{p}_1 + \int_{-1}^{t} g(u(s))ds - u(t) = \\
&= \mathrm{p}_1 + \int_{-1}^{t} \left(c_0 + 2\sum_{k=1}^{\infty} c_k T_k(s) \right) ds - \left(a_0 + 2\sum_{k=1}^{\infty} a_k T_k(t) \right) = \\
&= \mathrm{p}_1 + \Bigg[c_0 \Bigg(T_1(t) - \underbrace{T_1(-1)}_{=-1=-T_0(t)} \Bigg) + \frac{2}{4}c_1 \Bigg(T_2(t) + \underbrace{T_0(t)}_{=T_0(-1)} - (\underbrace{T_2(-1)}_{=1} - T_0(-1)) \Bigg) \\
&+2\sum_{k=2}^{\infty} c_k \left[\frac{1}{2}\left(\frac{T_{k+1}(t)}{k+1} - \frac{T_{k-1}(t)}{k-1} \right) \underbrace{-\frac{1}{2}\left(\frac{(-1)^{k+1}}{k+1} - \frac{(-1)^{k-1}}{k-1} \right)}_{=-\frac{(-1)^k}{k^2-1}} \right] \Bigg] \\
&- \left(a_0 + 2\sum_{k=1}^{\infty} a_k T_k(t) \right) = T_0(t) \left(\mathrm{p}_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 \right) \\
&+ 2T_1(t) \left(\frac{c_0}{2} - \frac{c_1}{2} \right) + 2T_2(t) \left(\frac{c_1}{4} - \frac{c_3}{4} \right) + 2\sum_{k=2}^{\infty} \frac{(c_{k-1} - c_{k+1})}{2k} T_k(t) - 2\sum_{k=1}^{\infty} a_k T_k(t) = \\
&= T_0(t) \left(\mathrm{p}_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 \right) + 2\sum_{k=2}^{\infty} \left(\frac{(c_{k-1} - c_{k+1})}{2k} - a_k \right) T_k(t),
\end{aligned}
$$

where the numbers marked in red are added artificially in order to clarify the pattern. As a result we are able to define

$$
\tilde{f}_k = \begin{cases} \mathrm{p}_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 & k = 0 \\ \frac{1}{2k}\left(c_{k-1} - c_{k+1} \right) - a_k & k \geq 1. \end{cases} \tag{3.48}
$$

Concerning the boundary condition $\mathcal{G}(\mathrm{p}_1, \mathrm{p}_2)$ we assume a general expansion

$$
\eta : X^s \to \mathbb{R}^p, \;\; \eta(\theta, a) = \mathcal{G}(\mathrm{p}_1, \mathrm{p}_2). \tag{3.49}
$$

We hence are ready to define the map $f = (f_k)_{k \geq k_0}$ on X^s componentwise by

$$
f_k = \begin{cases} \eta(\theta, a) & k = -1 \\ \mathrm{p}_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 & k = 0 \\ 2k a_k + (c_{k+1} - c_{k-1}) & k \geq 1 \end{cases} \tag{3.50}
$$

where $c_k = c_k(a)$ and we employed the rescaling $f_k = -2k\tilde{f}_k$ for $k \geq 1$.

Remark 3.2.5 *If we assume the second component of F to be given by* (3.24) *then we obtain*

$$f_k = \begin{cases} \eta(\theta, a) & k = -1 \\ -\mathrm{p}_2 + c_0 + \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} + a_0 & k = 0 \\ 2ka_k + (c_{k+1} - c_{k-1}) & k \geq 1 \end{cases} \tag{3.51}$$

where we apply for $k \geq 1$ the scaling $f_k = 2k\tilde{f}_k$ instead of the above considered scaling.

For the sake of concreteness let us derive the map η for the solution of IVPs and the computation of generic heteroclinics.

Remark 3.2.6 1. *In order to solve IVPs with initial value p_1 we do not have an explicit additional boundary condition. Thus we set $p = 0$, $k_0 = 0$ and f from* (3.50) *is given componentwise by*

$$f_k = \begin{cases} p_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 & k = 0 \\ 2ka_k + (c_{k+1} - c_{k-1}) & k \geq 1. \end{cases} \tag{3.52}$$

2. *Concerning the computation of generic heteroclinic orbits between hyperbolic fixed points $p_{1,2}$ with (un)stable manifolds of dimension $n_{u,s}$ we recall $\theta = (\alpha, \phi)$ and summarize* (3.28) *by*

$$\mathrm{p}_1(\theta) = Q(\Theta(\alpha)) \qquad \mathrm{p}_2(\theta) = P(\phi),$$

where $\mathcal{G}(\mathrm{p}_1, \mathrm{p}_2)$ can be written as

$$u(1) = P(\phi),$$

where Q and P are parametrizations of the corresponding unstable and stable manifolds. Hence, using the property $T_k(1) = 1$ for all $k \geq 0$ from Lemma 2.2.1, we obtain

$$f_k = \begin{cases} P(\phi) - (a_0 + 2\sum_{k=1}^{\infty} a_k) & k = -1 \\ Q(\Theta(\alpha)) + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty} \frac{(-1)^k c_k}{k^2-1} - a_0 & k = 0 \\ 2ka_k + (c_{k+1} - c_{k-1}) & k \geq 1 \end{cases} . \tag{3.53}$$

In the next result we summarize some important features of the above procedure. In particular we obtain that $f : X^s \to X^{s-1}$. In addition we realize that solutions $x \in X^s$ of the equation $f(x) = 0$ have strong regularity properties.

Lemma 3.2.2 *Let f be given by* (3.50) *and set $s \geq 2$ be arbitrary but fixed. Assume that there is a constant $C \in \mathbb{R}$ such that*

$$\sup_{k\geq 0}\{\|c_k\|_\infty \omega_k^s\} = C < \infty.$$

Then $f : X^s \to X^{s-1}$. Additionally if $x \in X^s$ is a solution of $f(x) = 0$ with f either given by (3.50) *or* (3.51)*, then u defined by* (3.43) *solves $F(\theta, u) = 0$ with F given by* (3.23) *or* (3.24) *respectively. Furthermore it follows that $x \in X^{s_0}$ for all $s_0 > 1$.*

Proof 3.2.2 *Let $(\theta, a) \in X^s$ be given. Then there exists a constant $A \in \mathbb{R}$ with* $\sup_{k\geq 0}\{\|a_k\|_\infty \omega_k^s\} < A < \infty$*. We compute for $k = 0$:*

$$\begin{aligned}
&\|\mathrm{p}_1 + c_0 - \frac{1}{2}c_1 - 2\sum_{k=2}^{\infty}\frac{(-1)^k c_k}{k^2-1} - a_0\|_\infty \leq \\
&\leq \|\mathrm{p}_1\|_\infty + \frac{1}{2}\|c_1\|_\infty + 2\sum_{k=1}^{\infty}\frac{\|c_k\|_\infty}{k^2-1} + \|a_0\|_\infty \leq \\
&\leq \|\mathrm{p}_1\|_\infty + \frac{C}{2} + 2C\underbrace{\sum_{k=1}^{\infty}\frac{1}{\omega_k^s(k^2-1)}}_{<\infty} + A = C_1 < \infty.
\end{aligned}$$

And for $k \geq 1$:

$$\begin{aligned}
\|c_{k+1} - c_{k-1} + 2ka_k\|_\infty \omega_k^{s-1} &\leq \left[\frac{C}{\omega_k^s}\left(1 + \underbrace{\frac{\omega_k^s}{\omega_{k-1}^s}}_{\leq 2^s}\right) + 2k\frac{A}{\omega_k^s}\right]\omega_k^{s-1} \\
&\leq \frac{(1+2^s)C}{\omega_k} + 2A \leq (1+2^s)C + 2A = C_2 < \infty.
\end{aligned}$$

This induces that

$$\sup_{k\geq 0}\|f_k\|\omega_k^{s-1} \leq \max(C_1, C_2) < \infty.$$

Taking $\|\eta(\theta, a)\|_\infty < \infty$ into account, it follows that $(f_k)_{k\geq k_0} \in X^{s-1}$ and hence $f : X^s \to X^{s-1}$.

Next assume that $x = (\theta, a) \in X^s$ solves $f(x) = 0$. This in particular implies that

$$a_k = \frac{1}{2k}(c_{k-1} - c_{k+1})$$

and hence we obtain uniformly in k that

$$\|a_k\|\omega_k^{s+1} \leq \frac{\omega_k^{s+1}}{2k}(1+2^s)\frac{C}{\omega_k^s} = \frac{1+2^s}{2}C < \infty.$$

This implies $a \in X^{s+1}$*. As a matter of fact* $1 \leq s_1 \leq s_2$ *implies* $\omega_k^{s_1} \leq \omega_k^{s_2}$ *and we know that*

$$X^{s_1} \subset X^{s_2} \tag{3.54}$$

for all $1 \leq s_1 \leq s_2$*. Hence* $a \in X^{s+1}$ *implies* $a \in X^{\tilde{s}}$ *for all* $s \leq \tilde{s} \leq s+1$*. Inductively we obtain that* $a \in X^{s_0}$ *for all* $s_0 \geq s$ *and by* (3.54) $a \in X^{s_0}$ *for all* $s_0 \geq 1$*. Finally we obtain by construction that if* $f(\theta, a) = 0$ *with* f *either given by* (3.50) *or* (3.51)*, then* u *defined by* (3.43) *solves* $F(\theta, u) = 0$ *with* F *given by* (3.23) *or* (3.24) *respectively.* □

Definition of the fixed point operator and radii polynomials As alluded to above the strategy to find solutions $x \in X^s$ of

$$f(x) = 0$$

with f generally given by either (3.50) or (3.51), is to consider an equivalent fixed point operator $T : \mathcal{X} \to \mathcal{X}$ whose fixed points are in one-to-one correspondence with the zeros of f. More precisely, we choose $\mathcal{X} = X^s$ and let the operator T be a Newton-like operator about an approximate solution $\bar{x}$ of f.

In order to compute this numerical approximation we introduce a Galerkin projection. Let $m > 1$ and define the finite dimensional projection $\Pi_m : X^s \to X^s_m$ by $\Pi_m x = (x_k)_{k=k_0}^{m-1}$. The Galerkin projection of f is defined by

$$f^{(m)} : X^s_m \to X^s_m : x_F \mapsto \Pi_m f(x_F, 0_\infty), \tag{3.55}$$

where $0_\infty = (I - \Pi_m)0$. Identifying $(x_F, 0_\infty)$ with $x_F \in X^s_m \cong \mathbb{R}^{p+dm}$ we think of $f^{(m)} : \mathbb{R}^{p+dm} \to \mathbb{R}^{p+dm}$. Now assume that we have computed $\bar{x}_F \in \mathbb{R}^{p+dm}$ such that $f^{(m)}(\bar{x}_F) \approx 0$ and let $\bar{x} = (\bar{x}_F, 0_\infty) \in X^s$. Let $B_{\bar{x}}(r) = \bar{x} + B(r)$, the closed ball in X^s of radius r centered at $\bar{x}$, where

$$B(r) = \left\{ x \in X^s : \|x\|_s = \sup_{k \geq k_0}\{\|x_k\|_\infty \omega_k^s\} \leq r \right\} = \prod_{k \geq k_0} \left[-\frac{r}{\omega_k^s}, \frac{r}{\omega_k^s} \right]^{n(k_0)}, \tag{3.56}$$

where $n(-1) = p$ and $n(k) = ds$ for $k \geq 0$. In order to define the fixed point operator T, we introduce $A_m \approx \left(Df^{(m)}(\bar{x}_F)\right)^{-1}$ a numerical inverse of $Df^{(m)}(\bar{x}_F)$. Assume that the finite dimensional matrix A_m is invertible (this hypothesis can be rigorously verified with interval arithmetic). Define the linear invertible operator $A : X^s \to X^{s+1}$ by

$$(Ax)_k = \begin{cases} (A_m(\Pi_m x))_k, & k = k_0, \dots, m-1 \\ \frac{1}{2k} x_k, & k \geq m. \end{cases} \tag{3.57}$$

Finally define the Newton-like operator $T : X^s \to X^s$ about the numerical solution $\bar{x}$ by

$$T(x) = x - Af(x). \tag{3.58}$$

The goal is to determine (if possible) a positive radius r of the ball $B_{\bar{x}}(r)$ so that $T : B_{\bar{x}}(r) \to B_{\bar{x}}(r)$ is a contraction. Assuming that such $r > 0$ exists, an application of the contraction mapping theorem yields the existence of a unique fixed point $\tilde{x}$ of T within the closed ball $B_{\bar{x}}(r)$. By invertibility of the linear operator A, one can conclude that $\tilde{x}$ is the unique solution of $f(x) = 0$ in the ball $B_{\bar{x}}(r)$. By construction, this unique solution represents a solution $u(t)$ of the general operator equation (3.26) with the second component either given by (3.23) or (3.24). Hence, all we need to do is to find $r > 0$ such that $T : B_{\bar{x}}(r) \to B_{\bar{x}}(r)$ is a contraction. This task is achieved with the notion of the radii polynomials (originally introduced in [15] to compute equilibria of PDEs), which provide an efficient way of constructing a set on which the contraction mapping theorem is applicable. As in section 3.2.1 their construction depends on the Y and Z bounds. Consider the bound $Y = (Y_k)_{k \geq k_0}$ satisfying

$$\left| \left[T(\bar{x}) - \bar{x} \right]_k \right| \preceq Y_k, \ \ k \geq k_0, \tag{3.59}$$

where the inequality is taken component-wise and where $Y_k \in \mathbb{R}^d_+$ for $k \geq 0$. If $k_0 = -1$, then $Y_{k_0} \in \mathbb{R}^p_+$. Consider the bound $Z(r) = (Z_k(r))_{k \geq k_0}$ satisfying

$$\sup_{\xi_1, \xi_2 \in B(r)} \left| \left[DT(\bar{x} + \xi_1)\xi_2 \right]_k \right| \preceq Z_k(r), \ \ k \geq k_0, \tag{3.60}$$

where again the inequality is taken component-wise and where $Z_k(r) \in \mathbb{R}^d_+$ for $k \geq 0$. If $k_0 = -1$, then $Z_{k_0}(r) \in \mathbb{R}^p_+$. If the vector field in (2.1) is polynomial, then it is possible to obtain a polynomial expansion in r for $Z_k(r)$. As a matter of fact, in this case, the degree of the polynomial $Z_k(r)$ is the same as the degree of the polynomial vector field $g(u)$. The reason for this can be found in Remark 3.2.4 in equation (3.46). A more detailed description of how (3.46) is involved in this fact can be found in the concrete derivation of the Z bounds in the application Section 4. Otherwise, that is if the analytic vector field $g(u)$ is not polynomial, a Taylor expansion can be considered in order to obtain a polynomial expression in r for $Z_k(r)$.

In addition to the analogues of the hypothesis **RP1.** to **RP4.** in the Spline case we make the following important assumptions. Assume that there exists a number $M \geq m$ where m is the *dimension* of the Galerkin projection (3.55) such that the bounds Y and Z satisfying (3.59) and (3.60) are such that

A1. $Y_k = 0 \in \mathbb{R}^n$ for all $k \geq M$.

A2. There exists a *uniform* polynomial bound $\bar{Z}_M(r)$ such that for all $k \geq M$,

$$Z_k(r) \preceq \frac{\bar{Z}_M(r)}{\omega_k^s}. \tag{3.61}$$

Before introducing the radii polynomials, let us briefly talk about the two above assumptions. If the vector field $g(u)$ is polynomial, then the nonlinear terms $c_k(\bar{a})$ are convolutions terms of the form

$$((\bar{a})_{j_1}(\bar{a})_{j_2} \cdots (\bar{a})_{j_\ell})_k$$

which are eventually equal to zero for large enough k since $\bar{a}_k = 0$ for $k \geq m$. Hence, by construction of A defined in (3.57) and of the bound Y as in (3.59), there exists M such that Y_k can be defined to be $0 \in \mathbb{R}^d$ for $k \geq M$. Again in case the vector field $g(u)$ is polynomial, there are some analytic convolution estimates (e.g. the ones developed in [24]) that allow computing $\bar{Z}_M(r)$ satisfying (3.61). The computation of the uniform polynomial bound $\bar{Z}_M(r)$ is presented explicitly in the examples of Section 4.

Definition 3.2.2 *Denote by $\mathbf{1_d} \in \mathbb{R}^d$ the vector whose components are all* 1. *We define the finite radii polynomials $(p_k(r))_{k\geq k_0}^{M-1}$ by*

$$p_k(r) = Y_k + Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d}, \;\; k = k_0, \ldots, M-1, \tag{3.62}$$

and the tail radii polynomial by

$$p_M(r) = \bar{Z}_M(r) - r\mathbf{1_d}. \tag{3.63}$$

The following result justifies the construction of the radii polynomials of Definition 3.2.2.

Theorem 3.2.2 *If there exists $\bar{r} > 0$ such that $p_k(\bar{r}) \prec 0$ for all $k = k_0, \dots, M$, then $T : B_{\bar{x}}(\bar{r}) \longrightarrow B_{\bar{x}}(\bar{r})$ is a contraction and therefore there exists a unique $\tilde{x} \in B_{\bar{x}}(\bar{r})$ such that $T(\tilde{x}) = \tilde{x}$. Hence, there exists a unique $\tilde{x} \in B_{\bar{x}}(\bar{r})$ such that $f(\tilde{x}) = 0$.*

Proof 3.2.3 *See Corollary 3.6 in [24].*

We wish to emphasize that the conditions (3.62) and (3.63) entail that $Y_k + Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d} \prec 0$ for all $k \geq k_0$. We summarize this in the following Proposition.

Proposition 3.2.1 *Let $p_k(r)$, $k \geq k_0$ be given by* (3.62) *and $p_M(r)$ by* (3.63)*. If $p_k(r) \prec 0$ for $k = k_0, \dots, M$, then*

$$Y_k + Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d} \prec 0 \quad \text{for all } k \geq k_0. \tag{3.64}$$

Proof 3.2.4 *For $k \geq M$ we have*

$$Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d} = \underbrace{Y_k}_{=0} + Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d}. \tag{3.65}$$

*And hence if $p_M(r) \prec 0 \Leftrightarrow \bar{Z}_M(r) \prec r\mathbf{1_d}$ we get by assumption **A2.** that*

$$Z_k(r) \preceq \frac{\bar{Z}_M(r)}{\omega_k^s} \prec \frac{r}{\omega_k^s}\mathbf{1_d}. \tag{3.66}$$

Using (3.65), (3.66) *implies*

$$Y_k + Z_k(r) - \frac{r}{\omega_k^s}\mathbf{1_d} \prec 0 \tag{3.67}$$

for all $k \geq M$. Together with (3.62) *this yields our claim. Note that this is essential for the proof in [24] and can be seen as one of the key steps in controlling the infinite dimensional error.*

In summary the strategy to rigorously compute solutions of $f(x) = 0$ with f either given by (3.50) or (3.51) is therefore to construct the radii polynomials of Definition 3.2.2, to verify (if possible) the hypothesis of Theorem 3.2.2, and to use the result of Lemma 3.2.2 to conclude that $u(t) = a_0 + 2\sum_{k\geq 1} a_k T_k(t)$ is a solution of $F = 0$ where F is either given by (3.23) or (3.24) respectively.

While the computation of the bound Y satisfying (3.59) is rather straightforward, the computation of the polynomial bound $Z(r)$ satisfying (3.60) is more involved. In order to simplify its computation, we introduce the linear invertible operator $A^\dagger : X^s \to X^{s+1}$ by

$$(A^\dagger x)_k = \begin{cases} (Df^{(m)}(\bar{x}_F)(\Pi_m x))_k, & k = k_0, \dots, m-1 \\ 2kx_k, & k \geq m. \end{cases} \tag{3.68}$$

and we use the factorization $T(x) = x - Af(x) = (I - AA^\dagger)x - A(f(x) - A^\dagger x)$. Letting $\xi_1 = wr$, $\xi_2 = vr$ with $w, v \in B(1)$, one has that

$$\begin{aligned} DT(\bar{x} + \xi_1)\xi_2 &= (I - AA^\dagger)\xi_2 - A\left(Df(\bar{x} + \xi_1)\xi_2 - A^\dagger \xi_2\right) \\ &= \left[(I - AA^\dagger)v\right] r - A\left(Df(\bar{x} + wr)vr - A^\dagger vr\right), \end{aligned} \tag{3.69}$$

where the first term is of the form ϵr, for $\epsilon = (I - AA^\dagger)v \in X^s$ very small, and where the coefficient of r in $[Df(\bar{x} + wr)vr - A^\dagger vr]_k$ should be small as the dimension of the Galerkin projection m is large. Hence, for m large enough, the coefficient in r of $Z_k(r)$ should be small. That should increase the chances of the coefficient of r in the radii polynomials defined in Definition 3.2.2 to be negative, and therefore increase the chances of verifying the hypothesis of Theorem 3.2.2. We give more details on both the computation of the Y and the Z bounds for the applications in Section 4.

3.3 Results on transversality

In the context of generic connecting orbits it is a natural question if the connecting orbit we proof to exist corresponds to a transversal intersection of the related invariant manifolds. If that is the case we will speak of a transversal connecting orbit. More precisely, suppose that we are in the setting of two hyperbolic fixed points p_1 and p_2 of (2.1) fulfilling the non-degeneracy condition $n_u + n_s = d + 1$ where $n_{u,s}$ are as usual the unstable dimension of p_1 and the stable dimension of p_2 respectively. The common idea of the results is the following.

Recall that our philosophy to compute connecting orbits is to encode the existence of a connecting orbit as the existence of a zero of some operator F defined on a Banach space X. Starting from an approximate solution $\bar{x}$ we ensure the existence of a genuine zero $\tilde{x}$ implying the existence of a connecting orbit together with rigorous normwise bounds on the error $\bar{x} - \tilde{x}$ which can also be used to derive phase space bounds on the connecting orbit itself. The idea to additionally ensure that the intersection

of the unstable and stable manifold is transversal consists in showing that transversality follows from the invertibility of the derivative $DF(\tilde{x})$ at the solution $\tilde{x}$. Note that we only know a ball in which $\tilde{x}$ is contained but do not know it exactly. We will nevertheless be able to make a statement about the invertibility of the derivative at that point and as a consequence about the transversality of the connections.
Before we investigate our different approaches more closely let us recall the notion of transversality. As we work in euclidian space $\mathbb{R}^d$ we restrict to the definition for that particular case.

Definition 3.3.1 (Transversality) *Suppose we are given two submanifolds M, N of $\mathbb{R}^d$. Then these two submanifolds intersect transversally if for every point $x \in M \cap N$ the tangent spaces T_xM and T_xN generate $\mathbb{R}^d$. In more detail we assume that for any particular $x \in M \cap N$, if $v_1, \dots, v_m$ is a basis for T_xM and $w_1, \dots, w_n$ for T_xN then we have that $v_1, \dots, v_m, w_1, \dots, w_n$ is a basis for $\mathbb{R}^d$.*

Note that in the general definition for submanifolds of a general manifold we need to have that the tangent spaces at every intersection point generate the tangent space of the ambient manifold at that point.

3.3.1 Discretization free approach

Suppose we are in the general setting of Theorem 3.1.1. Then we have the following result.

Corollary 3.3.1 (Bounded Invertibility of the Inverse) *Let $\|DF(\bar{x})^{-1}\|_{B(X)}$, $\bar{x}$, ϵ_{NK}, κ, r, and $\tilde{x}$ be as Theorem 3.1.1. In addition suppose that $r \leq 4\epsilon_{NK}$ and that the strict inequality*

$$4\epsilon_{NK}\kappa\|DF(\bar{x})^{-1}\|_{B(X)} < 1, \tag{3.70}$$

is satisfied. Let M be any constant with $4\epsilon_{NK}\kappa\|DF(\bar{x})^{-1}\|_{B(X)} \leq M < 1$. Then $DF(\tilde{x})$ is invertible and

$$\|DF(\tilde{x})^{-1}\|_{B(X)} \leq \frac{\|DF(\bar{x})^{-1}\|_{B(X)}}{1-M}. \tag{3.71}$$

Proof 3.3.1 *As a preparation we realize that*

$$\begin{aligned} DF(\tilde{x}) &= DF(\tilde{x}) + DF(\bar{x}) - DF(\bar{x}) = \\ &= DF(\bar{x})\left(DF(\bar{x})^{-1}DF(\tilde{x}) + I - I\right) = \\ &= DF(\bar{x})\left(I - (I - DF(\bar{x})^{-1}DF(\tilde{x}))\right) = \\ &= DF(\bar{x})\left(I - DF(\bar{x})^{-1}(DF(\bar{x}) - DF(\tilde{x}))\right). \end{aligned}$$

This motivates applying a Neumann Series argument to the expression

$$DF(\tilde{x}) = DF(\bar{x})\left[I - DF(\bar{x})^{-1}\left(DF(\bar{x}) - DF(\tilde{x})\right)\right].$$

Recalling Remark 3.1.2 we require to this end that

$$\|DF^{-1}(\bar{x})(DF(\bar{x}) - DF(\tilde{x}))\|_{B(X)} < 1.$$

Using the definition of κ and the fact that we require $r \leq 4\epsilon_{NK}$, we estimate in the following way:

$$\begin{aligned}\|DF(\bar{x})^{-1}(DF(\bar{x}) - DF(\tilde{x}))\|_{B(X)}\| &\leq \|DF(\bar{x})^{-1}\|_{B(X)}\|DF(\bar{x}) - DF(\tilde{x})\|_{B(X)}\\ &\leq \|DF(\bar{x})^{-1}\|_{B(X)}\kappa\|\bar{x} - \tilde{x}\|\\ &\leq \|DF(\bar{x})^{-1}\|_{B(X)}\kappa r\\ &\leq \|DF(\bar{x})^{-1}\|_{B(X)}\kappa 4\epsilon_{NK}.\end{aligned}$$

Thus by using (3.70)

$$\|DF^{-1}(\bar{x})(DF(\bar{x}) - DF(\tilde{x}))\|_{B(X)} < 1,$$

follows and the Neumann series Theorem yields invertibility of $DF(\tilde{x})$ together with (3.71). □

We now turn to the concrete consideration of the discretization free connection operator F given by (3.16) to compute generic connecting orbits between the hyperbolic equilibria $p_{1,2}$. Recall that F is defined on $X \cong \mathbb{R}^d$, where $d = n_s + n_u - 1$ with the stable and unstable dimensions $n_{u,s}$. Assume an approximate solution $\bar{x} = (\bar{\alpha}, \bar{\phi})$ with $F(\bar{\alpha}, \bar{\phi})$ to be given together with an exact solution $\tilde{x} = (\tilde{\alpha}, \tilde{\phi})$ obtained via the discretization-free procedure based on Theorem 3.1.1 and introduced in detail in Section 3.1.2. Corollary 3.3.1 ensures that $DF(\tilde{x})$ is invertible. We wish to connect this fact to the transversality of the intersection of the unstable and stable manifold of p_1 and p_2. Let us therefore investigate $DF(x)$. Using (3.16) we obtain

$$DF(x) = [\underbrace{DQ(\Theta(\alpha))D\Theta(\alpha)}_{\in\mathbb{R}^{d,n_u-1}}\,|\,\underbrace{-DP(\phi)}_{\in\mathbb{R}^{d,n_s}}] \in \mathbb{R}^{d,d}.$$

First using Corollary 3.3.1 we obtain that $DF(\tilde{x})$ is invertible. The idea is to show that the columns of $DF(\tilde{x})$ span the cartesian product

$$T_{Q(\Theta(\tilde{\alpha}))}W^u(p_1) \times T_{P(\tilde{\phi})}W^s(p_2).$$

By invertibility of $DF(\tilde{x})$ they span $\mathbb{R}^d$ and the tangent spaces would hence span $\mathbb{R}^d$. This entails transversality of the intersection. The idea of the following Theorem is to show that the vector $DQ(\Theta(\tilde{\alpha}))J_u\Theta(\tilde{\alpha})$, which completes the columns of $DQ(\Theta(\alpha))D\Theta(\alpha)$ to a basis of $T_{Q(\Theta(\tilde{\alpha}))}W^u(p_1)$ fulfills

$$DQ(\Theta(\tilde{\alpha}))J_u\Theta(\tilde{\alpha}) \in span(DP(\tilde{\phi}))$$

with J_u given in (2.7). Once this is shown the transversality follows. The main tool we use will be the invariance equation that the parametrizations P and Q fulfill by definition. Let us now state the result.

Theorem 3.3.1 *Suppose that the hypotheses of Theorem 3.1.2 are satisfied. In particular assume the validation values defined in Definition 3.1.2 and a corresponding zero $\tilde{x}$ to be given. If in addition the strict inequality given by*

$$\frac{4\epsilon_{NK}\kappa C_1}{1-M} < 1,$$

is fulfilled, then the connecting orbit from p_1 to p_2 through $\tilde{x} \in \mathbb{R}^d$ is transverse.

***Proof* 3.3.2** *Note that $P[\Theta_v(\tilde{\alpha})] = Q(\tilde{\phi})$ and denote $\tilde{x}(t)$ the orbit $\tilde{x}$ starting at $t = 0$. Then*

$$\tilde{x}'(0) = g[\tilde{x}(0)] = g[Q(\Theta_v(\tilde{\alpha}))] = DQ(\Theta_v(\tilde{\alpha}))J_u\Theta_v(\tilde{\alpha}),$$

where the last equality follows from the invariance equation (2.10) *for the parameterization and also*

$$\tilde{x}'(0) = g[\tilde{x}(0)] = g[P(\tilde{\phi})] = DP(\tilde{\phi})J_s\tilde{\phi}.$$

The matrices $J_{u,s}$ are defined in equation (2.7) *for the unstable and stable eigenvalues respectively. Then we have*

$$DP(\tilde{\phi})J_s\tilde{\phi} = DQ(\Theta_v(\tilde{\alpha}))J_u\Theta_v(\tilde{\alpha}). \tag{3.72}$$

This completes the proof.

3.3.2 Spline approach

Concerning the Spline approach for generic heteroclinic orbits we wish to take the same path as for the discretization free approach. That is we show the following: given a zero $\tilde{x} \in X = \mathbb{R}^p \times C_0([0,1],\mathbb{R}^d)$ of F as given in (3.29) that was shown to exist via Theorem 3.2.1, then the Fréchet derivative $DF(\tilde{x})$ is an invertible linear operator. Then we go on to show that this invertibility implies transversality of the connection. The following Corollary achieves the first step.

Corollary 3.3.2 *Assume that the hypotheses of Theorem 3.2.1 are satisfied and consider $\tilde{x}$ to be the unique fixed point of T within $B_{\bar{x}}(r) = \bar{x} + B(r,\omega)$. Then the linear operator $DF(\tilde{x}) : X \to X$ is invertible.*

Proof 3.3.3 *Recalling* (3.31) *and* (3.32)*, we define a norm on X as follows. Given $x \in X$, consider the weighed norm on X by*

$$\|x\|_X = \max\left\{ \|\Pi_m x\|_{X_m} \, , \, \frac{1}{\omega}\|\Pi_\infty x\|_{X_\infty} \right\}. \tag{3.73}$$

Recalling (3.35)*, the closed unit ball in X with respect to norm* (3.73) *is $B(1,\omega)$. Then, letting $\bar{x}$ be the center of the ball $B_{\bar{x}}(r)$, there exists $x_1 \in B(r,\omega)$ such that $\tilde{x} = \bar{x} + x_1$. Recalling* (3.41) *and* (3.42) *and from the fact that each radii polynomial $p_l(r) prec 0$, we then get that*

$$\begin{aligned}
\|DT(\tilde{x})\|_X &= \sup_{x \in B(1,\omega)} \|DT(\tilde{x})x\|_X = \frac{1}{r} \sup_{x \in B(1,\omega)} \|DT(x_* + x_1)xr\|_X \\
&= \frac{1}{r} \sup_{x \in B(1,\omega)} \left\{ \max\left\{ \|\Pi_m DT(x_* + x_1)xr\|_{X_k} \, , \, \frac{1}{\omega}\|\Pi_\infty DT(x_* + x_1)wr\|_{X_\infty} \right\}\right\} \\
&\leq \max\left\{ \frac{(Z_0)_1}{r}, \dots \frac{(Z_0)_p}{r}, \frac{(Z_1)_1}{r}, \dots, \frac{(Z_1)_{m+1}}{r} \dots \frac{(Z_d)_1}{r}, \dots, \frac{(Z_d)_{m+1}}{r}, \frac{Z_\infty}{r} \right\} \\
&< 1.
\end{aligned}$$

Using Neumann series, we get that the operator $I - DT(\tilde{x}) : X \to X$ is invertible. Since

$$T(x) = (\Pi_m - A_m \Pi_m F)(x) + \Pi_\infty(F(x) + (x)) = x - A_m \Pi_m F(x) + \Pi_\infty F(x),$$

then

$$I - DT(\tilde{x}) = -A_m \Pi_m DF(\tilde{x}) + \Pi_\infty DF(\tilde{x}).$$

Suppose that there exists $y \in X$ such that $DF(\tilde{x})y = 0$. Then $\Pi_m DF(\tilde{x})y = 0$ ($^{A_m} \overset{\text{invertible}}{\Longleftrightarrow} -A_m \Pi_m DF(\tilde{x})y = 0$) and $\Pi_\infty DF(\tilde{x})y = 0$. Hence

$$[I - DT(\tilde{x})]y = -A_m \Pi_m DF(\tilde{x})y + \Pi_\infty DF(\tilde{x})y = 0$$

which implies that $y = 0$ by invertibility of $I - DT(\tilde{x})$. That implies that $DF(\tilde{x})$ is injective. We want to show that $DF(\tilde{x})$ is surjective. Consider $w \in X$ (we want to construct $y \in X$ such that $w = DF(\tilde{x})y$). Let $w_m \overset{\text{def}}{=} \Pi_m w$ and $w_\infty \overset{\text{def}}{=} \Pi_\infty w$. Define $z_m = -A_m w_m$, $z_\infty = w_\infty$ and $z = z_m + z_\infty \in X$. We know by surjectivity of $I - DT(\tilde{x})$ that there exists $y \in X$ such that

$$z = [I - DT(\tilde{x})]y = -A_m \Pi_m DF(\tilde{x})y + \Pi_\infty DF(\tilde{x})y.$$

Hence, $z_m = \Pi_m z = -A_m \Pi_m DF(\tilde{x})y$ and $z_\infty = \Pi_\infty z = \Pi_\infty DF(\tilde{x})y$. The invertibility of A_m (see **RP4** *above) implies that $w_m = -(A_m)^{-1} z_m = \Pi_m DF(\tilde{x})y$. We can therefore conclude that $w = w_m + w_\infty = \Pi_m DF(\tilde{x})y + \Pi_\infty DF(\tilde{x})y = DF(\tilde{x})y$.*

Next suppose that $(\tilde{\alpha}, \tilde{\phi}, \tilde{u}) \in X$ is a zero of the operator F given by (3.29). Having Corollary 3.3.2 in mind we need to obtain the Fréchet derivative of the operator $F \colon X \to X$. Consider $(\alpha_1, \phi_1, u_1) \in X$. Computing the difference

$$F(\tilde{\alpha} + \alpha_1, \tilde{\phi} + \phi_1, \tilde{u} + u_1) - F(\tilde{\alpha}, \tilde{\phi}, \tilde{u}),$$

and neglecting the terms which are quadratic in (α_1, ϕ_1, u_1) leads to

$$DF[\tilde{\alpha}, \tilde{\phi}, \tilde{u}](\alpha_1, \phi_1, u_1) =$$
$$\begin{pmatrix} DQ(\tilde{\phi})\phi_1 - DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha})\alpha_1 - L\int_0^1 Dg[\tilde{u}(\tau)]u_1(\tau)\,d\tau \\ u_1(t) - DP(\Theta_\nu(\tilde{\alpha})D\Theta_\nu(\tilde{\alpha})\alpha_1 - L\int_0^t Dg[\tilde{u}(\tau)]u_1(\tau)\,d\tau \end{pmatrix}. \tag{3.74}$$

The next step in order to apply Corollary 3.3.2 is to characterize the kernel of DF.

Lemma 3.3.1 $(\alpha_1, \phi_1, u_1) \in ker(DF(\alpha, \phi, u))$ *if and only if*

$$u_1'(t) = LDg[u(t)]u_1(t), \tag{3.75}$$

$$u_1(0) = DP(\Theta_\nu(\alpha)D\Theta_\nu\alpha)\alpha_1 \quad \text{and} \quad u_1(1) = DQ(\phi)\phi_1. \tag{3.76}$$

Proof 3.3.4 *The proof follows by rewriting* (3.75) *in integral form and taking the boundary conditions* (3.76) *at $t = 0$ and at $t = 1$ into account.*

Theorem 3.3.2 *Suppose that $(\tilde{\alpha}, \tilde{\phi}, \tilde{u}) \in X$ is a zero of F, and that $DF(\tilde{\alpha}, \tilde{\phi}, \tilde{u})$ is invertible. Then the intersection of $W^u(p_1)$ and $W^s(p_2)$ is non-empty and transverse on orbit$(Q[\tilde{\phi}])$.*

Proof 3.3.5 *Let $\tilde{z} = P(\Theta_\nu(\tilde{\alpha})) \in W^u(p_1)$ and define $\hat{y} = \Phi(\tilde{z}, 1) = \tilde{u}(1)$. Then by the fact $F(\tilde{\alpha}, \tilde{\theta}, \tilde{u}) = 0$ it follows $\tilde{y} = Q(\tilde{\phi}) \in W^s(p_2)$. Furthermore it follows from the flow invariance of $W^u(p_1)$ and $W^s(p_2)$ that orbit$(\tilde{z})$ = orbit$(\tilde{y}) \subset W^u(p_1) \cap W^s(p_2)$, so that the intersection is non-empty.*

$\Phi(\tilde{z}, t) \in W^u(p_1)$ for any $t \in \mathbb{R}$, and by the chain rule

$$D_\alpha \Phi(P(\Theta_\nu(\tilde{\alpha})), t) = D\Phi(\tilde{z}, t)DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha}).$$

Then the columns of this matrix span the linear subspace of $T_{\Phi(\tilde{z},t)}W^u(p_1)$ perpendicular to the orbit of $\tilde{z}$ for any $t \in [0, 1]$. Since the columns of $-DQ(\tilde{\phi})$ span $T_{\tilde{y}}W^s(p_2)$ (and since the orbit passes through $\tilde{y}$) we have that the columns of the matrix

$$[D\Phi(\tilde{z}, 1)DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha}) \mid -DQ(\tilde{\phi})]$$

span $T_{\hat{y}}W^u(p_1)$ and $T_{\hat{y}}W^s(p_2)$.

Assume for the sake of contradiction that the intersection $W^u(p_1) \cap W^s(p_2)$ is not transverse at $\hat{y}$. Then $T_{\hat{y}}W^u(p_1)$ and $T_{\hat{y}}W^s(p_2)$ do not span $\mathbb{R}^n$ and there is a non-zero vector $\xi = (\xi_1, \xi_2) \in \mathbb{R}^{n_u-1} \times \mathbb{R}^{n_s} = \mathbb{R}^n$ so that

$$[D\Phi(\tilde{z}, 1)DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha}) \mid -DQ(\tilde{\phi})]\xi = 0.$$

or

$$M(\tilde{z}, 1)DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha})\xi_1 = DQ(\hat{\phi})\xi_2,$$

where $M(\tilde{z}, t)$ is the solution of the variational equation

$$\frac{d}{dt}M(\tilde{z}, t) = L\, Dg[\tilde{u}(t)]M(\tilde{z}, t) \qquad M(\tilde{z}, 0) = I.$$

If we define $\alpha_1 = \xi_1$, $\phi_1 = \xi_2$, and take $u_1 \colon [0, 1] \to \mathbb{R}^n$ to be

$$u_1(t) = M(\tilde{z}, t)DP(\Theta_\nu(\tilde{\alpha}))D\Theta_\nu(\tilde{\alpha})\alpha_1 \quad \text{for all} \quad t \in [0, 1],$$

then (α_1, ϕ_1, u_1) solves the boundary value problem (3.75)*. Thus $0 \neq (\alpha_1, \phi_1, u_1) \in ker(DF(\tilde{\alpha}, \tilde{\phi}, \tilde{u}))$ which is a contradiction as we assumed $DF(\tilde{\alpha}, \tilde{\phi}, \tilde{u})$ to be invertible.*

3.3.3 Chebyshev approach

Assume we are in the setting or Theorem 3.2.2 and recall Remark 3.2.4. We first make the following observation.

Lemma 3.3.2 *Let the bounds Y_k and Z_k be defined as in* (3.59) *and* (3.60)*. Define the sequence $Z(r) = (Z_k(r))_{k \geq k_0}$. Assume for some $M > 0$ the radii polynomials $p_k(r)$ $k = k_0, \ldots, M-1$ and $p_M(r)$ to be given as specified in Definition 3.2.2. Finally let the conditions of Theorem 3.2.2 be fulfilled. Then $\|Z\|_s < r$.*

Proof 3.3.6 *Using equation* (3.67) *of Proposition 3.2.4 and* $0 \preceq Y_k$ *for all* $k \geq k_0$ *we get*

$$\|Z_k(r)\| \leq \|Y_k + Z_k(r)\|_\infty < \frac{r}{\omega_k^s}$$

for all $k \geq k_0$. Thus

$$\|Z(r)\|_s = \sup_{k \geq k_0} \|Z_k(r)\|_\infty \omega_k^s < r.$$

□

This enables us to proof the following Theorem.

Theorem 3.3.3 *Assume the conditions of Theorem 3.2.2 to be fulfilled. In particular let $\tilde{x}$ such that $f(\tilde{x}) = 0$. Then $Df(\tilde{x})$ is an invertible linear operator.*

Proof 3.3.7 *Aiming to use a Neumann series argument similar to the one in the proof of Theorem 3.3.2 we first consider the operator norm $\|DT(\tilde{x})\|_s$ of $DT(\tilde{x})$ as map from X^s to itself. We thus compute*

$$\|DT(\tilde{x})\|_s = \sup_{\|v\|_s=1} \|DT(\tilde{x})v\|_s = \sup_{\|v\|_s=1} \|DT(\bar{x}+\xi_1)v\|_s$$

with $\xi_1 \in B_{\bar{x}}(r)$. We continue with

$$\begin{aligned} \|DT(\tilde{x})\|_s &= \frac{1}{r} \sup_{\|v\|_s=1} \|DT(\bar{x}+\xi_1)rv\|_s = \frac{1}{r} \sup_{\|\xi_2\|_s=r} \|DT(\bar{x}+\xi_1)\xi_2\|_s = \\ &= \frac{1}{r} \sup_{\|\xi_2\|_s=r} \sup_{k\geq k_0} \{\|(DT(\bar{x}+\xi_1)\xi_2)_k\|_\infty \omega_k^s\} \leq \\ &\leq \frac{1}{r} \sup_{\|\xi_2\|_s=r} \{\|Z_k(r)\|_\infty \omega_k^s\} \leq \frac{1}{r}\|Z(r)\|_s < \frac{1}{r} r = 1, \end{aligned}$$

where we used the definition of $Z_k(r)$ in 3.60 and the strict inequality follows from Lemma 3.3.2. By a Neumann series argument

$$\mathrm{I} - DT(\tilde{x}) = -ADf(\tilde{x})$$

is invertible. By invertibility of A we obtain that $Df(\tilde{x})$ is invertible. More precisely assume that $B \overset{\text{def}}{=} Df(\tilde{x})$ is not injective. Then there is an $x \neq 0$ with $Bx = 0$. But then $ABx = 0$, as A is injective which is a contradiction to invertibility of AB. To show surjectivity let $y \in X^s$ be given. By surjectivity of AB there is in particular an x such that $ABx = Ay$, which implies by invertibility of A that $Bx = y$. □

Chapter 4

Applications and approach comparison

In this chapter we describe several numerical applications to demonstrate the effectivity of our approach. We start in Section 4.1 by considering symmetric homoclinic orbits in the Gray-Scott system. This system was also considered in [60] where a similar approach to the one presented in this work was taken. The difference is that linear splines were used in the discretization and the formulation of the equivalent operator equation (3.1) used the second order formulation of the system. We first describe the Chebyshev approach to the Gray-Scott system written as a 4D first order system where the symmetry condition is employed to formulate the boundary conditions. In particular we extend some results on the existence of symmetric homoclinics obtained in [60].

Our next case study in Section 4.2 focuses on a generic first order system, the Lorenz equations. The success of our boundary value approach relies on a combination of the parametrization method together with the corresponding discretization. By first presenting applications of our discretization free approach, we take this opportunity to investigate the parametrization method together with the validation of the parametrization computations more closely. In the sequel we concentrate on the rigorous solution of initial value problems in the Lorenz system. This will in particular enable us to scrutinize the dependency of our algorithm on the discretization method in more detail and compare their performance. We finish by giving an application of the spline based boundary value approach in this generic first order system.

4.1 Symmetric connections in the Gray Scott system

We consider the Gray-Scott equation re-scaled by a time factor L given by

$$\begin{aligned} v_1'' &= L^2\left(v_1v_2^2 - \lambda(1-v_1)\right) \\ v_2'' &= L^2\left(\tfrac{1}{\gamma}(v_2 - v_1v_2^2)\right). \end{aligned} \tag{4.1}$$

More precisely, letting $u_1 = v_1$, $u_2 = v_1'$, $u_3 = v_2$, $u_4 = v_2'$ and $u = (u_1, u_2, u_3, u_4)$, we can re-write (4.1) as the following vector field:

$$\frac{du}{dt} = g(u) = \begin{pmatrix} u_2 \\ L^2\left(\lambda u_1 + u_1u_3^2 - \lambda\right) \\ u_4 \\ L^2\left(\frac{1}{\gamma}u_3 - \frac{1}{\gamma}u_1u_3^2\right) \end{pmatrix}. \tag{4.2}$$

(4.2) has a hyperbolic fixed point $q = (1,0,0,0)^T$ where the eigenvalues of $Dg(q)$ are given by

$$\mu_{1,2} = \pm L\sqrt{\lambda} \qquad \mu_{3,4} = \pm L\frac{1}{\sqrt{\gamma}}. \tag{4.3}$$

We first give some background information how (4.1) arises in the context of a reaction diffusion equation modeling a chemical reaction. In particular we concentrate on the physical significance of the symmetric homoclinics that we compute and describe the meaning of the parameters γ and λ. We also elaborate on the mathematical role of the parameters for the existence of exact analytic formulas for symmetric homoclinics as discussed in [27] and for the occurrence of resonances in the eigenvalues of $Dg(p)$. In particular we will find all possible resonances.

Then we describe the implementation of the Chebyshev approach for this problem.

4.1.1 Background on the Gray-Scott system and results

Background The Gray Scott model serves as a model for a continuously fed unstirred autocatalytic reaction. The homoclinic solutions we seek represent non-trivial stationary spatial patterns in the form of pulses. Let us give some more detailed explanations of what this means. Our main source is [27] and references therein.

The Gray-Scott model is a particular case of a two component reaction diffusion system. Reaction diffusion systems serve for example as a mathematical model for situations where chemical substances react with each other at certain rates while being able to distribute spatially in a prescribed manner. We stress that reaction diffusion systems in general have attracted much interest in the context of pattern formation [26, 51] and in the modeling of biological processes [41].

'Two component system' refers to the fact that two chemicals are involved in the reaction. Let us denote those by A and B with concentrations a and b. In addition the term 'continuously fed reaction' means that the substance A is supplied at a certain rate θ. Schematically this is described by

$$\begin{cases} \theta \xrightarrow{k_f} A + 2B \xrightarrow{k_1} 3B, & \text{rate} = k_1 ab^2 \\ \qquad\qquad B \xrightarrow{k_2} C, & \text{rate} = k_2 b, \end{cases} \tag{4.4}$$

where $k_{1,2}$ are positive reaction rates and the rate θ at which a is supplied is assumed to be of the form

$$\theta = k_f(a_0 - a).$$

This means θ is positive if the concentration a is below some threshold a_0 and negative otherwise. The mathematical model in the case where the concentrations $a(x,t)$ and $b(x,t)$ depend on one spatial dimension is given by the semilinear parabolic reaction diffusion PDE

$$\begin{aligned} \frac{\partial a}{\partial t} &= D_A \frac{\partial^2 a}{\partial x^2} - k_1 ab^2 + k_f(a_0 - a) \\ \frac{\partial b}{\partial t} &= D_B \frac{\partial^2 b}{\partial x^2} + k_1 ab^2 - k_2 b, \end{aligned} \tag{4.5}$$

where $D_{A,B}$ are diffusion coefficients of A and B. After a nondimensionalization (4.5) becomes

$$\begin{aligned} \frac{\partial \tilde{v}_1}{\partial t} &= \frac{\partial^2 \tilde{v}_1}{\partial x^2} - \tilde{v}_1 \tilde{v}_2^2 + R(1 - \tilde{v}_1) \\ \frac{\partial \tilde{v}_2}{\partial t} &= d \frac{\partial^2 \tilde{v}_1}{\partial x^2} + \tilde{v}_1 \tilde{v}_2^2 - S\tilde{v}_2 \end{aligned} \tag{4.6}$$

where the dimensionless constants R, S and d are defined by

$$R = \frac{k_f}{k_1 a_0^2}, \quad S = \frac{k_2}{k_1 a_0^2} \quad \text{and} \quad d = \frac{D_B}{D_A}.$$

For details on the rescaling to obtain $\tilde{v}_1, \tilde{v}_2$ we refer to [27]. (4.1) arises when we look for stationary, that is time-independent, solutions of (4.6). More precisely a further scaling is conducted (see [27]) and we obtain

$$\begin{aligned} v_1'' &= v_1 v_2^2 - \lambda(1 - v_1) \\ v_2'' &= \tfrac{1}{\gamma}(v_2 - v_1 v_2^2) \end{aligned} \quad ,$$

where

$$\lambda = \frac{R}{S^2} \quad \text{and} \quad \lambda = Sd \tag{4.7}$$

as the stationary equation from which (4.1) arises via a rescaling of time with a factor L.

Results Next we concentrate on considering the first order system (4.2) equivalent to (4.1) on the interval $[-1,1]$ and seek for evenly symmetric homoclinic solutions to the hyperbolic fixed point $q = (1,0,0,0)$.

Remark 4.1.1 *1. From now on we interpret the interval $[-1,1]$ as 'time' interval even if (4.1) is the stationary equation for the reaction diffusion PDE (4.6). As this is a mere notational issue no confusion should arise.*

2. We recall that for a function $v : \mathbb{R} \to \mathbb{R}$ by 'evenly symmetric' we mean 'evenly symmetric about $t = -1$' in this case, that is $v(-1+t) = v(-1-t)$ for all $t \in \mathbb{R}$, as the we shift the origin in time to $t = -1$ to be in the generic Chebyshev setting.

It is shown in [27] that for parameter values $\gamma\lambda = 1$ and $\lambda > 4$ there exists a family of evenly symmetric homoclinics. More precisely for all (λ, γ) in the parameter set

$$\mathcal{C}_0 = \left\{ (\gamma, \frac{1}{\gamma}) : \ 0 < \gamma < \frac{2}{9} \right\}$$

the functions given by

$$v_1(t) = 1 - \frac{3\gamma}{1 + Q\cosh(\frac{t+1}{\sqrt{\gamma}})} \quad \text{and} \quad v_2(t) = \frac{3}{1 + Q\cosh(\frac{t+1}{\sqrt{\gamma}})}, \tag{4.8}$$

with $Q(\gamma) = \sqrt{1 - \frac{9\gamma}{2}}$, induce an evenly symmetric homoclinic orbit to the fixed point p of (4.2). Let us make two remarks about the consequences of the relationship $\lambda\gamma = 1$.

Remark 4.1.2 *Taking into account the definition (4.7) of λ and γ the relation $\lambda\gamma = 1$ means that*

$$\frac{k_f}{D_B} = \frac{k_2}{D_A},$$

thereby relating the supply rate k_f of A and the reaction k_2 at which B disappears in terms of the diffusion rates $D_{A,B}$.

Remark 4.1.3 ***Resonances***
Recalling (2.1.6) *a resonance between the two stable eigenvalues* μ_2 *and* μ_4 *defined in* (4.3) *is given if there are positive integers* $m, n \geq 0$ *such that*

$$m\mu_2 + n\mu_4 = \mu_j \;\; j = 2, 4$$

Assuming $0 < \gamma < \frac{2}{9}$ *and*

$$\lambda\gamma = 1 \qquad \lambda > 4,$$

we observe that from $\lambda = \frac{1}{\gamma}$ *we have that* $\mu_2 = \mu_4$ *and hence there is a resonance between the stable eigenvalues* μ_2 *and* μ_4 *by setting* $m = 1$ *and* $n = 0$ *or vice versa. If one considers the general case of resonances between* μ_2 *and* μ_4 *we must have*

$$n\sqrt{\lambda} + m\frac{1}{\sqrt{\gamma}} = \sqrt{\lambda} \qquad \textit{or} \qquad n\sqrt{\lambda} + m\frac{1}{\sqrt{\gamma}} = \frac{1}{\sqrt{\gamma}}$$

for integers $m, n \geq 0$ *or equivalently*

$$\sqrt{\lambda\gamma} = \underbrace{\frac{-m}{n-1}}_{<0, unless \;\; n=0} \quad (n \neq 1) \qquad \textit{or} \qquad \sqrt{\lambda\gamma} = \underbrace{-\frac{m-1}{n}}_{<0, unless \;\; m=0} \quad (n \neq 0).$$

Hence we can get two families of resonance curves that contain all possible parameter values at which resonances can occur. More concretely these are given by

$$\mathcal{C}_m = \left\{(\lambda, \gamma) : \lambda\gamma = (m+1)^2\right\} \quad \textit{and} \quad \mathcal{C}_{\frac{1}{n}} = \left\{(\lambda, \gamma) : \lambda\gamma = \frac{1}{n^2}\right\}$$

where $m \geq 0$ *and* $n \geq 2$.

Furthermore Theorem C in [27] ascertains that the homoclinics persist if $\lambda\gamma = 1 + \epsilon$ for some positive ϵ and [60] investigates to a certain extent the magnitude of ϵ. More concretely Theorem 1.1 in [60] shows the existence of 30 homoclinic orbits on the line $\gamma = 0.15$ in parameter space. Let us now formulate a result guaranteeing the existence of 297 homoclinics for $\gamma \in \{0.14, 0.15, 0.16\}$, and for several different values of λ. As a preparation we introduce the notation

$$\Lambda^{\pm}_{\mathcal{I},\Delta_\lambda}(\gamma) = \left\{(\gamma, \lambda) : \lambda = \frac{1 \pm k\Delta_\lambda}{\gamma}, \; k \in \mathcal{I}\right\},$$

for an index set $\mathcal{I}$.

γ	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20
$L^+(\gamma)$	0.45	0.50	0.55	0.60	0.60	0.60	0.65	0.70	0.75	0.75	0.75
$L^-(\gamma)$	0.50	0.55	0.55	0.55	0.60	0.60	0.65	0.65	0.70	0.70	0.7
$K^+(\gamma)$	90	90	90	90	90	90	90	90	90	90	90
$K^-(\gamma)$	18	16	15	12	11	9	7	6	3	2	1

Table 4.1: Values of L and K in dependence of γ

Theorem 4.1.1 *Let $\Delta_\lambda = 0.03$ and $\gamma_i = 0.14 + (i-1)0.01$ for $i = 1,2,3$. Set $\mathcal{I}^+(\gamma_i) = \{1,\dots,90\}$ for $i = 1,2,3$ and $\mathcal{I}^-(\gamma_i) = \{1,\dots,K^-(\gamma_i)\}$ where $K^-(\gamma_i)$ is specified in Table 4.1 for $i = 1,2,3$. If*

$$(\lambda,\gamma) \in \bigcup_{i=1}^{3} \Lambda^+_{\mathcal{I}^+(\gamma_i),\Delta_\lambda}(\gamma_i) \cup \Lambda^-_{\mathcal{I}^-(\gamma_i),\Delta_\lambda}(\gamma_i),$$

there exists a ball $B_{\bar{x}}(\bar{r}_{\gamma,\lambda}) \subset X^s$ (with $f_{\gamma,\lambda}(\bar{x}) \approx 0$) containing a unique solution $\tilde{x} = (\tilde{\theta},\tilde{a})$ of $f_{\gamma,\lambda}(x) = 0$ corresponding to an even homoclinic solution of (4.1).

For a geometric representation of the result of Theorem 4.1.1, we refer to Figure 4.2 and Figure 4.3. The rigorous verification of Theorem 4.1.1 can be found in the MATLAB programs *proofLambdaplusγ.m* and *proofLambdaminusγ.m* with $\gamma = 014, 015, 016$, and relies on Theorem 3.2.2. All codes can be downloaded from [28]. The programs make use of the package Intlab [48] for the interval computations and of the package Chebfun [55]. Chebfun is used to compute the Chebyshev coefficients of the exact solutions (4.8) from which a continuation is performed. The main prerequisite for applying Theorem 3.2.2 is the construction of the radii polynomials (3.62) and (3.63).

Beside these rigorously verified homoclinic solutions we investigated a bigger region in parameter space by constructing the radii polynomials $p_1(r),\dots,p_M(r)$ without interval arithmetic and finding an $r > 0$ such that $p_i(r) \prec 0$ for all $i = 1,\dots,M$. The results are marked in black in Figure 4.1. More precisely set $\Delta_\lambda = 0.03$ and $\gamma_i = 0.10 + (i-1)0.1$ for $i = 1,\dots,11$ and let $K^\pm(\gamma_i)$ be specified by Table 4.1. Define $\mathcal{I}^+_{\mathrm{nr}}(\gamma_i) = \{1,\dots,90\} \cup \{110,\dots,K^+(\gamma_i)\}$ for $i \in \{1,2,3,4,5,8,9,10,11\}$, $\mathcal{I}^+_{\mathrm{nr}}(\gamma_i) = \{110,\dots,K^+(\gamma_i)\}$ for $i = 6,7,8$ and $\mathcal{I}^-_{\mathrm{nr}}(\gamma_i) = \{1,\dots,K^-(\gamma_i)\}$. Note that “nr" stands for non rigorous. We found symmetric homoclinic solutions for

$$(\gamma,\lambda) \in \bigcup_{i=1}^{11} \left(\Lambda^+_{\mathcal{I}^+_{\mathrm{nr}}(\gamma_i),\Delta_\lambda}(\gamma_i) \cup \Lambda^-_{\mathcal{I}^-_{\mathrm{nr}}(\gamma_i),\Delta_\lambda}(\gamma_i)\right).$$

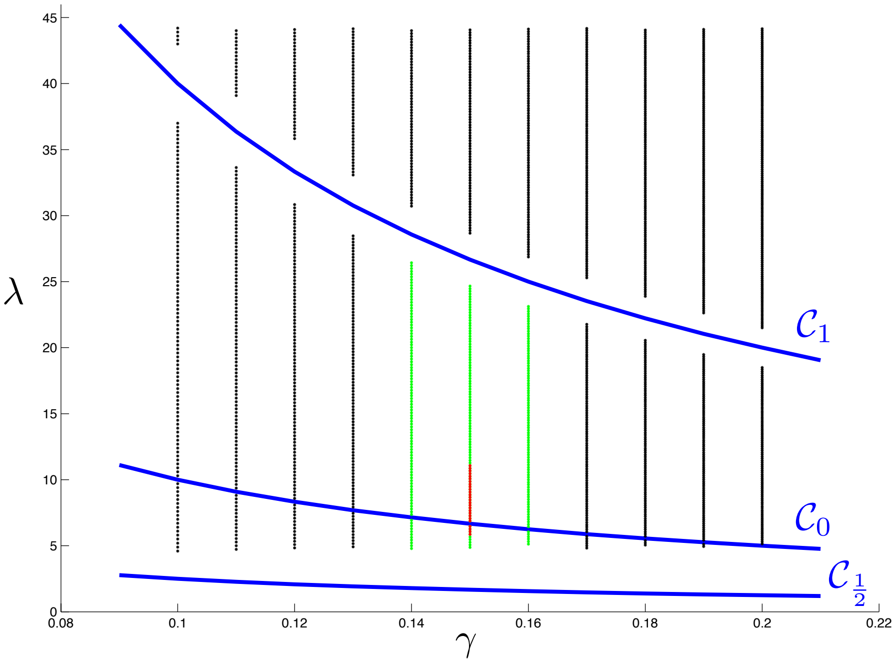


Figure 4.1: The green points indicate the region in parameter space at which the rigorous proof of existence of symmetric homoclinics was obtained by computing the radii polynomials with interval arithmetic. The red points indicate the region investigated in [60]. The black points are investigated using the radii polynomials computed without the use of interval arithmetic. Based on the discussion about resonances, we portrait the curve $\mathcal{C}_1$ and $C_{\frac{1}{2}}$ at which our rigorous method will necessarily fail. Note that $\mathcal{C}_0$ is the curve on which the exact homoclinics (4.8) exist.

These computations are carried out by the MATLAB program *nonrigoroushomoclinics.m* also to be found on [28]. We now give some details on the derivation of the bounds involved in the definition of the radii polynomials.

4.1.2 Formulation of the operator and construction of the radii polynomial bounds for the Chebyshev approach

In order to be able to compute the bounds $Y_{-1}, \ldots, Y_{M-1}, Z_{-1}(r), \ldots, Z_{M-1}(r) \in \mathbb{R}^4$ and $Z_M(r) \in \mathbb{R}^4$ specified in (3.59) and (3.60) we need to specify certain details for the general operator given by (3.51).

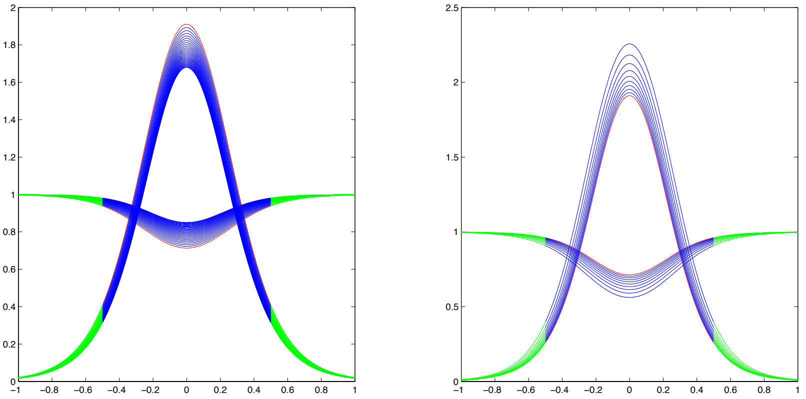


Figure 4.2: Thirty-nine homoclinics from Theorem 4.1.1, where $(\lambda, \gamma) \in \Lambda^{+}_{\{1,\ldots,30\},0.03}(0.15)$ on the left and $(\lambda, \gamma) \in \Lambda^{-}_{\{1,\ldots,9\},0.03}(0.15)$ on the right. The red solution corresponds to the exact homoclinic given by (4.8). Each couple (v_1, v_2) is the center of a ball in function space in which an exact solution is guaranteed to exist. The blue part over $[0, \frac{1}{2}]$ corresponds to the interval $[-1, 1]$ for the operator (4.10), which in turn corresponds to the rescaling of $[0, L^{\pm}(0.15)]$. The green part is added by using the conjugacy relation (see equation (57) in [60]) fulfilled by the parametrization P of $W^{s}_{\text{loc}}(p)$, where we integrate for 2 time units on the time scale of (4.10) and then rescale $[-1, 3]$ to the interval [0,1]. The part over $[-1, 0]$ is obtained using the symmetry.

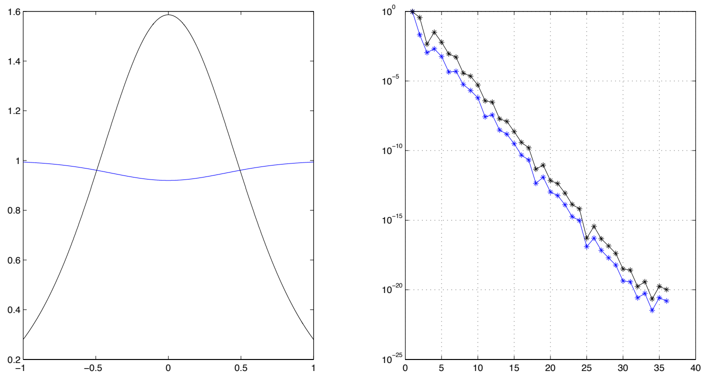


Figure 4.3: **(Left)** Components v_1 (black) and v_2 (blue) of the homoclinic solution of Theorem 4.1.1 corresponding to the parameter value $(\gamma, \lambda) = (0.15, \frac{1+89(0.03)}{0.15}) \in \Lambda^{+}_{\mathcal{I}^{+}(0.15),0.03}(0.15)$. The interval $[0, 1]$ corresponds to the rescaled interval $[-1, 1]$ of (4.10), corresponding itself in turn to a rescaling of $[0, 0.6]$. The interval $[-1, 0]$ is added by symmetry. **(Right)** The Chebyshev coefficients of v_1 (black) and v_2 (blue). Notice the fast decay of the coefficients to zero.

Formulation of the operator f for symmetric homoclinics

The first step is to compute the Chebyshev coefficients c_k of $g \circ u$ where g is given by (4.2). Recalling Remark 3.2.4 which is based on Lemma 2.2.2 we can directly derive that the Chebyshev coefficients (3.45) of (4.2) are given explicitly by

$$c_k = \begin{pmatrix} (a_2)_k \\ L^2\left(\lambda(a_1)_k + (a_1 * a_3^2)_k - \lambda\delta_{k,0}\right) \\ (a_4)_k \\ L^2\left(\frac{1}{\gamma}(a_3)_k - \frac{1}{\gamma}(a_1 * a_3^2)_k\right) \end{pmatrix}, \tag{4.9}$$

where $\delta_{k,0}$ is the Kronecker delta function and where

$$(a_1 * a_3^2)_k = \sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} (a_1)_{|k_1|}(a_3)_{|k_2|}(a_3)_{|k_3|}.$$

To complete the statement of (3.51) we need to specify the boundary conditions $\eta : X^s \to \mathbb{R}^p$.

We are interested in computing symmetric homoclinic orbits between $q = (1,0,0,0)^T$ and itself. Note that q has a two-dimensional stable manifold parametrized by two parameters. This yields that we are in the case $p = 2$. Thus we need to stipulate two additional boundary conditions. We use the fact that we seek even homoclinics to do so. Consider $P(\theta)$ to be a parameterization of the local stable manifold $W^s_{loc}(p)$ at the steady state q. In order to compute P we employ the parametrization method developed in [7, 8, 9] that we summarize in Section 2.1.2.

We thus interpret symmetric homoclinic orbits as solutions of a BVP with the boundary value $u(1) = P(\theta)$, that is $u(1) \in W^s_{loc}(p)$. We impose an even symmetry of the orbit (v_1, v_2) which implies that one imposes $v_1'(-1) = u_2(-1) = 0$ and $v_2'(-1) = u_4(-1) = 0$. Hence, the boundary condition (3.49) reads as $\mathcal{G}(u(-1), u(1)) = (u_2(-1), u_4(-1))^T \in \mathbb{R}^2$, $p_1 = P(\theta)$ and then the operator (3.26) is given by

$$F(\theta, u)(t) = \begin{pmatrix} u_2(-1) \\ u_4(-1) \\ u(t) + \displaystyle\int_t^1 \Psi(u(s))ds - P(\theta) \end{pmatrix}. \tag{4.10}$$

Using $T_k(-1) = (-1)^k$ yields

$$\eta(\theta, a) = ((a_2)_0 + 2\sum_{k=1}^{\infty}(-1)^k(a_2)_k, (a_4)_0 + 2\sum_{k=1}^{\infty}(-1)^k(a_4)_k).$$

Together with (4.9) we hereby obtain an explicit expression $f(x) = f_{\gamma,\lambda}(x)$ for the operator (3.51) tailored to the problem of finding even homoclinics in the Gray-Scott system.

Computation of the *Y*- and *Z*-bounds

Let us now derive the quantities $Y_{-1}, \dots, Y_{M-1}$, $Z_{-1}(r), \dots, Z_{M-1}(r) \in \mathbb{R}^4$ and $Z_M(r) \in \mathbb{R}^4$ specified in (3.59) and (3.60). The important ingredients are summarized in Tables 4.2 and 4.3. For notational brevity we omit the $*$ in the convolutions sums in Tables 4.2 and 4.3 whenever convenient. The technical details of their derivation are our next focus of attention.

Assume a dimension m for the Galerkin projection to be given. We start by explicitly stating the Galerkin projection $f^{(m)} : \mathbb{R}^{4m+2} \to \mathbb{R}^{4m+2}$ specified in (3.55). Let $x_F = (\theta, a_0, \dots, a_{m-1}) = (\theta, a_F) \in \mathbb{R}^{4m+2}$ then $f^{(m)}(x_F)$ is defined by

$$\begin{aligned}
\eta^{(m)}(x_F) &= \begin{pmatrix} (a_2)_0 + 2\sum_{j=1}^{m-1}(-1)^j (a_2)_j \\ (a_4)_0 + 2\sum_{j=1}^{m-1}(-1)^j (a_4)_j \end{pmatrix} \\
f_0^{(m)}(x_F) &= a_0 + \left(c_0^{(m)} + \frac{c_1^{(m)}}{2} - 2\sum_{j=2}^{m-1} \frac{c_j^{(m)}}{j^2-1} \right) - P(\theta) \\
f_k^{(m)}(x_F) &= 2ka_k + \left(c_{k+1}^{(m)} - c_{k-1}^{(m)} \right) \quad k = 1, \dots, m-2 \\
f_{m-1}^{(m)}(x_F) &= 2(m-1)a_{m-1} + \left(\begin{pmatrix} 0 \\ L^2 (a_1 * a_3^2)_m^{(m)} \\ 0 \\ L^2 \left(-\frac{1}{\gamma}(a_1 * a_3^2)_m^{(m)} \right) \end{pmatrix} - c_{m-2}^{(m)} \right)
\end{aligned}$$

where we set

$$c_k^{(m)} = \begin{pmatrix} (a_2)_k \\ L^2 \left(\lambda (a_1)_k - \lambda \delta_{k,0} \right) \\ (a_4)_k \\ L^2 \left(\frac{1}{\gamma}(a_3)_k \right) \end{pmatrix} + \begin{pmatrix} 0 \\ L^2 (a_1 * a_3^2)_k^{(m)} \\ 0 \\ L^2 \left(-\frac{1}{\gamma}(a_1 * a_3^2)_k^{(m)} \right) \end{pmatrix},$$

with the finite convolution sums

$$(a_1 * a_3^2)_k^{(m)} = \sum_{\substack{k_1+k_2+k_3=k \\ |k_i|<m}} (a_1)_{k_1} (a_2)_{k_2} (a_3)_{k_3}.$$

	$k=-1$
$\tilde{Z}_1^{-1}$	$(s-1)\begin{pmatrix}\frac{1}{(m-1)s-1}\\ \frac{1}{(m-1)^{s-1}}\end{pmatrix}$
	$k=0$
$\tilde{Z}_1^{0}$	$\left[\begin{pmatrix}0\\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_0^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_0^M\right]\end{pmatrix}+\epsilon_0^3\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0\\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert^I)_1^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_1^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_1^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_1^M\right]\end{pmatrix}+\frac{\epsilon_1^3}{2}\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}+2\sum_{j=2}^{m-1}\frac{1}{j^2-1}\left(\begin{pmatrix}0\\ L^2(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_0^M\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_0^M\right]\end{pmatrix}+\epsilon_j^3\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}\right)+ +\sum_{j=m}^{M-2}\frac{1}{j^2-1}\left(\frac{1}{\omega_j^s}\begin{pmatrix}1\\ L^2\lambda\\ 1\\ \frac{L^2}{\gamma}\end{pmatrix}+\begin{pmatrix}0\\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)_0^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_0^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)_0^M\right]\end{pmatrix}+\epsilon_j^3\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}\right)+ \frac{2}{((M-1)^2-1)(s-1)(M-2)^{s-1}}\left(\begin{pmatrix}1\\ L^2\lambda\\ 1\\ \frac{L^2}{\gamma}\end{pmatrix}+\begin{pmatrix}0\\ L^2(\Sigma_{33}^{M-1}+2\Sigma_{13}^{M-1})\\ 0\\ \frac{L^2}{\gamma}(\Sigma_{33}^{M-1}+2\Sigma_{13}^{M-1})\end{pmatrix}\right)\right]+\Lambda$
$\tilde{Z}_2^{0}$	$\left[\begin{pmatrix}0\\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_0^M+2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)_0^{M-1}+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_0^M)\\ 0\\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_0^M+2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)_0^{M-1}+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_0^M\right]\end{pmatrix}+2\epsilon_0^3\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0\\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_1^M+2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)_1^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_1^M)\\ 0\\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_1^M+2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)_1^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_1^M\right]\end{pmatrix}+\epsilon_1^3\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}+2\sum_{j=2}^{M-2}\frac{1}{j^2-1}\left(\begin{pmatrix}0\\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_j^M+2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)_j^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_j^M)\\ 0\\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_j^M+2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)_j^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_j^M\right]\end{pmatrix}+2\epsilon_j^3\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}\right)+\frac{2\alpha_{M-1}^3}{((M-1)^2-1)(s-1)(M-2)^{s-1}}\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}\right]$
$\tilde{Z}_3^{0}$	$\left[\begin{pmatrix}0\\ L^2((\vert w_3\vert^2\vert v_3\vert)_0^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_0^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)_0^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_0^M\right]\end{pmatrix}+9\epsilon_0^3\begin{pmatrix}0\\ 1\\ 0\\ \frac{1}{\gamma}\end{pmatrix}+\frac{1}{2}\begin{pmatrix}0\\ L^2((\vert w_3\vert^2\vert v_3\vert)_1^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_1^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)_1^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_1^M\right]\end{pmatrix}+\frac{9}{2}\epsilon_1^3\begin{pmatrix}0\\ L^2\\ 0\\ \frac{L^2}{\gamma}\end{pmatrix}+2\sum_{j=2}^{M-2}\frac{1}{j^2-1}\left(\begin{pmatrix}0\\ L^2((\vert w_3\vert^2\vert v_3\vert)_j^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_j^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)_j^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_j^M\right]\end{pmatrix}+9\epsilon_j^3\begin{pmatrix}0\\ L^2\\ 0\\ \frac{L^2}{\gamma}\end{pmatrix}\right)+\frac{6\alpha_{M-1}^3}{((M-1)^2-1)(s-1)(M-2)^{s-1}}\begin{pmatrix}0\\ L^2\\ 0\\ \frac{L^2}{\gamma}\end{pmatrix}\right]$
	$k=1,\dots,m-1$
$\tilde{Z}_1^{k}$	$\left[\begin{pmatrix}0\\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert^I)_{k+1}^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_{k+1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_{k+1}^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_{k+1}^M\right]\end{pmatrix}+\epsilon_{k+1}^3\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}+\begin{pmatrix}0\\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert^I)_{k-1}^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_{k-1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert^I)_{k-1}^M+2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3^I\vert)_{k-1}^M\right]\end{pmatrix}+\epsilon_{k-1}^3\begin{pmatrix}0\\ L^2(A_3+2A_1A_3)\\ 0\\ \frac{L^2}{\gamma}(A_3+2A_1A_3)\end{pmatrix}\right]$
$\tilde{Z}_2^{k}$	$\left[\begin{pmatrix}0\\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_{k+1}^M+2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)_{k+1}^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_{k+1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_{k+1}^M+2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)_{k+1}^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_{k+1}^M\right]\end{pmatrix}+2\epsilon_{k+1}^3\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}+\begin{pmatrix}0\\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_{k-1}^M+2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)_{k-1}^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_{k-1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)_{k-1}^M+2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)_{k-1}^M+2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)_{k-1}^M\right]\end{pmatrix}+2\epsilon_{k-1}^3\begin{pmatrix}0\\ L^2(4A_3+2A_1)\\ 0\\ \frac{L^2}{\gamma}(4A_3+2A_1)\end{pmatrix}\right]$
$\tilde{Z}_3^{k}$	$\left[\begin{pmatrix}0\\ L^2((\vert w_3\vert^2\vert v_3\vert)_{k+1}^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_{k+1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)_{k+1}^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_{k+1}^M\right]\end{pmatrix}+9\epsilon_{k+1}^3\begin{pmatrix}0\\ L^2\\ 0\\ \frac{L^2}{\gamma}\end{pmatrix}+\begin{pmatrix}0\\ L^2((\vert w_3\vert^2\vert v_3\vert)_{k-1}^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_{k-1}^M)\\ 0\\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)_{k-1}^M+2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)_{k-1}^M\right]\end{pmatrix}+9\epsilon_{k-1}^3\begin{pmatrix}0\\ L^2\\ 0\\ \frac{L^2}{\gamma}\end{pmatrix}\right]$

Table 4.2: Formulas for $\tilde{Z}_l^k$, $k=0,\dots,m-1$

	$m \leq k \leq M-1$
$\tilde{Z}_1^k$	$\Bigg[\frac{1}{\omega^s_{k+1}} \begin{pmatrix} 1 \\ L^2\lambda \\ 1 \\ \frac{L^2}{\gamma} \end{pmatrix} + \begin{pmatrix} 0 \\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert)^M_{k+1} + 2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)^M_{k+1}) \\ 0 \\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert)^M_{k+1} + 2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)^M_{k+1}\right] \end{pmatrix} + \epsilon^3_{k+1} \begin{pmatrix} 0 \\ L^2(A_3 + 2A_1A_3) \\ 0 \\ \frac{L^2}{\gamma}(A_3 + 2A_1A_3) \end{pmatrix}$ $+ \frac{1}{\omega^s_{k-1}} \begin{pmatrix} 1 \\ L^2\lambda \\ 1 \\ \frac{L^2}{\gamma} \end{pmatrix} + \begin{pmatrix} 0 \\ L^2((\vert\bar{a}_3\vert^2\vert v_1\vert)^M_{k-1} + 2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)^M_{k-1}) \\ 0 \\ \frac{L^2}{\gamma}\left[(\vert\bar{a}_3\vert^2\vert v_1\vert)^M_{k-1} + 2(\vert\bar{a}_1\vert\vert\bar{a}_3\vert\vert v_3\vert)^M_{k-1}\right] \end{pmatrix} + \epsilon^3_{k-1} \begin{pmatrix} 0 \\ L^2(A_3 + 2A_1A_3) \\ 0 \\ \frac{L^2}{\gamma}(A_3 + 2A_1A_3) \end{pmatrix} \Bigg]$
$\tilde{Z}_2^k$	$\Bigg[\begin{pmatrix} 0 \\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)^M_{k+1} + 2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)^M_{k+1} + 2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)^M_{k+1}) \\ 0 \\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)^M_{k+1} + 2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)^M_{k+1} + 2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)^M_{k+1}\right] \end{pmatrix} + 2\epsilon^3_{k+1} \begin{pmatrix} 0 \\ L^2(4A_3 + 2A_1) \\ 0 \\ \frac{L^2}{\gamma}(4A_3 + 2A_1) \end{pmatrix}$ $+ \begin{pmatrix} 0 \\ L^2(2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)^M_{k-1} + 2(\vert\bar{a}_3\vert w_1\vert\vert v_3\vert)^M_{k-1} + 2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)^M_{k-1}) \\ 0 \\ \frac{L^2}{\gamma}\left[2(\vert\bar{a}_3\vert\vert w_3\vert\vert v_1\vert)^M_{k-1} + 2(\vert\bar{a}_3\vert\vert w_1\vert\vert v_3\vert)^M_{k-1} + 2(\vert\bar{a}_1\vert\vert w_3\vert\vert v_1\vert)^M_{k-1}\right] \end{pmatrix} + 2\epsilon^3_{k-1} \begin{pmatrix} 0 \\ L^2(4A_3 + 2A_1) \\ 0 \\ \frac{L^2}{\gamma}(4A_3 + 2A_1) \end{pmatrix} \Bigg]$
$\tilde{Z}_3^k$	$\Bigg[\begin{pmatrix} 0 \\ L^2((\vert w_3\vert^2\vert v_3\vert)^M_{k+1} + 2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)^M_{k+1}) \\ 0 \\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)^M_{k+1} + 2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)^M_{k+1}\right] \end{pmatrix} + 9\epsilon^3_{k+1} \begin{pmatrix} 0 \\ L^2 \\ 0 \\ \frac{L^2}{\gamma} \end{pmatrix} + \begin{pmatrix} 0 \\ L^2((\vert w_3\vert^2\vert v_3\vert)^M_{k-1} + 2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)^M_{k-1}) \\ 0 \\ \frac{L^2}{\gamma}\left[(\vert w_3\vert^2\vert v_3\vert)^M_{k-1} + 2(\vert w_1\vert\vert w_3\vert\vert v_3\vert)^M_{k}\right] \end{pmatrix} + 9\epsilon^3_{k-1} \begin{pmatrix} 0 \\ L^2 \\ 0 \\ \frac{L^2}{\gamma} \end{pmatrix} \Bigg]$
	$k \geq M$
$\tilde{Z}_1^M$	$(1 + (\frac{M}{M-1})^s) \Bigg[\begin{pmatrix} 1 \\ L^2\lambda \\ 1 \\ \frac{L^2}{\gamma} \end{pmatrix} + \begin{pmatrix} 0 \\ L^2(\Sigma_{33}^{M-1} + 2\Sigma_{13}^{M-1}) \\ 0 \\ \frac{L^2}{\gamma}(\Sigma_{33}^{M-1} + 2\Sigma_{13}^{M-1}) \end{pmatrix} \Bigg]$
$\tilde{Z}_2^M$	$(1 + (\frac{M}{M-1})^s)\alpha^3_{M-1} \begin{pmatrix} 0 \\ L^2(4A_3 + 2A_1) \\ 0 \\ \frac{L^2}{\gamma}(4A_3 + 2A_1) \end{pmatrix}$
$\tilde{Z}_3^M$	$(1 + (\frac{M}{M-1})^s)3\alpha^3_{M-1} \begin{pmatrix} 0 \\ L^2 \\ 0 \\ \frac{L^2}{\gamma} \end{pmatrix}$

Table 4.3: Formulas for $\tilde{Z}_l^k$, $k = m, \ldots, M$

Note that as $|k_i| < m-1$ for $i = 1,2,3$

$$(a_1 * a_3^2)_k^{(m)} = 0 \text{ whenever } k > 3(m-1). \tag{4.11}$$

Assume that a numerical approximation $\bar{x}_F$ with $f^{(m)}(\bar{x}_F) \approx 0$ is given and define $\bar{x} = (\bar{x}_F, 0_\infty)$. First we notice that by (4.11), setting $M = 3m-1$ suffices to fulfill assumption **A.1** from Section 3.2.2. This can be seen by realizing that for $k \geq m+1$ we have that

$$\begin{aligned}(T\bar{x} - \bar{x})_k &= -(Af(\bar{x})) = \\ &= \frac{-L^2}{2k}\left(\begin{pmatrix} 0 \\ (\bar{a}_1 * \bar{a}_3^2)_{k+1}^{(m)} \\ 0 \\ -\frac{1}{\gamma}(\bar{a}_1 * \bar{a}_3^2)_{k+1}^{(m)} \end{pmatrix} - \begin{pmatrix} 0 \\ (\bar{a}_1 * \bar{a}_3^2)_{k-1}^{(m)} \\ 0 \\ -\frac{1}{\gamma}(\bar{a}_1 * \bar{a}_3^2)_{k-1}^{(m)} \end{pmatrix}\right).\end{aligned} \tag{4.12}$$

Recalling (4.11) we hereby see that $(T\bar{x} - \bar{x})_k = 0$ for $k-1 \geq 3(m-1)$, hence for $k \geq 3m-1$. Thus we set $M = 3m-1$ and define $\bar{M} = M-1 = 3m-2$. Our first goal is to compute bounds $Y_{-1}, \ldots, Y_{\bar{M}}$ such that

$$|(T\bar{x} - \bar{x})_k| \leq Y_k$$

for $k = -1, \ldots, \bar{M}$. We therefore define

$$\bar{x}_{\bar{M}} = (\bar{\theta}, \bar{a}_0, \ldots, \bar{a}_{m-1}, \underbrace{0_4, \ldots 0_4}_{2m-2 \text{ times}})$$

and compute $y = f(\bar{x}_{\bar{M}}, 0_\infty) = (f^{(\bar{M})}(\bar{x}_{\bar{M}}), 0_\infty)$. Then we set Y_k as

$$Y_k = \begin{cases} (|A_m||y_F|)_k & k = -1, \ldots m-1 \\ \frac{|y_k|}{2k} & k = m, \ldots, \bar{M} \end{cases}. \tag{4.13}$$

We now compute polynomials $Z_k(r) = \sum_{l=1}^3 Z_l^k r^l \in \mathbb{R}^4$ $(k = -1, \ldots, \bar{M})$ such that

$$\sup_{\xi_1, \xi_2 \in B_r(0)} |(DT(\bar{x} + \xi_1)\xi_2)_k| \preceq Z_k(r)$$

for all $k = -1, \ldots, \bar{M}$. In order to do so we use the splitting given in (3.69). Let

$$\xi_1 = r(\theta, w) \text{ and } \xi_2 = r(\phi, v) \tag{4.14}$$

with$(\theta, w), (\phi, v) \in B_1(0)$.

Our first step is to compute polynomials $z_k(r) = \sum_{l=1}^{3} z_l^k r^l$ such that

$$Df_k(\bar{x}+\xi_1)\xi_2 - (A^\dagger\xi_2)_k = z_k(r) \tag{4.15}$$

for $k \geq -1$. The componentwise estimation of $|z_k(r)|$ as major step to obtain $Z_k(r)$ is postponed to a separate consideration. We note that we have to distinguish the cases $k=-1$, $k=0$, $1 \leq k \leq m-2$, $k=m-1$, $m \leq k$.

Derivation of $z_k(r)$ Let us start with the cases $-1 \leq k \leq m-1$. For these components we realize that

$$Df_k(\bar{x}+\xi_1)\xi_2 - A^\dagger\xi_2 = Df_k(\bar{x}+\xi_1)\xi_2 - Df_k^{(m)}(\bar{x}_F)\xi_{2_F}.$$

First we consider $k=-1$ in (4.15) and compute for $x=(\theta_x,a), y=(\theta_y,b) \in X^s$ and $z_F=(\theta_z,d_F) \in \mathbb{R}^{4m+2}$:

$$\begin{aligned}
D\eta(x)y - D\eta^{(m)}(z_F)y_F &= \frac{d}{dt}\eta(x+ty)|_{t=0} - \frac{d}{dt}\eta^{(m)}(z_F+ty_F)|_{t=0} \\
= \frac{d}{dt}&\begin{pmatrix} ((a+tb)_2)_0 + \sum_{j=1}^{\infty}((a+tb)_2)_j \\ ((a+tb)_4)_0 + \sum_{j=1}^{\infty}((a+tb)_4)_j \end{pmatrix}|_{t=0} \\
- \frac{d}{dt}&\begin{pmatrix} ((d_F+tb_F)_2)_0 + \sum_{j=1}^{m-1}((d_F+tb_F)_2)_j \\ ((d_F+tb_F)_4)_0 + \sum_{j=1}^{m-1}((d_F+tb_F)_4)_j \end{pmatrix}|_{t=0} = \begin{pmatrix} \sum_{j=m}^{\infty}(b_2)_j \\ \sum_{j=m}^{\infty}(b_4)_j \end{pmatrix}.
\end{aligned}$$

Setting $x=\bar{x}+\xi_1$, $z_F=\bar{x}_F$ and $y=\xi_2$ we obtain, recalling (4.14)

$$z_1^{-1} = \begin{pmatrix} \sum_{j=m}^{\infty}(v_2)_k \\ \sum_{j=m}^{\infty}(v_4)_k \end{pmatrix}. \tag{4.16}$$

In addition as $z_l^{-1}=0$ for $l=2,3$, we note that

$$z_{-1}(r) = rz_1^{-1}. \tag{4.17}$$

We remark that in order to compute $Z_{-1}(r)$, that is among others to compute a componentwise estimate on $|z_1(r)|$, we will crucially use that $\xi_2 \in B_1(0) \subset X^s$. This fact gives us information about the decay behavior of the sequences $v_{2,4}$ and enables us to estimate the infinite tail series in a standard way.

Let us continue with computing for $x=(\theta_x,a), y=(\theta_y,b) \in X^s$ and $z_F=(\theta_z,d_F) \in \mathbb{R}^{4m+2}$

$$Df_0(x)y - Df_0^{(m)}(z_F)y_F = \frac{d}{dt}f_0(x+ty)|_{t=0} - \frac{d}{dt}f_0^{(m)}(d_F+tb_F)|_{t=0}. \tag{4.18}$$

Keep in mind that we will later want to apply these calculations with $x = \bar{x} + \xi_1$, $z_F = \bar{x}_F$ and $y = \xi_2$ with $\xi_{1,2}$ given by (4.14). In this context it is essential that we compute first

$$Dc_k(a)b = \frac{d}{dt}c_k(a + tb)|_{t=0} \quad \text{and} \quad Dc_k^{(m)}(d_F)b_F = \frac{d}{dt}c_k^{(m)}(d_F + tb_F)|_{t=0}$$

for certain $k \geq 0$. In order to achieve this we have to compute

$$\begin{aligned}
&\frac{d}{dt}((a_1 + tb_1) * (a_3 + tb_3)^2)_k|_{t=0} \\
&= \sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} \frac{d}{dt}(a_1 + tb_1)_{|k_1|}(a_3 + tb_3)_{|k_2|}(a_3 + tb_3)_{|k_3|}|_{t=0} \\
&= \sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} \frac{d}{dt}\Big[(a_1)_{|k_1|}(a_3)_{|k_2|}(a_3)_{|k_3|} + t(b_1)_{|k_1|}(a_3)_{k_2}(a_3)_{|k_3|} + \\
&\qquad\qquad t(a_1)_{|k_1|}(b_3)_{|k_2|}(a_3)_{|k_3|} + t(a_1)_{|k_1|}(a_3)_{|k_2|}(b_3)_{|k_3|} + \\
&\qquad\qquad t^2(b_1)_{|k_1|}(b_3)_{|k_2|}(a_3)_{|k_3|} + t^2(b_1)_{|k_1|}(a_3)_{|k_2|}(b_3)_{|k_3|} + \\
&\qquad\qquad t^2(a_1)_{|k_1|}(b_3)_{|k_2|}(b_3)_{|k_3|} + t^3(b_1)_{|k_1|}(b_3)_{|k_2|}(b_3)_{|k_3|}\Big]_{t=0} \\
&= \sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} (b_1)_{|k_1|}(a_3)_{|k_2|}(a_3)_{|k_3|} + (a_1)_{|k_1|}(b_3)_{|k_2|}(a_3)_{|k_3|} + (a_1)_{|k_1|}(a_3)_{|k_2|}(b_3)_{|k_3|} \\
&= (a_3^2 * b_1)_k + 2(a_1 * a_3 * b_3)_k.
\end{aligned}$$

In the same manner we obtain that

$$\frac{d}{dt}((d_1 + tb_1) * (d_3 + tb_3)^2)_k^{(m)}|_{t=0} = (d_3^2 * b_1)_k^{(m)} + 2(d_1 * d_3 * b_3)_k^{(m)}.$$

This yields

$$Dc_k(a)b = \begin{pmatrix} (b_2)_k \\ L^2\lambda(b_1)_k \\ (b_4)_k \\ \frac{L^2}{\gamma}(b_3)_k \end{pmatrix} + \begin{pmatrix} 0 \\ L^2\left((a_3^2 * b_1)_k + 2(a_1 * a_3 * b_3)_k\right) \\ 0 \\ \frac{-L^2}{\gamma}\left((a_3^2 * b_1)_k + 2(a_1 * a_3 * b_3)_k\right) \end{pmatrix} \tag{4.19}$$

and

$$Dc_k^{(m)}(d_F)b_F = \begin{pmatrix} (b_2)_k \\ L^2\lambda(b_1)_k \\ (b_4)_k \\ \frac{L^2}{\gamma}(b_3)_k \end{pmatrix} + \begin{pmatrix} 0 \\ L^2\left((d_3^2 * b_1)_k^{(m)} + 2(d_1 * d_3 * b_3)_k^{(m)}\right) \\ 0 \\ \frac{-L^2}{\gamma}\left((d_3^2 * b_1)_k^{(m)} + 2(d_1 * d_3 * b_3)_k^{(m)}\right) \end{pmatrix} \tag{4.20}$$

for $k = 1, \dots, m-1$. We now go on to consider

$$\begin{aligned}Df_0(x)y - Df_0^{(m)}(z_F)y_F = &\underbrace{\frac{d}{dt}(a_0 + tb_0 - d_0 - tb_0)|_{t=0}}_{=0} + \\ &(Dc_0(a)b - Dc_0^{(m)}(d_F)b_F) + \frac{1}{2}(Dc_1(a)b - Dc_1^{(m)}(d_F)b_F) \\ &- 2\sum_{j=2}^{m-1} \frac{1}{j^2-1}(Dc_j(a)b - Dc_j^{(m)}(d_F)b_F) - 2\sum_{j=m}^{\infty} \frac{1}{j^2-1} Dc_j(a)b \\ &- DP(\theta_x)\theta_y.\end{aligned} \tag{4.21}$$

We are now ready to set $x = \bar{x} + r(\theta, w)$, $z_F = \bar{x}_F$ and $y = r(\phi, v)$ with $(\theta, w), (\phi, v) \in B_1(0) \subset X^s$ in (4.21). In order to get a polynomial expansion

$$Df_0(\bar{x} + r(\theta, w))r(\phi, v) - Df_0^{(m)}(\bar{x}_F)r(\phi, v_F) = \sum_{l=1}^{3} z_l^0 r^l \tag{4.22}$$

we hence need to consider

$$Dc_j(\bar{a} + rw)rv - Dc_j^{(m)}(\bar{a}_F)rv_F$$

for $0 \leq j \leq m-1$. Recalling (4.19) and (4.20) let us first realize that

$$\begin{aligned}(\bar{a}_3 + rw_3)^2 * rv_1 =& r\bar{a}_3^2 * v_1 + 2r^2\bar{a}_3 * w_3 * v_1 + r^3 w_3^2 * v_1 \\ (\bar{a}_1 + rw_1) * (\bar{a}_3 + rw_3) * rv_3 =& r\bar{a}_1 * \bar{a}_3 * v_3 \\ &+ r^2(\bar{a}_3 * w_1 * v_3 + \bar{a}_1 * w_3 * v_1) \\ &+ r^3 w_1 * w_3 * v_3.\end{aligned}$$

Thus we obtain

$$Dc_j(\bar{a} + rw)rv = \sum_{l=1}^{3} \beta_l^j r^j$$

with β_l^j (l=1,2,3) defined for every $j \geq 0$ by

$$\beta_1^j = \begin{pmatrix} (v_2)_j \\ L^2\lambda(v_1)_j \\ (v_4)_j \\ \frac{L^2}{\gamma}(v_3)_j \end{pmatrix} + \begin{pmatrix} 0 \\ L^2\left((\bar{a}_3^2 * v_1)_j + 2(\bar{a}_1 * \bar{a}_3 * v_3)_j\right) \\ 0 \\ \frac{-L^2}{\gamma}\left((\bar{a}_3^2 * v_1)_j + 2(\bar{a}_1 * \bar{a}_3 * v_3)_j\right) \end{pmatrix}$$

$$\beta_2^j = \begin{pmatrix} 0 \\ L^2\left(2(\bar{a}_3 * w_3 * v_1)_j + 2(\bar{a}_3 * w_1 * v_3)_k + 2(\bar{a}_1 * w_3 * v_1)_j\right) \\ 0 \\ \frac{-L^2}{\gamma}\left(2(\bar{a}_3 * w_3 * v_1)_j + 2(\bar{a}_3 * w_1 * v_3)_k + 2(\bar{a}_1 * w_3 * v_1)_j\right) \end{pmatrix} \tag{4.23}$$

$$\beta_3^j = \begin{pmatrix} 0 \\ L^2\left((w_3^2 * v_3)_j + 2(w_1 * w_3 * v_3)_j\right) \\ 0 \\ \frac{-L^2}{\gamma}\left((w_3^2 * v_3)_j + 2(w_1 * w_3 * v_3)_j\right) \end{pmatrix}.$$

Now assume $0 \leq j \leq m-1$. Then

$$Dc_j^{(m)}(\bar{a}_F)rv_F = r\left(\begin{pmatrix} (v_2)_j \\ L^2\lambda(v_1)_j \\ (v_4)_j \\ \frac{L^2}{\gamma}(v_3)_j \end{pmatrix} + \begin{pmatrix} 0 \\ L^2\left((\bar{a}_3^2 * v_1)_j^{(m)} + 2(\bar{a}_1 * \bar{a}_3 * v_3)_j^{(m)}\right) \\ 0 \\ \frac{-L^2}{\gamma}\left((\bar{a}_3^2 * v_1)_j^{(m)} + 2(\bar{a}_1 * \bar{a}_3 * v_3)_j^{(m)}\right) \end{pmatrix}\right). \tag{4.24}$$

Remark 4.1.4 ***Difference of infinite and finite cubic convolution sums***
Let sequences $a_{1,3} \in \Omega^s$ and $v_{1,3} \in \Omega^s$ be given. Then we consider the difference $(a_3^2 v_1)_k - (a_3^2 v_1)_k^{(m)}$ of the infinite and finite convolution sums:

$$\sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} - \sum_{\substack{k_1+k_2+k_3=k \\ |k_i|<m}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} =$$

$$\sum_{\substack{-m+1\leq k_1 \leq m-1 \\ |k_i|\geq m \ \ (i=2,3) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} + \sum_{\substack{|k_1|\geq m \\ k_i \in \mathbb{Z} \ \ (i=2,3) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} +$$

$$\sum_{\substack{-m+1\leq k_2 \leq m-1 \\ |k_i|\geq m \ \ (i=1,3) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} + \sum_{\substack{|k_2|\geq m \\ k_i \in \mathbb{Z} \ \ (i=1,3) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} +$$

$$\sum_{\substack{-m+1\leq k_3 \leq m-1 \\ |k_i|\geq m \ \ (i=1,2) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} + \sum_{\substack{|k_3|\geq m \\ k_i \in \mathbb{Z} \ \ (i=1,2) \\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|}.$$

We make the following important observation. If $(a_3)_k = 0$ for $k \geq m$ then we *obtain that*

$$\sum_{\substack{k_1+k_2+k_3=k\\ k_i\in\mathbb{Z}}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{k_3} - \sum_{\substack{k_1+k_2+k_3=k\\ |k_i|<m}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} =$$

$$= \sum_{\substack{|k_3|\geq m\\ k_i\in\mathbb{Z}\ \ (i=1,2)\\ k_1+k_2+k_3=k}} (a_3)_{|k_1|}(a_3)_{|k_2|}(v_1)_{|k_3|} = ((a_3)^2 v_1^I)_k,$$

where we define for a sequence $a \in \Omega_s$

$$a_k^I = \begin{cases} 0 & k \leq m-1 \\ a_k & k \geq m \end{cases}. \tag{4.25}$$

Note that we encounter exactly the case $(a_3)_k = 0$ for $k \geq m$ when we compute

$$Dc_j(\bar{a} + rw)rv - Dc_j^{(m)}(\bar{a}_F)rv_F.$$

In an analogue fashion we have in the case $(a_1)_k = (a_3)_k = 0$ for $k \geq m$

$$\sum_{\substack{k_1+k_2+k_3=k\\ k_i\in\mathbb{Z}}} (a_1)_{|k_1|}(a_3)_{|k_2|}(v_3)_{k_3} - \sum_{\substack{k_1+k_2+k_3=k\\ |k_i|<m}} (a_1)_{|k_1|}(a_3)_{|k_2|}(v_3)_{|k_3|} =$$

$$= \sum_{\substack{|k_3|\geq m\\ k_i\in\mathbb{Z}\ \ (i=1,2)\\ k_1+k_2+k_3=k}} (a_1)_{|k_1|}(a_3)_{|k_2|}(v_3)_{|k_3|} = (a_1 a_3 v_3^I).$$

Using (4.23), (4.24) and Remark 4.1.4, we obtain for $0 \leq j \leq m-1$

$$Dc_j(\bar{a} + rw)rv - Dc_j^{(m)}(\bar{a}_F)rv_F = \sum_{l=1}^{3} \kappa_l^j r^l \tag{4.26}$$

with κ_l^j defined by

$$\kappa_1^j = \begin{pmatrix} 0 \\ L^2\left((\bar{a}_3^2 * v_1^I)_j + 2(\bar{a}_1 * \bar{a}_3 * v_3^I)_j\right) \\ 0 \\ \frac{-L^2}{\gamma}\left[(\bar{a}_3^2 * v_1^I)_j + 2(\bar{a}_1 * \bar{a}_3 * v_3^I)_j\right] \end{pmatrix} \tag{4.27}$$

$$\kappa_{2,3}^j = \beta_{2,3}^j.$$

Using (4.26) we can specify the polynomial expansion given in (4.22) by

$$\begin{aligned} &Df_0(\bar{x} + r(\theta, w))r(\phi, v) - Df_0^{(m)}(\bar{x}_F)r(\phi, v_F) = \\ &\left[\sum_{l=1}^{3} \kappa_l^0 r^l + \frac{1}{2}\sum_{l=1}^{3} \kappa_l^1 r^l \right. \\ &\left. -2\sum_{j=2}^{m-1} \frac{1}{j^2-1}\left(\sum_{l=1}^{3} \kappa_l^j r^l\right) - 2\sum_{j=m}^{\infty} \frac{1}{j^2-1}\left(\sum_{l=1}^{3} \beta_l^j r^l\right)\right] = \sum_{j=1}^{3} z_j^0 r^j \end{aligned} \tag{4.28}$$

with

$$\begin{aligned} z_1^0 &= \left[\kappa_1^0 + \frac{1}{2}\kappa_1^1 - 2\sum_{j=2}^{m-1} \frac{\kappa_1^j}{j^2-1} - 2\sum_{j=m}^{\infty} \frac{\beta_1^j}{j^2-1}\right] - DP(\bar{\theta} + r\theta)\phi \\ z_l^0 &= \left[\beta_l^0 + \frac{1}{2}\beta_l^1 - 2\sum_{j=2}^{\infty} \frac{\beta_l^j}{j^2-1}\right] \quad (l = 2, 3) \end{aligned} \tag{4.29}$$

where $0 \leq r < r_{appr}$ with an apriori bound r_{appr} on r.

We now go ahead to consider the cases $1 \leq k \leq m-1$. We obtain for $x = (\theta_a, a)$ and $y = (\theta_b, b)$

$$Df_k(x)y = 2kb_k + (Dc_{k+1}(a)b - Dc_{k-1}(a)b).$$

This yields for $k = 1, \ldots, m-1$ and $\xi_{1,2}$ specified in (4.14)

$$\begin{aligned} & Df_k(\bar{x} + \xi_1)\xi_2 - A^\dagger\xi_2 = Df_k(\bar{x} + \xi_1)\xi_2 - Df_k^{(m)}(\bar{x}_F)\xi_2 = \\ & = L\left[(Dc_{k+1}(\bar{a} + rw)rv - Dc_{k+1}^{(m)}(\bar{a}_F)rw_F) - (Dc_{k-1}(\bar{a} + rw)rv - Dc_{k-1}^{(m)}(\bar{a}_F)rv_F)\right] \\ & = \sum_{l=1}^{3}(\kappa_l^{k+1} - \kappa_l^{k-1})r^l = \sum_{l=1}^{3} z_l^k r^l, \end{aligned}$$

where we define for $j = 1, 2, 3$

$$z_l^k = \begin{cases} (\kappa_l^{k+1} - \kappa_l^{k-1}) & l = 1 \\ (\beta_l^{k+1} - \beta_l^{k-1}) & l = 2, 3 \end{cases} \tag{4.30}$$

with κ_1^k and $\beta_{2,3}^k$ given as in (4.27) and (4.23).

Finally for $k \geq m$ we have that

$$\begin{aligned} & Df_k(\bar{x} + \xi_1)\xi_2 - A^\dagger\xi_2 = Df_k(\bar{x} + \xi_1)\xi_2 - 2krv_k = \\ & = [Dc_{k+1}(\bar{a} + rw)rv - Dc_{k-1}(\bar{a} + rw)rv] = \\ & = \sum_{l=1}^{3}(\beta_l^{k+1} - \beta_l^{k-1})r^l = \sum_{l=1}^{3} z_l^k r^l, \end{aligned}$$

where we define for $j = 1, 2, 3$

$$z_l^k = (\beta_l^{k+1} - \beta_l^{k-1}) \tag{4.31}$$

with β_l^k given as in (4.23).

Our next goal is to compute bounds $\tilde{Z}_k(r) = \sum_{l=1}^{3} \tilde{Z}_l^k r^l$ such that

$$|z_k(r)| \preceq \tilde{Z}_k(r)$$

for $k = -1, \dots, \bar{M}$, where we recall the notation $\preceq$ given in Definition 3.1.1. Similar notation will be applied for strict inequalities. We note that the results can be found in Tables 4.2 and 4.3.

Derivation of $\tilde{Z}_k(r)$ and $Z_k(r)$ Let us start with $k = -1$. Recalling (4.16) and in addition that $w, v \in B_1(0) \subset \Omega_s$ implies in particular that $|(v_i)_k| \leq \frac{1}{\omega_k^s}$ for $i = 2, 4$. Hence by applying integral estimates in each component we obtain

$$|z_1^{-1}| \preceq \begin{pmatrix} \sum_{j=m}^{\infty} \frac{1}{j^s} \\ \sum_{j=m}^{\infty} \frac{1}{j^s} \end{pmatrix} \preceq \begin{pmatrix} \int_{m-1}^{\infty} \frac{1}{j^s} dj \\ \int_{m-1}^{\infty} \frac{1}{j^s} dj \end{pmatrix} = (s-1) \begin{pmatrix} \frac{1}{(m-1)^{s-1}} \\ \frac{1}{(m-1)^{s-1}} \end{pmatrix}$$

We can therefore define

$$\tilde{Z}_1^{-1} = (s-1) \begin{pmatrix} \frac{1}{(m-1)^{s-1}} \\ \frac{1}{(m-1)^{s-1}} \end{pmatrix} \qquad \tilde{Z}_l^{-1} = 0 \ \ (l = 2, 3) \tag{4.32}$$

The decay information on the sequences w, v will continue to stay crucial for our estimates. To define $\tilde{Z}_k$ for $k = 0, \dots, \bar{M}$ we first derive bounds B_l^k $(l = 1, 2, 3)$, K_1^k and $B_l^{\bar{M}}$ $(l = 1, 2, 3)$ such that:

$$\begin{aligned} |\beta_l^k| &\preceq B_l^k \quad l = 1, 2, 3, \quad k = 0, \dots, M \\ |\kappa_1^k| &\preceq K_1^k \qquad 0 \leq k \leq m-1 \\ |\beta_l^k| &\preceq \frac{B_l^{\bar{M}}}{\omega_k^s} \quad l = 1, 2, 3 \quad k \geq \bar{M}. \end{aligned} \tag{4.33}$$

To achieve this we will need to bound general cubic convolution sums. Realize that every convolution sum can be split in the following way. Let $l \in \{1, 2, 3\}$, $M > 0$ and $a, b, c \in \Omega^s$ sequences. Then

$$\sum_{\substack{k_1+k_2+k_3=k \\ k_i \in \mathbb{Z}}} a_{|k_1|} b_{|k_2|} c_{|k_3|} = \sum_{\substack{k_1+k_2+k_3=k \\ |k_i| < M \ (i=1,\dots,l)}} a_{|k_1|} b_{|k_2|} c_{|k_3|} + \sum_{\substack{k_1+k_2+k_3=k \\ \max_{i=1,\dots,l} |k_i| \geq M}} a_{|k_1|} b_{|k_2|} c_{|k_3|}.$$

The benefit of this operation is that in our case the sequences a, b will be finite and thus by choosing $l = 3$ the first sum is finite and can be computed via FFT and the second sum can be estimated by the following results obtained in [25]. The next Lemma is a special case of Lemma **A.4** in [25].

Lemma 4.1.1 *Let $s \geq 2$ be a decay rate and $M_1 \geq M_2 \geq 6$. Fix $l \in \{1,2,3\}$. Then for $0 \leq k \leq M_2 - 1$ there are computable numbers $\epsilon_k^3 = \epsilon_k^3(M_1, M_2)$ such that*

$$\left| \sum_{\substack{k_1+k_2+k_3=k \\ \max_{i=1,\dots,l} |k_i| \geq M_2}} a_{|k_1|} b_{|k_2|} c_{|k_3|} \right| \leq l(ABC)\epsilon_k^3,$$

where $A = \|a\|_s$, $B = \|b\|_s$ and $C = \|c\|_s$.

Proof 4.1.1 *See [25].*

We set $A_{1,3} = \max\limits_{i=0,\dots,m-1} \{|(\bar{a}_{1,3})_i| \omega_k^s\}$ and proceed to compute B_l^k for $k = 0, \dots, M$. Recalling (4.23) we aim to use Lemma 4.1.1 with $M_1 = M + 1$ and $M_2 = M = \bar{M} + 1$. Hence we obtain:

$$\begin{aligned}
|\beta_1^k| &\preceq \frac{1}{\omega_k^s} \begin{pmatrix} 1 \\ L^2\lambda \\ 1 \\ \frac{L^2}{\gamma} \end{pmatrix} + \begin{pmatrix} 0 \\ L^2 \left((|\bar{a}_3|^2 * |v_1|)_k^{(M)} + 2(|\bar{a}_1| * |\bar{a}_3| * |v_3|)_k^{(M)} \right) \\ 0 \\ \frac{L^2}{\gamma} \left[(|\bar{a}_3|^2 * |v_1|)_k^{(M)} + 2(|\bar{a}_1| * |\bar{a}_3| * |v_3|)_k^{(M)} \right] \end{pmatrix} + \\
&\quad \epsilon_k^3 \begin{pmatrix} 0 \\ L^2 \left(A_3^2 + 2A_1A_3 \right) \\ 0 \\ \frac{L^2}{\gamma} (A_3^2 + 2A_1A_3) \end{pmatrix} \stackrel{\text{def}}{=} B_1^k \\
|\beta_2^k| &\preceq \begin{pmatrix} 0 \\ L^2 \left(2(|\bar{a}_3| * |w_3| * |v_1|)_k^{(M)} + 2(|\bar{a}_3| * |w_1| * |v_3|)_k^{(M)} + 2(|\bar{a}_1| * |w_3| * |v_1|)_k^{(M)} \right) \\ 0 \\ \frac{L^2}{\gamma} \left[2(|\bar{a}_3| * |w_3| * |v_1|)_k^{(M)} + 2(|\bar{a}_3| * |w_1| * |v_3|)_k^{(M)} + 2(|\bar{a}_1| * |w_3| * |v_1|)_k^{(M)} \right] \end{pmatrix} + \\
&\quad 2\epsilon_k^3 \begin{pmatrix} 0 \\ L^2 \left(4A_3 + 2A_1 \right) \\ 0 \\ \frac{L^2}{\gamma} (4A_3 + 2A_1) \end{pmatrix} \stackrel{\text{def}}{=} B_2^k \\
|\beta_3^k| &\preceq \begin{pmatrix} 0 \\ L^2 \left((|w_3|^2 * |v_3|)_k^{(M)} + 2(|w_1| * |w_3| * |v_3|)_k^{(M)} \right) \\ 0 \\ \frac{L^2}{\gamma} \left[(|w_3|^2 * |v_3|)_k^{(M)} + 2(|w_1| * |w_3| * |v_3|)_k^{(M)} \right] \end{pmatrix} + 9\epsilon_k^3 \begin{pmatrix} 1 \\ L^2 \\ 1 \\ L^2 \end{pmatrix} \stackrel{\text{def}}{=} B_3^k
\end{aligned} \tag{4.34}$$

Recalling (4.27) we go on by computing K_1^k for $k = 0, \dots, m-1$

$$\begin{aligned} |\kappa_1^k| \preceq & \begin{pmatrix} 0 \\ L^2 \left((|\bar{a}_3|^2 * |v_1|^I)_k^{(M)} + 2(|\bar{a}_1| * |\bar{a}_3| * |v_3^I|)_k^{(M)} \right) \\ 0 \\ \frac{L^2}{\gamma} \left[(|\bar{a}_3|^2 * |v_1|^I)_k^{(M)} + 2(|\bar{a}_1| * |\bar{a}_3| * |v_3^I|)_k^{(M)} \right] \end{pmatrix} + \\ & \epsilon_k^3 \begin{pmatrix} 0 \\ L^2 \left(A_3 + 2A_1A_3 \right) \\ 0 \\ \frac{L^2}{\gamma} (A_3 + 2A_1A_3) \end{pmatrix} \stackrel{\text{def}}{=} K_1^k. \end{aligned} \tag{4.35}$$

Concerning the uniform bounds $B_l^{\bar{M}}$ for $l = 1, 2, 3$ specified in (4.33) we use Lemma 2.2.5 with $n = 3$ and $M_1 = \bar{M}$. In particular we wish to remind the reader of equation (2.43). Thus we obtain

$$\begin{aligned} |\beta_1^k| &\preceq \frac{1}{\omega_k^s} \begin{pmatrix} 1 \\ L^2\lambda \\ 1 \\ \frac{L^2}{\gamma} \end{pmatrix} + \frac{\alpha_{\bar{M}}^3}{\omega_k^s} \begin{pmatrix} 0 \\ L^2(A_3^2 + 2A_1A_3) \\ 0 \\ \frac{L^2}{\gamma}(A_3^2 + 2A_1A_3) \end{pmatrix} \stackrel{\text{def}}{=} \frac{B_1^{\bar{M}}}{\omega_k^s} \\ |\beta_2^k| &\preceq \frac{\alpha_{\bar{M}}^3}{\omega_k^s} \begin{pmatrix} 0 \\ L^2 \left(4A_3 + 2A_1 \right) \\ 0 \\ \frac{L^2}{\gamma}(4A_3 + 2A_1) \end{pmatrix} \stackrel{\text{def}}{=} \frac{B_2^{\bar{M}}}{\omega_k^s} \\ |\beta_3^k| &\preceq \frac{3\alpha_{\bar{M}}^3}{\omega_k^s} \begin{pmatrix} 0 \\ L^2 \\ 0 \\ L^2 \end{pmatrix} \stackrel{\text{def}}{=} \frac{B_3^{\bar{M}}}{\omega_k^s}. \end{aligned} \tag{4.36}$$

Let us now estimate $|z_l^0|$ for $l = 1, 2, 3$. To this end we first assume that we have a bound $\Lambda \in \mathbb{R}^4$ such that

$$|DP^s(\bar{\theta} + r\theta)\phi| \leq \Lambda$$

for all r with $0 < r < r_{apriori}$, where $r_{apriori}$ is an apriori bound on r. The details of achieving this via an argument based on the Mean Value Theorem are explained in [60]. The reasoning in 3.1.2 is very similar in spirit. Recalling (4.28) we have

$$\begin{aligned}
|z_1^0| &\preceq L\left[K_1^0+\frac{1}{2}K_1^1+\sum_{j=2}^{m-1}K_1^j\frac{1}{k^2-1}+\sum_{j=m}^{\infty}B_1^{\bar{M}}\underbrace{\frac{\alpha_{\bar{M}}^3}{j^s(j^2-1)}}_{\leq\frac{\alpha_{\bar{M}}^3}{(\bar{M}^2-1)j^s}}\right]+\Lambda\\
&\preceq L\Bigg[K_1^0+\frac{1}{2}K_1^1+\sum_{j=2}^{m-1}K_1^j\frac{1}{j^2-1}\\
&\quad+\sum_{j=2}^{m-1}B_1^j\frac{1}{j^2-1}+B_1^{\bar{M}}\frac{\alpha_{\bar{M}}^3}{(\bar{M}^2-1)(s-1)(\bar{M}-1)^{(s-1)}}\Bigg]+\Lambda\overset{\text{def}}{=}\tilde{Z}_1^0\\
|z_l^0| &\preceq L\left[B_l^0+\frac{1}{2}B_l^1+\sum_{j=2}^{\bar{M}}B_l^k\frac{1}{j^2-1}+B_l^{\bar{M}}\frac{\alpha_{\bar{M}}^3}{(\bar{M}^2-1)(s-1)(\bar{M}-1)^{(s-1)}}\right]\overset{\text{def}}{=}\tilde{Z}_l^0.
\end{aligned} \tag{4.37}$$

Then for $k=1,\dots,m-1$ we have for $l=1,2,3$ that

$$|z_l^k|\preceq(K_l^{k+1}+K_l^{k-1})\overset{\text{def}}{=}\tilde{Z}_l^k. \tag{4.38}$$

And for $k=m\dots,\bar{M}$ and $l=1,2,3$ we get

$$|z_l^k|\preceq L(B_l^{k+1}+B_l^{k-1})\overset{\text{def}}{=}\tilde{Z}_l^k. \tag{4.39}$$

Taking (4.32), (4.37), (4.38) and (4.39) together we obtain polynomial expansions

$$|Df_k(\bar{x}+rw_1)-A^{\dagger}rw_2|\preceq\tilde{Z}_k(r)=\sum_{l=1}^{3}\tilde{Z}_l^k r^l \tag{4.40}$$

for all $k=-1,\dots,\bar{M}$.

By definition of A and $A^{\dagger}$ there is a δ such that for all $k\geq-1$

$$|((\mathrm{I}-AA^{\dagger})\xi_2)_k|\preceq r\delta\in\mathbb{R}^4$$

where the inequality is to understood componentwise. We now define for $l=1,2,3$ vectors $V_l=(\tilde{Z}_l^{-1},\tilde{Z}_l^0,\dots,\tilde{Z}_l^{m-1})\in\mathbb{R}^{4m+2}$ to obtain for $k=-1,\dots,m-1$

$$\begin{aligned}
Z_1^k&=(|A_m|V_1)_k+\delta\\
Z_l^k&=(|A_m|V_j)_k\quad j=2,3
\end{aligned} \tag{4.41}$$

and for $k=m,\dots,\bar{M}$

$$Z_l^k=\frac{1}{2k}\tilde{Z}_l^k\quad k=m,\dots,\bar{M} \tag{4.42}$$

and hereby have that for all $k = -1, \dots, \bar{M}$

$$|(DT(\bar{x} + \xi_1)\xi_2)_k| \preceq \sum_{l=1}^{3} Z_l^k r^l.$$

As the right hand side is independent of $\xi_{1,2}$ we can take the supremum over all $\xi_{1,2} \in B_r(0) \subset X^s$ and obtain

$$\sup_{\xi_1,\xi_2 \in B_r(0)} |(DT(\bar{x} + \xi_1)\xi_2)_k| \preceq \sum_{l=1}^{3} Z_l^k r^l$$

for all $k = -1, \dots, \bar{M}$.

We now are able to combine equations (4.57), (4.68) and (4.69) to define the radii polynomials for $k = -1, \dots, \bar{M}$ specified in (3.62) by setting

$$p_k(r) = Y_k + \sum_{l=1}^{3} Z_l^k r^l - \frac{r}{\omega_k^s} \mathbf{1_4}. \tag{4.43}$$

Let us consider the tail radii polynomial from (3.63). We then seek a bound $Z_M(r)$ such that

$$\sup_{w_1,w_2 \in B_r(0)} |(DT(\bar{x} + rw_1)w_2)_k| \preceq \frac{1}{\omega_k^s} Z_M(r)$$

for $k \geq M$. The result can be found in Table 4.3. To this end let us compute bounds such that

$$|z_l^k| \preceq \frac{1}{\omega_k^s} \tilde{Z}_l^M$$

for $k \geq M$. Recall that $z_l^k = \beta_l^{k+1} - \beta_l^{k-1}$ with β_l^k specified in (4.23). For $l = 2, 3$ we aim to use (4.36). Concerning the crucial term for $l = 1$ we will follow a more refined reasoning we specify momentarily. Thus we obtain for $l = 2, 3$:

$$|z_l^k| \preceq L\left(\frac{B_l^{\bar{M}}}{\omega_{k+1}^s} + \frac{B_l^{\bar{M}}}{\omega_{k-1}^s}\right) \preceq B_l^{\bar{M}} \frac{1}{\omega_k^s}\left(1 + (1 + \frac{1}{\bar{M}})^s\right) \overset{\text{def}}{=} \frac{1}{\omega_k^s} \tilde{Z}_l^M \tag{4.44}$$

where we applied the following trick:

$$\begin{aligned} \frac{1}{(k+1)^s} + \frac{1}{(k-1)^s} &\leq \frac{1}{k^s} + \frac{1}{(k-1)^s} = \frac{1}{k^s}\left(1 + \frac{k^s}{(k-1)^s}\right) \\ &\leq \frac{1}{k^s}\left(1 + \frac{M^s}{(M-1)^s}\right) = \frac{1}{k^s}\left(1 + \frac{(\bar{M}+1)^s}{\bar{M}^s}\right) \\ &= \frac{1}{k^s}\left(1 + (1 + \frac{1}{\bar{M}})^s\right). \end{aligned}$$

We could proceed in an analogue fashion for the linear terms but we can achieve sharper bounds by applying the following reasoning. Remembering (4.30) we first need to consider for $k \geq M$ the convolution term

$$\begin{aligned}
\left|(\bar{a}_3^2 v_1)_{k+1}\right| &= \left|\sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (\bar{a}_3)_{|k_1|} (\bar{a}_3)_{|k_2|} v_1)_{k-k_1-k_2}\right| \\
&\leq \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_3|)_{|k_1|} (|\bar{a}_3|)_{|k_2|} \frac{1}{(k+1-k_1-k_2)^s} \\
&= \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_3|)_{|k_1|} (|\bar{a}_3|)_{|k_2|} \frac{k^s}{(k+1-k_1-k_2)^s} \\
&\leq \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_3|)_{|k_1|} (|\bar{a}_3|)_{|k_2|} \max(\frac{M^s}{(M+1-k_1-k_2)^s}, 1) \\
&\overset{\text{def}}{=} \frac{1}{\omega_k^s} \Sigma_{33}^{M+1}.
\end{aligned} \tag{4.45}$$

In an analogue fashion we have that

$$\begin{aligned}
|(\bar{a}_3^2 (b_2)_1)_{k-1}| &\leq \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_3|)_{|k_1|} (|\bar{a}_3|)_{|k_2|} \max(\frac{(M-2)^s}{(M-1-k_1-k_2)^s}, 1) \\
&\overset{\text{def}}{=} \frac{1}{\omega_k^s} \Sigma_{33}^{M-1} \\
|(\bar{a}_1 \bar{a}_3 (b_2)_1)_{k-1}| &\leq \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_1|)_{|k_1|} (|\bar{a}_1|)_{|k_2|} \max(\frac{M^s}{(M+1-k_1-k_2)^s}, 1) \\
&\overset{\text{def}}{=} \frac{1}{\omega_k^s} \Sigma_{13}^{M+1} \\
|(\bar{a}_1 \bar{a}_3 (b_2)_1)_{k-1}| &\leq \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{m-1} \sum_{k_2=-m+1}^{m-1} (|\bar{a}_1|)_{|k_1|} (|\bar{a}_1|)_{|k_2|} \max(\frac{(M-2)^s}{(M-1-k_1-k_2)^s}, 1) \\
&\overset{\text{def}}{=} \frac{1}{\omega_k^s} \Sigma_{13}^{M-1}.
\end{aligned} \tag{4.46}$$

In the estimates above the following elementary reasoning is crucial. Consider $\tau_k = (\frac{k}{k+1-c})$ for a constant $c \in \mathbb{R}$ and k such that τ_k is well defined. Then we have

$$\begin{aligned}
\frac{k}{k+1-c} &\geq 1 \quad c \geq 1 \\
\frac{k}{k+1-c} &\leq 1 \quad c \leq 1.
\end{aligned}$$

So in the first case $\tau_{k+1} \leq \tau_k$ whenever defined and so we have $\tau_k \leq \tau_M$. In the second case we $\tau_k \leq 1$ for all k where it is defined and we can use this to estimate the sum.

Thus inserting (4.45) and (4.46) into (4.30) we obtain the estimate

$$\begin{aligned}|z_1^k| &\preceq \frac{1}{\omega_k^s}\left(\left(1+(1+\frac{1}{\bar{M}})^s\right)\begin{pmatrix}1\\ \lambda\\ 1\\ \frac{1}{\gamma}\end{pmatrix}+\begin{pmatrix}0\\ \Sigma_{33}^{M+1}+2\Sigma_{13}^{M+1}+\Sigma_{33}^{M-1}+2\Sigma_{13}^{M-1}\\ 0\\ \frac{1}{\gamma}\left(\Sigma_{33}^{M+1}+2\Sigma_{13}^{M+1}+\Sigma_{33}^{M-1}+2\Sigma_{13}^{M-1}\right)\end{pmatrix}\right)\\ &\overset{\text{def}}{=} \frac{1}{\omega_k^s}\tilde{Z}_1^M.\end{aligned} \tag{4.47}$$

The next step is to apply A to the estimates on $|Df_k(\bar{x}+rw_1)rw_2)) - (A^\dagger rw_2)_k|$. As

$$\frac{1}{2k} \le \frac{1}{2M}$$

for $k \ge M$ this leads to setting

$$Z_l^M = \frac{1}{2M}\tilde{Z}_l^M \tag{4.48}$$

for $l = 1,2,3$. Then defining $Z_M(r) = \sum_{l=1}^3 Z_l^M r^l$ we finally have that

$$\sup_{w_1,w_2} |(DT(\bar{x}+\xi_1)\xi_2)_k| \preceq \frac{1}{\omega_k^s} Z_M(r)$$

for $\xi_{1,2}$ given in (4.14) and all $k \ge M$.

4.2 The Lorenz system

In this section we consider the well-known Lorenz equations (see e.g. [52]) given by the 3D ODE system

$$\begin{aligned}\dot{x} &= \sigma(y-x)\\ \dot{y} &= \rho x - y - xz\\ \dot{z} &= xy - \beta z,\end{aligned} \tag{4.49}$$

where β, ρ and σ are real parameters. The system has the equilibria

$$q_1 = (0,0,0) \qquad q_{2,3} = (\pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho-1),$$

where $q_{2,3}$ exist for $\beta(\rho-1) \ge 0$. The classical studies conducted on the system consist in fixing the parameters $\beta = \frac{8}{3}$ and $\sigma = 10$ and consider ρ as a bifurcation parameter. For an extensive overview of analytical and numerical studies we refer the reader to [52].

We begin our treatment of the Lorenz system with the discretization-free approach. We fix the parameter σ at the non-classical value $\sigma = -2.2$ and let $\beta = \frac{8}{3}$. In order to study the connecting dynamics in the vicinity of the pitchfork bifurcation at $\rho = 1$ we vary ρ starting from $\rho = 1.33$ and investigate connecting orbits from the origin to the secondary equilibrium q_2. Our focus of attention is twofold. First we investigate the validation procedure for $\rho = 1.33$ more closely. In particular we elaborate on the usage of the conjugacy equation (2.9) for the integration of the flow on the stable manifold of q_2. Second we implement a simple continuation scheme to show the success of our procedure for several values of ρ in the interval $[1.33, 3.2]$.

We continue our study with an application of our methods to the rigorous solution of initial value problems at the classical parameter values $\beta = \frac{8}{3}$, $\sigma = 10$ and $\rho = 28$. This is motivated by the fact that we can concentrate on the role of the discretization method. We use this opportunity to compare the spline and the spectral approach. See in particular 4.2.2. We finish by presenting an application of the spline based boundary value approach to compute connecting orbits at the parameter values $\beta = \frac{8}{3}$ and $\sigma = -2.2$ for several values of ρ.

4.2.1 Discretization-free approach

In this section we describe the proof of the following Theorem.

Theorem 4.2.1 *Let the parameters $\beta = \frac{8}{3}$ and $\sigma = -2.2$ in the Lorenz system* (4.49) *be fixed. Define the parameter set for ρ by*

$$U = \{\rho : \rho = 1.33 + k0.01 \ \textit{with} \ k = 0, \dots, 186\}.$$

Then for every $\rho \in U$ there exists a tranverse connecting orbit from the origin to the secondary equilibrium q_2.

We first like to give some more details about the computation of the parametrization maps P and Q for the case of the Lorenz equations.

Details on the computation of the parametrization We focus our attention on the parametrization P of the stable manifold of the secondary equilibrium q_2 for ρ fixed at 1.33.

Assume the parameter values $\beta = \frac{8}{3}$, $\sigma = -2.2$ and $\rho = 1.33$ to be chosen. Denote by $\lambda^s_{1,2}$ the stable eigenvalues of $Dg(q_2)$ with corresponding eigenvectors $\xi^s_{1,2}$, where g is the Lorenz vector field. Note that we have

$\lambda^s_{1,2} = -1.5198 \pm 0.3893i \in \mathbb{C}$ with $\lambda^s_1 = \overline{\lambda^s_2}$ and $\xi^s_1 = \overline{\xi^s_2} \in \mathbb{C}^3$. Hence $m_s = 0$ and $l_s = 1$. This directly implies that $\lambda^s_{1,2}$ are non-resonant. More precisely assume that for a general $\lambda \in \mathbb{C} \setminus \mathbb{R}$ and for $k_{1,2} \in \mathbb{N}$

$$k_1\bar{\lambda} + k_2\lambda = \lambda \Leftrightarrow k_1\bar{\lambda} = (1-k_2)\lambda.$$

Comparing real and imaginary parts directly yields

$$\begin{aligned} k_1 &= 1-k_2 \\ k_1 &= -(1-k_2), \end{aligned}$$

which entails $k_1 = 0$ and $k_2 = 1$, which is of course a tautology. A similar reasoning applies to $k_1\bar{\lambda} + k_2\lambda = \bar{\lambda}$. Together it follows that in particular $\lambda^s_{1,2}$ are non-resonant. Hence we can use (2.15) in order to compute the coefficients $(a_k)_{|k|\geq 0}$, $a_k = ((a_k)_1, (a_k)_2, (a_k)_3) \in \mathbb{C}^3$, $k = (k_1, k_2) \in \mathbb{N}^2$ necessary to define $P : \mathbb{R}^2 \supset V_{\nu_s} \to \mathbb{R}^3$ by

$$\begin{aligned} P(\phi) = P(\phi_1, \phi_2) = f(\phi_1 + i\phi_2, \phi_1 - i\phi_2) = \\ = \sum_{\substack{k_1+k_2=0 \\ k_i \geq 0}}^{\infty} a_{(k_1,k_2)}(\phi_1 + i\phi_2)^{k_1}(\phi_1 - i\phi_2)^{k_2}. \end{aligned}$$

The constant and linear constraints yield $a_{0,0} = q_2$, $a_{(1,0)} = \xi^s_1$ and $a_{(0,1)} = \xi^s_2$. Next we give a concrete expression for the coefficients c_k involved in (2.15). In order to do so the following remark is helpful.

Remark 4.2.1 (Cauchy-Product in two variables) *Omitting details of convergence questions for the moment, we need to be able to compute Cauchy-products of the form*

$$\begin{aligned} &\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} a_{(k_1,k_2)} z_1^{k_1} z_2^{k_2} \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} b_{(k_1,k_2)} z_1^{k_1} z_2^{k_2} = \\ &= \sum_{k_1=0}^{\infty} \underbrace{\sum_{k_2=0}^{\infty} a_{(k_1,k_2)} z_2^{k_2}}_{\overset{\text{def}}{=}\alpha_{k_1}} z_1^{k_1} \sum_{k_1=0}^{\infty} \underbrace{\sum_{k_2=0}^{\infty} b_{(k_1,k_2)} z_2^{k_2}}_{\overset{\text{def}}{=}\beta_{k_1}} z_1^{k_1} = \\ &= \sum_{k_1=0}^{\infty} \left(\sum_{n_1=0}^{k_1} \alpha_{n_1}\beta_{k_1-n_1} \right) z_1^{k_1}. \end{aligned}$$

Computing the Cauchy-Product $\alpha_{n_1}\beta_{k_1-n_1}$ we get

$$\begin{aligned} \alpha_{n_1}\beta_{k_1-n_1} &= \sum_{k_2=0}^{\infty} a_{(n_1,k_2)} z_2^{k_2} \sum_{k_2=0}^{\infty} b_{(k_1-n_1,k_2)} z_2^{k_2} = \\ &= \sum_{k_2=0}^{\infty} \left(\sum_{n_2=0}^{k_2} a_{(n_1,n_2)} b_{(k_1-n_1,k_2-n_2)} \right) z_2^{k_2}. \end{aligned}$$

In summary we get

$$\begin{aligned}
&\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} a_{(k_1,k_2)} z_1^{k_1} z_2^{k_2} \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} b_{(k_1,k_2)} z_1^{k_1} z_2^{k_2} = \\
&= \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty} \left(\sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} a_{(n_1,n_2)} b_{(k_1-n_1,k_2-n_2)} \right) z_1^{k_1} z_2^{k_2}.
\end{aligned} \tag{4.50}$$

Note that the change from summation over $k_1 + k_2 = k$, $k_{1,2} \geq 0$ to the double sum over $k_{1,2}$ corresponds to a change of summation order that we assume to be possible by absolute convergence of the series involved.

Recalling the Lorenz vector field $g : \mathbb{R}^3 \to \mathbb{R}^3$ stipulated in (4.49) and using (4.50) the left hand side of the functional equation (2.11) for f reads explicitly as

$$\begin{aligned}
&\sum_{\substack{k_1+k_2=0\\ k_i \geq 0}}^{\infty} \left(\begin{pmatrix} \sigma\left((a_{(k_1,k_2)})_2 - (a_{(k_1,k_2)})_1 \right) \\ \rho (a_{(k_1,k_2)})_1 - (a_{(k_1,k_2)})_2 \\ -\beta (a_{(k_1,k_2)})_3 \end{pmatrix} + \right. \\
&\left. \begin{pmatrix} 0 \\ -\sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_3 \\ \sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_2 \end{pmatrix} \right) z_1^{k_1} z_2^{k_2}.
\end{aligned} \tag{4.51}$$

Concerning the right hand side we get

$$\begin{aligned}
&\left(\sum_{k_1=1}^{\infty}\sum_{k_2=0}^{\infty} a_{(k_1,k_2)} k_1 z_1^{k_1-1} z_2^{k_2} \right) \lambda_1^s z_1 + \left(\sum_{k_1=0}^{\infty}\sum_{k_2=1}^{\infty} a_{(k_1,k_2)} k_2 z_1^{k_1} z_2^{k_2-1} \right) \lambda_2^s z_2 = \\
&= a_{(1,0)} \lambda_1^s z_1^{k_1} + a_{(0,1)} \lambda_2^s z_2^{k_2} + \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} a_{(k_1,k_2)} \left(k_1 \lambda_1^s + k_2 \lambda_2^s \right) z_1^{k_1} z_2^{k_2}.
\end{aligned} \tag{4.52}$$

Matching like powers we rediscover the linear constraints for $(k_1, k_2) = (0, 0)$ and $(k_1, k_2) \in \{(1, 0), (0, 1)\}$. In preparation for matching like powers for $k_1 + k_2 \geq 2$ we realize

$$\begin{aligned}
&\begin{pmatrix} 0 \\ -\sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_3 \\ \sum_{n_1=0}^{k_1}\sum_{n_2=0}^{k_2} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_2 \end{pmatrix} \\
&= \begin{pmatrix} 0 \\ -\sum_{n_1=0}^{k_1'}\sum_{n_2=0}^{k_2'} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_3 \\ \sum_{n_1=0}^{k_1'}\sum_{n_2=0}^{k_2'} (a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_2 \end{pmatrix} + \\
&\begin{pmatrix} 0 & 0 & 0 \\ -(a_{(0,0)})_3 & 0 & -(a_{(0,0)})_1 \\ (a_{(0,0)})_2 & (a_{(0,0)})_1 & 0 \end{pmatrix} a_{(k_1 k_2)},
\end{aligned}$$

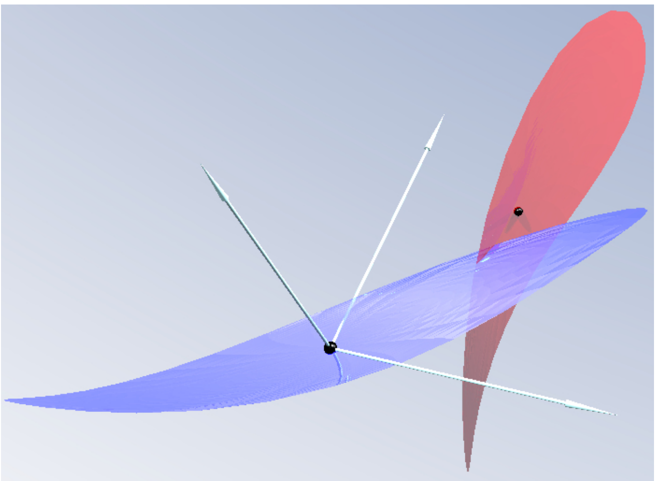


Figure 4.4: Local Stable (red) and local unstable (blue) manifolds for Lorenz when $\sigma = -2.2$, $\beta = 8/3$ and $\rho = 1.33$. Black spheres denote the location of the fixed points.

where we define the notation $\sum_{n_1=0}^{k_1'}\sum_{n_2=0}^{k_2'}$ to denote that

$$(n_1, n_2) \notin \{(0,0), (k_1, k_2)\}.$$

Finally we match like powers for $k_1 + k_2 \geq 2$ by comparing (4.52) to (4.51) and get the homological equation

$$\begin{pmatrix} -\sigma - (k_1\lambda_1^s + k_2\lambda_2^s) & \sigma & 0 \\ \rho - (a_{(0,0)})_3 & -1 - (k_1\lambda_1^s + k_2\lambda_2^s) & -(a_{(0,0)})_1 \\ (a_{(0,0)})_2 & (a_{(0,0)})_1 & -\beta - (k_1\lambda_1^s + k_2\lambda_2^s) \end{pmatrix} a_{(k_1 k_2)} = c_{(k_1,k_2)} \tag{4.53}$$

with

$$c_{(k_1,k_2)} = \begin{pmatrix} 0 \\ \sum_{n_1=0}^{k_1'}\sum_{n_2=0}^{k_2'}(a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_3 \\ -\sum_{n_1=0}^{k_1'}\sum_{n_2=0}^{k_2'}(a_{(n_1,n_2)})_1 (a_{(k_1-n_1,k_2-n_2)})_2 \end{pmatrix}.$$

We identify the structure

$$(Dg(q_1) - (\lambda_1^s k_1 + \lambda_2^s k_2)\mathbf{1}_{3,3})\, a_{(k_1,k_2)} = c_{(k_1,k_2)},$$

as stated in Lemma 2.1.1.

Validation of connection at $\beta = \frac{8}{3}$, $\sigma = -2.2$ and $\rho = 1.33$ Next we discuss the validation process more closely. We start by choosing parametrization orders $M = 45$ at q_1 and $N = 35$ at q_2. Note that the choice of a higher parametrization order for the unstable manifold of the origin is motivated by the fact that the unstable eigenvalues at the origin are given by $\lambda^u_{1,2} = 0.6000 \pm 0.6049i$ and thus the spectral gap μ from Theorem 2.1.1 is smaller than for $\lambda^s_{1,2}$ given above, necessitating a higher order to get the same accuracy. Furthermore we set the domain sizes to $\nu_u = 0.725$ and $\nu_s = 0.575$. As $m_u = m_s = 0$ (i.e. we have complex conjugate eigenvalues at both q_1 and q_2) the norm in parameter space $V_{\nu_{u,s}} \subset \mathbb{R}^2$ is the euclidian norm. Figure 4.4 shows the image of V_{ν_s} under the truncated stable parametrization P_N and of V_{ν_u} under the truncated unstable parametrization Q_M suggesting the existence of a transversal intersection. For these domains and parametrization orders we obtain the following numerical defects in the functional equation (2.11):

$$\|g \circ f_N - Df_N\Lambda_s\|_{\Sigma,\nu_s} \leq 6.29 \times 10^{-14}$$

and

$$\|g \circ h_M - Dh_M\Lambda_u\|_{\Sigma,\nu_u} \leq 1.09 \times 10^{-13},$$

where $\Lambda_{u,s} \in \mathbb{C}^{2,2}$ are diagonal matrices with the (un)stable eigenvalues on the diagonal. Using Theorem 2.1.1 we get a-posteriori error bounds $\delta_u = 3.3 \times 10^{-14}$ and $\delta_s = 3.5 \times 10^{-14}$.

To proof the existence of a connection we go on to set a phase condition in unstable parameter space. We pick a circle of radius $\nu = 0.70796507495989$, i.e. we set

$$\Theta(\alpha) = \nu \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix}.$$

This completes the concrete formulation of F given by (3.16). Using a classical Newton iteration we obtain an approximate solution

$$\bar{\alpha} = -1.08544433208255 \quad \text{(radians)}$$

and

$$\bar{\phi} = (-0.018554373780656, 0.548268655433034)$$

with

$$\|F(\bar{\alpha}, \bar{\phi})\|_\infty < 2.25 \times 10^{-14}.$$

By applying the validation Theorem 3.1.2 we obtain rigorous error bounds on $(\bar{\alpha}, \bar{\phi})$ and get the radius $\bar{r} = 2.72 \times 10^{-12}$ such that we can guarantee a unique solution with $\|\Theta(\tilde{\alpha}) - \Theta(\bar{\alpha})\|_2 < \bar{r}$ and $\|\tilde{\phi} - \bar{\phi}\|_2 < \bar{r}$. The corresponding computations are carried out by *Disfreerho133.m* to be

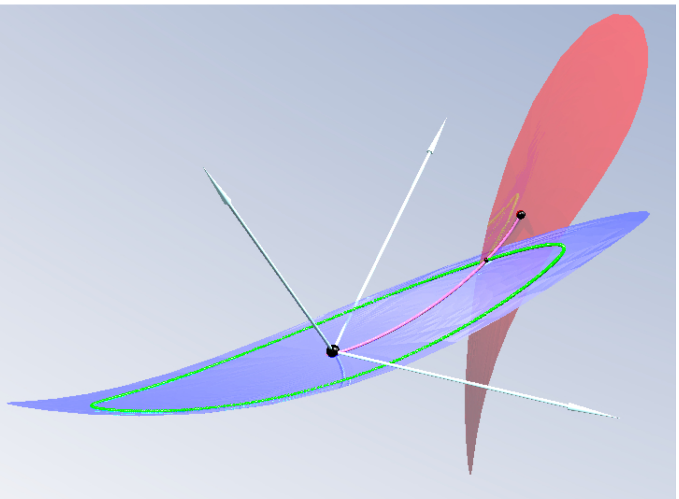


Figure 4.5: Validated *transversal connecting orbit* for Lorenz when $\sigma = -2.2$, $\beta = 8/3$ and $\rho = 1.33$. The image of the phase condition Θ is shown as a green circular arc on the local stable manifold. The solution of the discretization-free operator F is shown as a black dot on the intersection of the manifolds and the green phase arc. The pink arc is the image under the parameterizations of the flow in parameter space.

found at [28]. In addition we can compute a representation of the connecting orbit by flowing $\Theta(\bar{\alpha})$ in V_{ν_u} and $\bar{\phi}$ in V_{ν_s} under the linear flows of $J_{u,s}$ (induced by $\Lambda_{u,s}$) and lifting to phase space via (2.9). The result is shown in Figure 4.5. We now go on to scrutinize this process more closely and in particular compare this integration of the nonlinear flow via the conjugated linear flow to integration of the flow in phase space.

Nonlinear flow integration in phase space vs linear integration in parameter space Suppose we would approximate the (un)stable manifold of $q_{1,2}$ by a linear approximation as done in the classical approach of projected boundary conditions used in [1, 19]. We ask the following questions:

1. Which domain sizes $\nu_{u,s}^{lin}$ do we have to choose in order to get the same a-posteriori accuracy, that is $\delta_{u,s} \cong 3 \times 10^{-14}$?

2. How long do we need to integrate the linear flow in parameter spaces $V_{\nu_{u,s}}$ in order to flow the approximations $\bar{\alpha}$ and $\bar{\phi}$ from above to get to these smaller neighborhoods $V_{\nu_{u,s}^{lin}}$?

3. How does standard numerical integration behave in comparison?

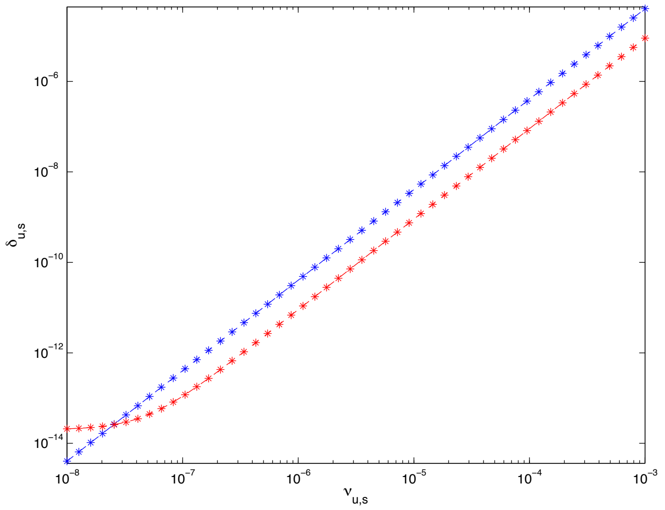


Figure 4.6: Dependence of $\delta_{u,s}$ on $\nu^{lin}_{u,s}$, where δ_u is shown in blue and δ_s is depicted in red.

Concerning the first question we resort to a heuristic study. Figure 4.6 shows the dependence of the a-posteriori errors $\delta_{u,s}$ on $\nu^{lin}_{u,s}$. That is we keep the parametrization order $N = M = 1$ fixed, i.e. conduct only linear approximation of the manifolds, and vary $\nu^{lin}_{u,s}$. We find that we can choose $\nu^{lin}_u = \nu^{lin}_s = 3 \times 10^{-8}$.

Let us now turn to the second question. Recalling 2.7 we use that the linear flow in stable parameter space is induced by the matrix

$$\exp(J_s) = \exp(Re(\lambda_1^s)) \begin{pmatrix} \cos(\gamma_s) & -\sin(\gamma_s) \\ \sin(\gamma_s) & \cos(\gamma_s) \end{pmatrix}$$

where $\exp(iIm(\lambda_1^s)) = (\cos(\gamma_s) + i\sin(\gamma_s))$. A similar formular holds for $\exp(J_u)$. As a consequence $\exp(J_{u,s})$ have the structure $\exp(J_{u,s}) = \exp(Re(\lambda_1^{u,s}))D_{u,s}$ where $D_{u,s}$ is an eucledian isometry. In particular we get that

$$\|\exp(J_u t)\varphi\|_2 = e^{0.6t}\|\varphi\|_2, \qquad \text{and} \qquad \|\exp(J_s t)\phi\|_2 \approx e^{-1.52t}\|\phi\|_2.$$

Thus the unstable and stable parameters $\bar{\varphi} = \Theta(\bar{\alpha})$ and $\bar{\phi}$ can be flown in

$$-28.3 \approx \frac{1}{0.6}\log\left(\frac{3\times 10^{-8}}{0.71}\right) \qquad \text{and} \qquad 11.0 \approx \frac{-1}{1.52}\log\left(\frac{3\times 10^{-8}}{0.55}\right),$$

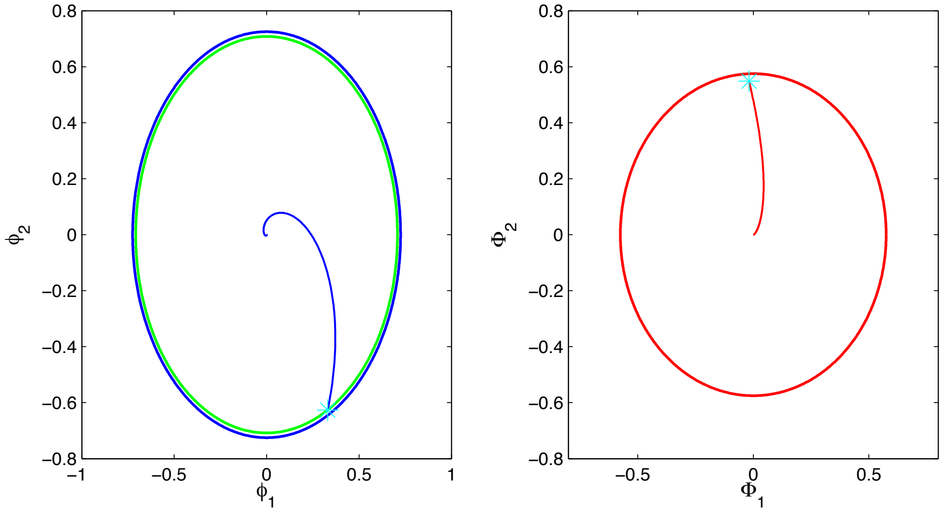


Figure 4.7: Flow in parameter space. On the left hand side the unstable parameter space is shown. The blue circle corresponds to V_{ν_u} together with the green circle illustrating the phase condition. The cyan star is $\Theta(\bar{\alpha})$ and the blue curve depicts its orbit under the linear flow induced by J_u. On the right hand side we see the corresponding picture with V_{ν_s} and $\bar{\phi}$.

time units into the neighborhoods $V_{\nu^{lin}_{u,s}}$. Figure 4.7 depicts the phase portrait of these linear integrations. Lifting these orbits via the conjugacy yields Figure 4.5.

In order to answer the third question we compare with standard numerical integration in phase space by flowing

$$q_u = Q_M(\exp(-28.3J_u)\Theta(\bar{\alpha})) \in \mathbb{R}^3$$

forward in time and check the convergence behavior by monitoring

$$d(t) = \|\Phi(q_u, t) - q_1\|_2. \tag{4.54}$$

On Figure 4.8 we see that in the beginning we can observe convergence towards q_1, but after 32 time units in the vicinity of the equilibrium numerical errors cause the orbit to switch to its local unstable manifold. This effect is circumvented completely during the integration using the conjugacy relation, where can integrate over $28.3 + 11 = 39.3$ time units without any instability effects.

Proof of Theorem 4.2.1 To proof Theorem 4.2.1 we implement a simple continuation scheme. The continuation begins with $\rho = 1.33$, $\nu_u = 0.725$,

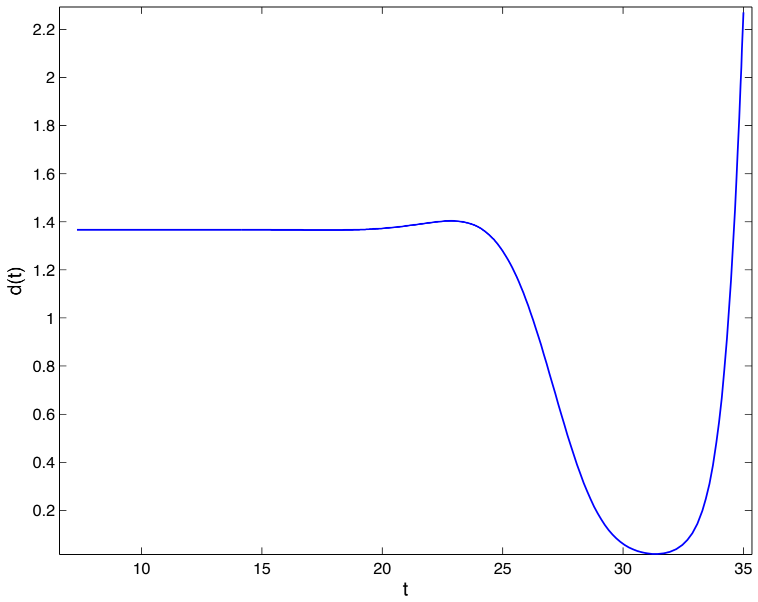


Figure 4.8: The time evolution of $d(t)$ ((4.54)) is depicted, where the integration is conducted by a standard Matlab 'ode45' integrator with $AbsTol = 10^{-16}$. We see convergence towards q_2 before at time $t \approx 32$ the minimum distance of 1.8×10^{-2} is reached. The numerical error causes then divergence along the unstable manifold of q_1. Note that the unstable eigenvalue of q_1 is $\lambda^u \approx 1.57$.

$\nu_s = 0.575$, and a phase circle fixed at $\nu = 0.70796507495989$. Each step of the continuation increases the previous value of ρ by 0.01, the previous value of ν_u by 0.0079, the previous value of ν_s by 0.0108, and the previous value of ν by 0.0079. In each computation the value of N and M are held at 30. For each value of the parameters the origin q_1 has two dimensional unstable manifold and the secondary equilibrium q_2 has two dimensional stable manifold, both with complex conjugate eigenvalues. The transversality follows by Theorem 3.3.1. The proof is completed by running *Disfreecontinuation.m* to be found at [28]. Some of the the results are summarized in Table 4.4, and seven of the resulting orbits are illustrated in Figure 4.9. It takes about three and a half hours for all 187 proofs to complete by running .

The proofs reported in Table 4.4 can be produced by running the program *DisfreecontinuationII.m* [28] without computing all 187 parameter values. This program runs in a much shorter time.

ρ	δ_u	δ_s	$\bar{r}$	Proof Time
1.35	2.26×10^{-13}	2.78×10^{-14}	1.36×10^{-11}	86.5 (sec)
1.45	3.03×10^{-13}	4.00×10^{-14}	2.86×10^{-12}	87.1 (sec)
1.55	4.83×10^{-13}	3.70×10^{-14}	3.00×10^{-12}	87.2 (sec)
1.65	8.49×10^{-13}	3.89×10^{-14}	4.87×10^{-12}	87.0 (sec)
1.75	1.54×10^{-12}	5.55×10^{-14}	8.65×10^{-12}	87.1 (sec)
1.85	2.88×10^{-12}	6.33×10^{-14}	1.60×10^{-11}	87.0 (sec)
1.95	5.47×10^{-12}	7.70×10^{-14}	3.05×10^{-11}	87.0 (sec)
2.05	1.05×10^{-11}	8.53×10^{-14}	5.84×10^{-11}	87.1 (sec)
2.15	2.03×10^{-11}	1.13×10^{-13}	1.15×10^{-10}	87.3 (sec)
2.25	3.98×10^{-11}	1.23×10^{-13}	2.28×10^{-10}	89.5 (sec)
2.35	7.95×10^{-11}	1.54×10^{-13}	4.62×10^{-10}	95.5 (sec)
2.45	1.59×10^{-10}	1.72×10^{-13}	9.39×10^{-10}	102.6 (sec)
2.55	3.16×10^{-10}	2.15×10^{-13}	1.91×10^{-9}	104.5 (sec)
2.65	6.20×10^{-10}	2.37×10^{-13}	3.84×10^{-9}	96.8 (sec)
2.75	1.20×10^{-9}	2.78×10^{-13}	7.58×10^{-9}	91.1 (sec)
2.85	2.27×10^{-9}	3.34×10^{-13}	1.50×10^{-8}	91.2 (sec)

Table 4.4: Proof of short-connections for sixteen different values of ρ.

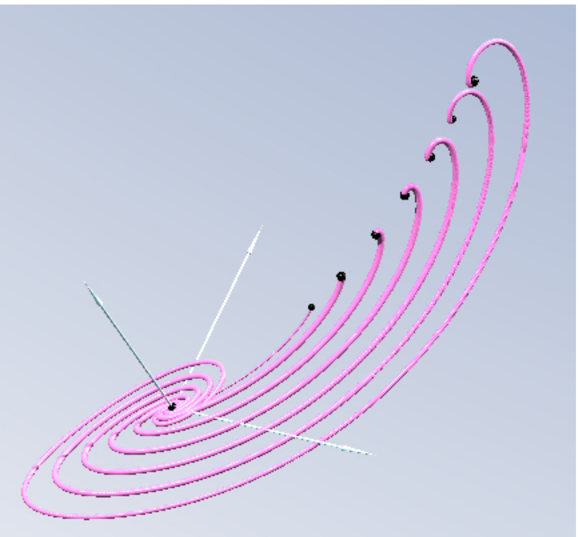


Figure 4.9: Seven validated *transversal connecting orbits* for Lorenz with $\sigma = -2.2$, $\beta = 8/3$ and ρ taking the values $1.35, 1.55, 1.75, 2.05, 2.25, 2.55$, and 2.75.

4.2.2 Rigorous solutions of IVPs

Next we consider rigorous solutions of IVPs using both the spline and the chebyshev approach. This will enable us to compare the two discretization methods explicitly. We start by presenting the Chebyshev approach and go on to compare the Spline approach.

Chebyshev approach

Recall that the operator f is given by (3.52) and using Lemma 2.2.2 we get that the chebyshev coefficients c_k of $g \circ u$ are explicitly defined by

$$c_k = L \begin{pmatrix} \sigma((a_2)_k - (a_1)_k) \\ \rho(a_1)_k - (a_2)_k - (a_1 a_3)_k \\ (a_1 a_2)_k - \beta(a_3)_k \end{pmatrix} \tag{4.55}$$

with

$$(a_n a_m)_k = \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} (a_n)_{|k_1|} (a_m)_{|k_2|}$$

for $n = 1$, $m = 1, 2$ and $k \geq 0$. Finally assuming an initial value p_1 to be given this yields an explicit expression for the IVP operator f. We emphasize again that in this case there are no invariant manifolds involved and the operator f has no dependence on parametrizations. Let us go on to present some rigorous numerical results for different initial values p_1.

Theorem 4.2.2 *Consider*

$$\begin{aligned} p_1^1 &= (8.102574164767477, 9.551574461919124, 24.429705657930224) \\ p_1^2 &= (-0.208252089096454, -0.454566900892446, 0) \\ p_1^3 &= (4.102702069909453, 8.936495309135337, 0.5789130478426856). \end{aligned}$$

Let $s = 2$. For $p_0 \in \{p_1^1, p_1^2, p_1^3\}$ consider the IVP-operator f given by (3.52) *with c_k as in* (4.55)*. For each L in Table 4.5 there exists a unique solution $\tilde{x} \in X^s$ of $f(x) = 0$ in a ball $B_{\bar{x}}(\bar{r}_{p_1^{1,2,3}}) \subset X^s$ of radius $\bar{r}_{p_1^{1,2,3}}$ centered at an approximate solution $\bar{x}$.*

L	0.5	1	1.5	2	2.5	3
$m_{p_1^1}$	50	100	200	250	300	500
$m_{p_1^2}$	300	300	300	350	500	failed
$m_{p_1^3}$	150	200	300	400	500	600
$\bar{r}_{p_1^1}$	2.61×10^{-9}	1.27×10^{-8}	2.85×10^{-8}	8.77×10^{-8}	4.53×10^{-7}	1.03×10^{-6}
$\bar{r}_{p_1^2}$	1.92×10^{-7}	6.81×10^{-7}	1.49×10^{-6}	2.60×10^{-6}	4.81×10^{-6}	–
$\bar{r}_{p_1^3}$	1.07×10^{-7}	1.31×10^{-7}	6.29×10^{-7}	1.09×10^{-6}	1.40×10^{-6}	5.17×10^{-6}

Table 4.5: Given $p_1^{1,2,3}$ and for a fixed L, these are corresponding values of the Galerkin projection dimension $m_{p_1^{1,2,3}}$ and the radius $\bar{r}_{p_1^{1,2,3}}$ around the approximate solution $\bar{x}$ in X^s for which the radii polynomials approach was successful.

Before we discuss the proof via an application of Theorem 3.2.2 we comment on the choice of the initial conditions. p_1^1 is chosen to lie approximately on the unstable manifold of the *positive eye* equilibrium

$$(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho - 1),$$

p_1^2 lies approximately on the unstable manifold of the origin whereas p_1^3 is taken randomly according to a uniform distribution in $[-10, 10] \times [-10, 10] \times [-10, 10]$. As one can see in Table 4.5, the data of the verification method depends strongly on the choice of the initial condition. We assume that this stems from the presence of poles of the complex extension of the solutions $u : [-1, 1] \to \mathbb{R}^3$ of (4.49) whose position in the complex plane changes depending on the initial condition and the scaling factor L. By Theorem 2.2.2 this influences the decay rate of the Chebyshev coefficients. This is illustrated in Figure 4.10. We refer to Figure 4.11 for a representation in phase space of two solutions of Theorem 4.2.2.

The proof of Theorem 4.2.2 can be found in the MATLAB programs *proofLorenz1.m*, *proofLorenz2.m* and *proofLorenz3.m* at [28]. It relies on Theorem 3.2.2 and uses the package Intlab [48] for the interval computations and the package Chebfun [55]. In order to apply Theorem 3.2.2 the construction of the radii polynomials as defined in (3.62) and (3.63) is crucial. After the following remark we aim to give some details about the derivation of the bounds defined in (3.59), (3.60) and (3.61) involved in the construction of the polynomials.

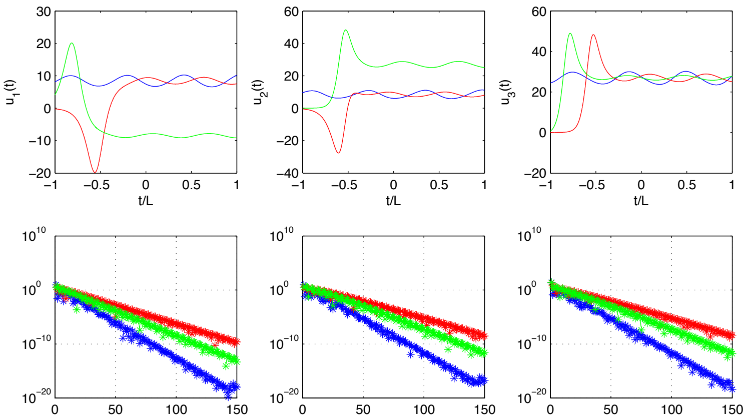
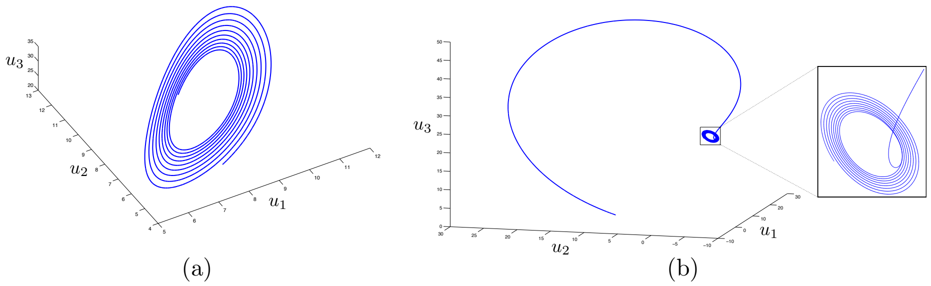


Figure 4.10: Comparison of the componentwise solution profiles of a solution $u : [-1,1] \to \mathbb{R}^3$ of the Lorenz equations for the initial condition p_1^1 (blue), p_1^2 (red) and p_1^3 (green) for $L = 1$ and of the decay rates of their Chebyshev coefficient sequences.



(a) (b)

Figure 4.11: Profile in phase space of the solution (u_1, u_2, u_3) of the Lorenz equations starting at (a) the initial condition p_1^1; (b) the initial condition p_1^3.

Remark 4.2.2 *Consider an approximate solution $\bar{x}$ and a corresponding unique genuine solution $\tilde{x} \in B_{\bar{x}}(r) \subset X^s$ of $f(x) = 0$ for a decay rate $s > 1$ and a radius $r > 0$. Via the expansion* (3.43) *the sequences of Chebyshev coefficients $\bar{x}$*

and $\tilde{x}$ correspond to functions $\bar{u}$ and $\tilde{u}$ respectively, where $\tilde{u}$ solves (4.49) *with respective initial condition p_1. Given $s > 1$, the inequality $\|\bar{x} - \tilde{x}\|_s \leq r$ can be used to get that*

$$\begin{aligned}
\|\bar{u} - \tilde{u}\|_\infty &\overset{\text{def}}{=} \sup_{t\in[-1,1]} \|\bar{u}(t) - \tilde{u}(t)\|_\infty \\
&\leq \|\bar{a}_0 - \tilde{a}_0\|_\infty + 2 \sup_{t\in[-1,1]} \sum_{k=1}^{\infty} \|\bar{a}_k - \tilde{a}_k\|_\infty \underbrace{|T_k(t)|}_{\leq 1} \\
&\leq \left(1 + 2\sum_{k=1}^{\infty} \frac{1}{\omega_k^s}\right) r \leq \left(3 + \frac{2}{s-1}\right) r.
\end{aligned}$$

We now turn to the computations of the bounds involved in the construction of the radii polynomials.

Derivation of the Y and Z bounds Recalling (3.52) we do not have additional boundary conditions and thus are in the case $p = 0$ and $k_0 = 0$. As we consider a 3D ODE we set $d = 3$. Let us now derive the quantities $Y_0, \ldots, Y_{M-1}$, $Z_0(r), \ldots, Z_{M-1}(r) \in \mathbb{R}^3$ and $Z_M(r) \in \mathbb{R}^3$.

Assume a dimension m for the Galerkin projection explained in (3.55) to be given. We start by explicitly stating the Galerkin projection

$$f^{(m)} : \mathbb{R}^{3m} \to \mathbb{R}^{3m}.$$

Let $x_F = (a_0, \ldots, a_{m-1}) = a_F \in \mathbb{R}^{3m}$ then $f^{(m)}(x_F)$ is given component-wise by

$$\begin{aligned}
f_0^{(m)}(x_F) &= p_1 - a_0 + c_0^{(m)} + \frac{c_1^{(m)}}{2} - 2\sum_{j=2}^{m-1} \frac{1}{j^2-1} c_j^{(m)} \\
f_k^{(m)}(x_F) &= 2ka_k + c_{k+1}^{(m)} - c_{k-1}^{(m)} \;\; k = 1, \ldots, m-2 \\
f_{m-1}^{(m)}(x_F) &= 2(m-1)a_{m-1} + \begin{pmatrix} 0 \\ -L(a_1a_3)_m^{(m)} \\ L(a_1a_2)_m^{(m)} \end{pmatrix} - c_{m-2}^{(m)}
\end{aligned}$$

where we define

$$c_k^{(m)} = L\left(\begin{pmatrix} \sigma((a_2)_k - (a_1)_k) \\ \rho(a_1)_k - (a_2)_k \\ -\beta(a_3)_k \end{pmatrix} + \begin{pmatrix} 0 \\ -(a_1a_3)_k^{(m)} \\ (a_1a_2)_k^{(m)} \end{pmatrix}\right)$$

with the finite convolution sums

$$(a_1a_2)_k^{(m)} = \sum_{\substack{k_1+k_2=k\\|k_i|<m}} (a_1)_{|k_1|}(a_2)_{|k_2|}$$
$$(a_1a_3)_k^{(m)} = \sum_{\substack{k_1+k_2=k\\|k_i|<m}} (a_1)_{|k_1|}(a_3)_{|k_2|}.$$

Note that as $|k_i| \leq m-1$ for $i = 1, 2$

$$(a_1a_2)_k^{(m)} = 0 \text{ whenever } k > 2(m-1). \tag{4.56}$$

Assume that a numerical approximation $\bar{x}_F$ with $f^{(m)}(\bar{x}_F) \approx 0$ is given and define $\bar{x} = (\bar{x}_F, 0_\infty)$. In a similar fashion as in (4.12) we notice that by (4.56), setting $M = 2m-1$ suffices to fulfill assumption **A.1** from Section 3.2.2.

Setting $\bar{M} = M-1$ our next goal is to compute bounds $Y_0, \ldots, Y_{\bar{M}}$ such that

$$|(T\bar{x}-\bar{x})_k| \preceq Y_k$$

for $k = 0, \ldots, \bar{M}$. We therefore define

$$\bar{x}_{\bar{M}} = (\bar{a}_0, \ldots, \bar{a}_{m-1}, \underbrace{0_3, \ldots 0_3}_{m-2 \text{ times}})$$

and compute $y = f^{(\bar{M})}(\bar{x}_{\bar{M}})$. Then we set Y_k as

$$Y_k = \begin{cases} (|A_m||y_F|)_k & k = 0, \ldots m-1 \\ \frac{|y_k|}{2k} & k = m, \ldots, \bar{M}. \end{cases} \tag{4.57}$$

Recalling (3.69) our next step is to compute polynomials $z_k(r) = z_1^k r + z_2^k r^2$ such that

$$Df_k(\bar{x}+rw)rv - (A^\dagger rv)_k = z_k(r)$$

for $k \geq 0$. The componentwise estimation of $|z_k(r)|$ as a major step to obtain $Z_k(r)$ is postponed to a separate consideration. We note that we have to distinguish the cases $k = 0$, $1 \leq k \leq m-2$, $m \leq k$.

Starting with $k = 0$ we compute for arbitrary $x = a \in X^s$, $y = b \in X^s$ and $z_F = d_F \in \mathbb{R}^{3m}$

$$Df_0(x)y - Df_0^{(m)}(z_F)y_F = \frac{d}{dt}f_0(x+ty)|_{t=0} - \frac{d}{dt}f_0^{(m)}(z_F+ty_F)|_{t=0}$$

Keep in mind that we will later want to apply these calculations with $x = \bar{x} + \xi_1$, $z_F = \bar{x}_F$ and $y = \xi_2$ with $\xi_i \in B_r(0) \subset X^s$. In this context it is essential that we compute first

$$Dc_k(a)b = \frac{d}{dt} c_k(a + tb)|_{t=0} \quad \text{and} \quad Dc_k^{(m)}(d_F)b_F = \frac{d}{dt} c_k^{(m)}(d_F + tb_F)|_{t=0}$$

for every $k \geq 0$. In order to achieve this we have to consider

$$\begin{aligned}
&\frac{d}{dt}((a_1 + tb_1)(a_3 + tb_3))_k|_{t=0} = \\
&= \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} \frac{d}{dt}(a_1 + tb_1)_{|k_1|}(a_3 + tb_3)_{|k_2|}|_{t=0} = \\
&= \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} \frac{d}{dt} \left[(a_1)_{|k_1|}(a_3)_{|k_2|} + t((a_1)_{|k_1|}(b_3)_{|k_2|} + (a_3)_{|k_1|}(b_1)_{|k_2|}) + \right. \\
&\qquad\qquad \left. t^2 (b_1)_{|k_1|}(b_3)_{|k_2|} \right] |_{t=0} = \\
&= \sum_{\substack{k_1+k_2=k \\ k_i \in \mathbb{Z}}} (a_1)_{|k_1|}(b_3)_{|k_2|} + (a_3)_{|k_1|}(b_1)_{|k_2|} = (a_3 b_1)_k + (a_1 b_3)_k
\end{aligned}$$

and in an analogue way

$$\frac{d}{dt}((a_1 + tb_1)(a_2 + tb_2))_k|_{t=0} = (a_1 b_2)_k + (a_2 b_1)_k.$$

This entails in particular

$$\begin{aligned}
\frac{d}{dt}((a_1 + tb_1)(a_3 + tb_3))_k^{(m)}|_{t=0} &= (a_3 b_1)_k^{(m)} + (a_1 b_3)_k^{(m)} \\
\frac{d}{dt}((a_1 + tb_1)(a_2 + tb_2))_k^{(m)}|_{t=0} &= (a_1 b_2)_k^{(m)} + (a_2 b_1)_k^{(m)}.
\end{aligned}$$

Hence we obtain

$$Dc_k(a)b = L \left[\begin{pmatrix} (\sigma((b_2)_k - (b_1)_k) \\ \rho(b_1)_k - (b_2)_k \\ -\beta(b_3)_k \end{pmatrix} + \begin{pmatrix} 0 \\ -((a_3 b_1)_k + (a_1 b_3)_k) \\ (a_1 b_2)_k + (a_2 b_1)_k \end{pmatrix} \right]$$

and for $k = 0, \dots m - 1$

$$Dc_k^{(m)}(d_F)b_F = L \left[\begin{pmatrix} (\sigma((b_2)_k - (b_1)_k) \\ \rho(b_1)_k - (b_2)_k \\ -\beta(b_3)_k \end{pmatrix} + \begin{pmatrix} 0 \\ -((d_3 b_1)_k^{(m)} + (d_1 b_3)_k^{(m)}) \\ (d_1 b_2)_k^{(m)} + (d_2 b_1)_k^{(m)} \end{pmatrix} \right].$$

We now go on to consider

$$Df_0(x)y - Df_0^{(m)}(d_F)y_F = \Big[(Dc_0(a)b - Dc_0^{(m)}(d_F)b_F) - \frac{1}{2}(Dc_1(a)b - Dc_1^{(m)}(d_F)b_F) - 2\sum_{j=2}^{m-1}\frac{1}{j^2-1}(Dc_j(a)b - Dc_j^{(m)}(d_F)b_F) -2\sum_{j=m}^{\infty}\frac{1}{j^2-1}Dc_j(a)b\Big]. \tag{4.58}$$

Let us next set $x = \bar{x} + \xi_1$, $d_F = \bar{x}_F$ and $y = \xi_2$ where $\xi_1 = rw$ and $\xi_2 = rv$. We hence obtain for $k \geq 0$

$$Dc_k(\bar{x} + rw)rv = \beta_1^k r + \beta_2^k r^2$$

with

$$\begin{aligned} \beta_1^k &= L\left[\begin{pmatrix}\sigma(v_2)_k - (v_1)_k) \\ \rho(v_1)_k - (v_2)_k \\ -\beta(v_3)_k\end{pmatrix} + \begin{pmatrix} 0 \\ -(\bar{a}_3 v_1)_k - (\bar{a}_1 v_3)_k \\ (\bar{a}_1 v_2)_k + (\bar{a}_2 v_2)_k\end{pmatrix}\right] \\ \beta_2^k &= L\begin{pmatrix} 0 \\ -(w_3 v_1)_k - (w_1 v_3)_k \\ (w_1 v_2)_k + (w_2 v_1)_k\end{pmatrix}. \end{aligned} \tag{4.59}$$

In addition we have

$$Dc_k(\bar{a} + rw)rv - Dc_k^m(\bar{a}_F)rv_F = \kappa_1^k r + \kappa_2^k r^2$$

with

$$\kappa_1^k = L\begin{pmatrix} 0 \\ -(\bar{a}_3 v_1^I)_k - (\bar{a}_1 v_3^I)_k \\ (\bar{a}_1 v_2^I)_k + (\bar{a}_2 v_2^I)_k\end{pmatrix} \quad k = 0, \ldots, m-1, \tag{4.60}$$

where

$$(v_{1,2,3}^I)_k = \begin{cases} 0 & 0 \leq k \leq m-1 \\ (v_{1,2,3})_k & k \geq m \end{cases}$$

and $\kappa_2^k = \beta_2^k$ for $k = 1, \ldots, m-1$.

We hereby obtain

$$\begin{aligned} &Df_0(\bar{x} + rw)rv - Df_0^{(m)}(\bar{x}_F)rv_F = \\ &= \sum_{l=1}^{2}\kappa_l^0 r^l + \frac{1}{2}\sum_{l=1}^{2}\kappa_l^1 r^l + \sum_{j=2}^{m-1}\frac{1}{j^2-1}\left(\sum_{l=1}^{2}\kappa_l^j r^l\right) + \sum_{j=m}^{\infty}\frac{1}{j^2-1}\left(\sum_{l=1}^{2}\beta_l^j r^l\right) \\ &= \sum_{l=1}^{2} z_l^0 r^l \end{aligned}$$

with

$$\begin{aligned} z_l^0 &= \left[\kappa_l^0 - \frac{1}{2}\kappa_l^1 - 2\sum_{j=2}^{m-1}\frac{1}{j^2-1}\kappa_l^j - 2\sum_{j=m}^{\infty}\frac{1}{j^2-1}\beta_l^j\right] \qquad l=1,2 \\ &= \begin{cases} \left[\kappa_l^0 - \frac{1}{2}\kappa_l^1 - 2\sum_{j=2}^{m-1}\frac{1}{j^2-1}\kappa_l^j - 2\sum_{j=m}^{\infty}\frac{1}{j^2-1}\beta_l^j\right] & l=1 \\ \left[\beta_l^0 - \frac{1}{2}\beta_l^1 - 2\sum_{j=2}^{\infty}\frac{1}{j^2-1}\beta_l^j\right] & l=2 \end{cases}. \end{aligned} \tag{4.61}$$

We now go ahead to consider the cases $1 \le k \le m-1$. We obtain

$$\begin{aligned} &Df_k(\bar{x}+rw)rv - (A^\dagger rv)_k = Df_k(\bar{x}+rw)rv - Df_k^{(m)}(\bar{x}_F)rv_F = \\ &= \Big(Dc_{k+1}(\bar{a}+rw)rv - Dc_{k+1}^{(m)}(\bar{a}_F)rv_F) \\ &\quad -(Dc_{k-1}(\bar{a}+rw)rv - Dc_{k-1}^{(m)}(\bar{a}_F)rv_F\Big) \\ &= \sum_{l=1}^{2}(\kappa_l^{k+1} - \kappa_l^{k-1})r^l = \sum_{l=1}^{2} z_l^k r^l, \end{aligned}$$

where we define for $l = 1, 2$

$$\begin{aligned} z_l^k &= (\kappa_l^{k+1} - \kappa_l^{k-1}) \\ &= \begin{cases} (\kappa_l^{k+1} - \kappa_l^{k-1}) & l=1 \\ (\beta_l^{k+1} - \beta_l^{k-1}) & l=2 \end{cases} \end{aligned} \tag{4.62}$$

with κ_l^k and β_l^k given as in (4.60) and (4.59).

Finally for $k \ge m$ we have that

$$\begin{aligned} &Df_k(\bar{x}+rw)rv - (A^\dagger rv)_k = Df_k(\bar{x}+rw)rv - 2krv_k = \\ &= L\left[Dc_{k+1}(\bar{a}+rw)rv - Dc_{k-1}(\bar{a}+rw)rv\right] \\ &= \sum_{l=1}^{2}(\beta_l^{k+1} - \beta_l^{k-1})r^l = z_1^k r + z_2^k r^2, \end{aligned}$$

where we define for $l = 1, 2$

$$z_l^k = \beta_l^{k+1} - \beta_l^{k-1} \tag{4.63}$$

with β_l^k given as in (4.59) . Summarizing (4.61) and (4.63) we now have

$$Df_k(\bar{x}+rw)rv - (A^\dagger rv)_k = z_1^k r + z_2^k r^2$$

for all $k \ge 0$. An overview can be found in Table 4.6.

	$k = 0$
z_1^0	$L\left[\begin{pmatrix}0\\-(\bar{a}_3v_1^I)_0-(\bar{a}_1v_3^I)_0\\(\bar{a}_1v_2^I)_0+(\bar{a}_2v_2^I)_0\end{pmatrix}-\frac{1}{2}\begin{pmatrix}0\\-(\bar{a}_3v_1^I)_1-(\bar{a}_1v_3^I)_1\\(\bar{a}_1v_2^I)_1+(\bar{a}_2v_2^I)_1\end{pmatrix}-2\sum_{j=2}^{m-1}\frac{1}{j^2-1}\begin{pmatrix}0\\-(\bar{a}_3v_1^I)_j-(\bar{a}_1v_3^I)_j\\(\bar{a}_1v_2^I)_j+(\bar{a}_2v_2^I)_j\end{pmatrix}\right.$ $\left.-2\sum_{j=m}^{M-1}\frac{1}{j^2-1}\left[\begin{pmatrix}\sigma((v_2)_j-(v_1)_j)\\\rho(v_1)_j-(v_2)_j\\-\beta(v_3)_j\end{pmatrix}+\begin{pmatrix}0\\-(\bar{a}_3v_1)_j-(\bar{a}_1v_3)_j\\(\bar{a}_1v_2)_j+(\bar{a}_2v_2)_j\end{pmatrix}\right]-2\sum_{j=M}^{\infty}\frac{1}{j^2-1}\left[\begin{pmatrix}\sigma((v_2)_j-(v_1)_j)\\\rho(v_1)_j-(v_2)_j\\-\beta(v_3)_j\end{pmatrix}+\begin{pmatrix}0\\-(\bar{a}_3v_1)_j-(\bar{a}_1v_3)_j\\(\bar{a}_1v_2)_j+(\bar{a}_2v_2)_j\end{pmatrix}\right]\right]$
z_2^0	$L\left[\begin{pmatrix}0\\-(w_3v_1)_0-(w_1v_3)_0\\(w_1v_2)_0+(w_2v_1)_0\end{pmatrix}-\frac{1}{2}\begin{pmatrix}0\\-(w_3v_1)_0-(w_1v_3)_0\\(w_1v_2)_0+(w_2v_1)_0\end{pmatrix}-2\sum_{j=2}^{\infty}\frac{1}{j^2-1}\begin{pmatrix}0\\-(w_3v_1)_j-(w_1v_3)_j\\(w_1v_2)_j+(w_2v_1)_j\end{pmatrix}\right]$
	$k = 1,\dots,m-1$
z_1^k	$L\left[\begin{pmatrix}0\\-(\bar{a}_3v_1^I)_{k+1}-(\bar{a}_1v_3^I)_{k+1}\\(\bar{a}_1v_2^I)_{k+1}+(\bar{a}_2v_2^I)_{k+1}\end{pmatrix}-\begin{pmatrix}0\\-(\bar{a}_3v_1^I)_{k-1}-(\bar{a}_1v_3^I)_{k-1}\\(\bar{a}_1v_2^I)_{k-1}+(\bar{a}_2v_2^I)_{k-1}\end{pmatrix}\right]$
z_2^k	$L\left[\begin{pmatrix}0\\-(w_3v_1)_{k+1}-(w_1v_3)_{k+1}\\(w_1v_2)_{k+1}+(w_2v_1)_{k+1}\end{pmatrix}-\begin{pmatrix}0\\-(w_3v_1)_{k-1}-(w_1v_3)_{k-1}\\(w_1v_2)_{k-1}+(w_2v_1)_{k-1}\end{pmatrix}\right]$
	$k \geq m$
z_1^k	$L\left[\left[\begin{pmatrix}\sigma((v_2)_{k+1}-(v_1)_{k+1})\\\rho(v_1)_{k+1}-(v_2)_{k+1}\\-\beta(v_3)_{k+1}\end{pmatrix}+\begin{pmatrix}0\\-(\bar{a}_3v_1)_{k+1}-(\bar{a}_1v_3)_{k+1}\\(\bar{a}_1v_2)_{k+1}+(\bar{a}_2v_2)_{k+1}\end{pmatrix}\right]-\left[\begin{pmatrix}\sigma((v_2)_{k-1}-(v_1)_{k-1})\\\rho(v_1)_{k-1}-(v_2)_{k-1}\\-\beta(v_3)_{k-1}\end{pmatrix}+\begin{pmatrix}0\\-(\bar{a}_3v_1)_{k-1}-(\bar{a}_1v_3)_{k-1}\\(\bar{a}_1v_2)_{k-1}+(\bar{a}_2v_2)_{k-1}\end{pmatrix}\right]\right]$
z_2^k	$L\left[\begin{pmatrix}0\\-(w_3v_1)_{k+1}-(w_1v_3)_{k+1}\\(w_1v_2)_{k+1}+(w_2v_1)_{k+1}\end{pmatrix}-\begin{pmatrix}0\\-(w_3v_1)_{k-1}-(w_1v_3)_{k-1}\\(w_1v_2)_{k-1}+(w_2v_1)_{k-1}\end{pmatrix}\right]$

Table 4.6: Formulas for z_k^l

Our next goal is to compute polynomials $\tilde{Z}_k(r) = \tilde{Z}_1^k r + \tilde{Z}_2^k r^2 \in \mathbb{R}^3$ such that

$$|z_k(r)| \preceq \tilde{Z}_k(r)$$

for $k = 0, \dots, \bar{M}$ and $\tilde{Z}_M(r) = \tilde{Z}_1^M r + \tilde{Z}_2^M r^2 \in \mathbb{R}^3$ such that

$$|z_l^k| \preceq \frac{\tilde{Z}_l^M}{\omega_k^s}$$

for all $k \geq M$. In order to do so we start by computing bounds $B_l^k \in \mathbb{R}^3$ and $K_l^k \in \mathbb{R}^3$ such that

$$\begin{aligned}|\beta_l^k| &\preceq B_l^k \quad k = 0, \dots, M \text{ and } l = 1,2\\ |\kappa_1^k| &\preceq K_1^k \quad k = 0, \dots, m\end{aligned}$$

and $\bar{B}_l^M$ with

$$|\beta_l^k| \preceq \frac{\bar{B}_l^M}{\omega_k^s} \qquad (l = 1,2)$$

for all $k \geq \bar{M}$. Let us start with B_1^k for $k = 1, \dots, M$ and K_1^k for $k = 1, \dots, m$.

Consider therefore the finite sums

$$\begin{aligned}
\left|(\bar{a}_1 v_3^{(I)})_k\right| &\leq \sum_{k_1=-m+1}^{m-1} (|\bar{a}_1|)_{|k_1|} (v_3^{(I)})_{|k-k_1|} \overset{\text{def}}{=} \Sigma_1^{k,(I)} \\
\left|(\bar{a}_2 v_1^{(I)})_k\right| &\leq \sum_{k_1=-m+1}^{m-1} (|\bar{a}_2|)_{|k_1|} (v_1^{(I)})_{|k-k_1|} \overset{\text{def}}{=} \Sigma_2^{k,(I)} \\
\left|(\bar{a}_3 v_1^{(I)})_k\right| &\leq \sum_{k_1=-m+1}^{m-1} (|\bar{a}_3|)_{|k_1|} (v_1^{(I)})_{|k-k_1|} \overset{\text{def}}{=} \Sigma_3^{k,(I)},
\end{aligned}$$

where we recall that $(v_i)_k = \frac{1}{\omega_k^s}$ for $i = 1, 2, 3$. Using this notation we can estimate

$$\begin{aligned}
|\beta_1^k| &\leq L \left[\frac{1}{k^s} \begin{pmatrix} 2\sigma \\ \rho + 1 \\ \beta \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_1^k + \Sigma_3^k \\ \Sigma_1^k + \Sigma_2^k \end{pmatrix} \right] \overset{\text{def}}{=} B_1^k \quad k = 0, \dots, M \\
|\kappa_1^k| &\leq L \begin{pmatrix} 0 \\ \Sigma_1^{k,I} + \Sigma_3^{k,I} \\ \Sigma_1^{k,I} + \Sigma_2^{k,I} \end{pmatrix} \overset{\text{def}}{=} K_1^k \quad k = 0, \dots, m.
\end{aligned}$$

In order to obtain B_1^M let us consider for $k \geq \bar{M}$

$$\begin{aligned}
\left| \sum_{k_1=-m+1}^{m-1} (\bar{a}_1)_{|k_1|} (v_3)_{|k-k_1|} \right| &\leq \sum_{k_1=-m+1}^{m-1} (|\bar{a}_1|)_{|k_1|} \frac{1}{(k-k_1)^s} \\
&\leq \frac{1}{\omega_k^s} \sum_{k_1=-m+1}^{0} (|\bar{a}_1|)_{|k_1|} + \frac{1}{\omega_k^s} \sum_{k_1=1}^{m-1} \frac{(|\bar{a}_1|)_k}{(1 - \frac{k_2}{(M-1)})^s} \\
&= \frac{1}{\omega_k^s} \left((|\bar{a}_1|)_0 + \sum_{k_1=1}^{m-1} (|\bar{a}_1|)_{k_1} \left(1 + \frac{1}{(1 - \frac{k_2}{(M-1)})^s} \right) \right) \\
&\overset{\text{def}}{=} \frac{1}{\omega_k^s} \Sigma_1^{M-1},
\end{aligned}$$

where we use that

$$\frac{1}{1 - \frac{k_1}{k}} \leq \begin{cases} 1 & k_1 \leq 0 \\ \frac{1}{1 - \frac{k_1}{(M-1)}} & k_1 \geq 0 \text{ and } k \geq M - 1 \end{cases}.$$

Similarly we set for $i = 2, 3$

$$\Sigma_i^{M-1} = \left((|\bar{a}_i|)_0 + \sum_{k_1=1}^{m-1} (|\bar{a}_i|)_{k_1} \left(1 + \frac{1}{(1 - \frac{k_2}{(M-1)})^s} \right) \right).$$

So we obtain for $k \geq \bar{M}$

$$|\beta_1^k| \preceq \frac{1}{\omega_k^s} L \left[\begin{pmatrix} 2\sigma \\ \rho+1 \\ \beta \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_1^{M-1} + \Sigma_3^{M-1} \\ \Sigma_1^{M-1} + \Sigma_2^{M-1} \end{pmatrix} \right] \overset{\text{def}}{=} \frac{1}{\omega_k^s} \bar{B}_1^M.$$

In order to compute B_2^k for $k = 0, \ldots, M$ and $\bar{B}_2^M$ we employ estimates whose detailed explanation can be found in [25] and that are summarized in Lemma 2.2.4. Using the constants α_k^2 for $k = 0, \ldots M$ defined in 2.2.4 we obtain that

$$\left| \sum_{k_1+k_{(2)}=k} \frac{1}{\omega_{k_1}\omega_{k_2}} \right| \leq \begin{cases} \alpha_k^2 & k = 0, \ldots, M-1 \\ \frac{\alpha_M^2}{\omega_k^s} & k \geq M-1 \end{cases}.$$

This enables us to estimate for $k = 0, \ldots, M$

$$|\beta_2^k| \preceq L\alpha_k^2 \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \overset{\text{def}}{=} B_2^k$$

and for $k \geq M$

$$|\beta_2^k| \preceq \frac{L\alpha_M^2}{\omega_k^s} \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} \overset{\text{def}}{=} \frac{1}{\omega_k^s} \bar{B}_2^M.$$

We are now ready to estimate:

$$\begin{aligned} |z_l^0| &\preceq \begin{cases} \Big[K_1^0 + \frac{1}{2}K_1^1 + 2\sum_{j=2}^{m-1} \frac{1}{j^2-1} K_1^j + 2\sum_{j=m}^{\bar{M}-1} \frac{1}{j^2-1} B_1^j + & \\ 2\frac{B_1^M}{((M-1)^2-1)(M-2)^{s-1}(s-1)} \Big] & l = 1 \\ \Big[B_2^0 + \frac{1}{2}B_2^1 + 2\sum_{j=2}^{\bar{M}-1} \frac{1}{j^2-1} B_2^j + 2\frac{B_2^M}{((M-1)^2-1)(M-2)^{s-1}(s-1)} \Big] & l = 2 \end{cases} \\ &\overset{\text{def}}{=} \begin{cases} \tilde{Z}_l^0 & l = 1 \\ \tilde{Z}_l^0 & l = 2 \end{cases}. \end{aligned} \tag{4.64}$$

For $k = 1, \ldots, m-1$ we have that

$$\begin{aligned} |z_l^k| &\preceq \begin{cases} (K_l^{k+1} + B_l^{k-1}) & l = 1 \\ (B_l^{k+1} + B_l^{k-1}) & l = 2 \end{cases} \\ &\overset{\text{def}}{=} \begin{cases} \tilde{Z}_l^k & l = 1 \\ \tilde{Z}_l^k & l = 2 \end{cases}. \end{aligned} \tag{4.65}$$

and for $k = m \ldots, \bar{M}$

$$|z_l^k| \preceq (B_l^{k+1} - B_l^{k-1}) \overset{\text{def}}{=} \tilde{Z}_l^k \qquad (l = 1, 2). \tag{4.66}$$

For $k \geq M$ we apply the following reasoning:

$$|z_l^k| \preceq \frac{1}{(k+1)^s} \bar{B}_l^M + \frac{1}{(k-1)^s} \bar{B}_l^M \preceq \frac{1}{\omega_k^s} \left(1 + (1 + \frac{1}{M})^s \right) \bar{B}_l^M \overset{\text{def}}{=} \frac{1}{\omega_k^s} \tilde{Z}_l^M.$$

As a result we obtained polynomial expansions $\tilde{Z}_k(r) = \tilde{Z}_1^k r + \tilde{Z}_2^k r^2$ such that

$$|Df_k(\bar{x} + ru)rv - A^{\dagger}rv| \preceq \tilde{Z}_k(r) \tag{4.67}$$

for all $k = 0, \ldots, \bar{M}$ and $\tilde{Z}_M(r) = \tilde{Z}_1^M r + \tilde{Z}_2^M r^2$ such that

$$|Df_k(\bar{x} + ru)rv - A^{\dagger}rv| \preceq \frac{1}{\omega_k^s} \tilde{Z}_M(r)$$

for all $k \geq M$. A summary is found in Table 4.7.

	$k = 0$
$\tilde{Z}_1^0$	$L \left[\begin{pmatrix} 0 \\ \Sigma_3^{0,I} + \Sigma_1^{0,I} \\ \Sigma_1^{0,I} + \Sigma_2^{0,I} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ \Sigma_3^{1,I} + \Sigma_1^{1,I} \\ \Sigma_1^{1,I} + \Sigma_2^{1,I} \end{pmatrix} + 2 \sum_{j=2}^{m-1} \frac{1}{j^2-1} \begin{pmatrix} 0 \\ \Sigma_3^{j,I} + \Sigma_1^{j,I} \\ \Sigma_3^{j,I} + \Sigma_1^{j,I} \end{pmatrix} \right.$ $\left. + 2 \sum_{j=m}^{M-2} \frac{1}{j^2-1} \left[\frac{1}{\omega_j^s} \begin{pmatrix} 2\lvert\sigma\rvert \\ \lvert\rho\rvert + 1 \\ \lvert\beta\rvert \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_3^j + \Sigma_1^j \\ \Sigma_2^j + \Sigma_2^j \end{pmatrix} \right] + \frac{2}{((M-1)^2-1)(s-1)(M-2)^{s-1}} \left[\begin{pmatrix} 2\lvert\sigma\rvert \\ \lvert\rho\rvert + 1 \\ \lvert\beta\rvert \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_3^{M-1} + \Sigma_1^{M-1} \\ \Sigma_1^{M-1} + \Sigma_2^{M-1} \end{pmatrix} \right] \right]$
$\tilde{Z}_2^0$	$L \left[\frac{2\alpha_0^{2,M}}{\omega_0^s} + \frac{\alpha_1^2}{\omega_1^s} + 2 \sum_{j=2}^{M-1} \frac{2\alpha_j^2}{(j^2-1)\omega_j^s} + \frac{4\alpha_{M-1}^2}{((M-1)^2-1)(s-1)(M-2)^{s-1}} \right] \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$
	$k = 1, \ldots, m-1$
$\tilde{Z}_1^k$	$L \left[\begin{pmatrix} 0 \\ \Sigma_3^{k+1,I} + \Sigma_1^{k+1,I} \\ \Sigma_1^{k+1,I} + \Sigma_2^{k+1,I} \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_3^{k-1,I} + \Sigma_1^{k-1,I} \\ \Sigma_1^{k-1,I} + \Sigma_2^{k-1,I} \end{pmatrix} \right]$
$\tilde{Z}_2^k$	$L \left[\frac{\alpha_{k+1}^2}{\omega_{k+1}^s} + \frac{\alpha_{k-1}^2}{\omega_{k-1}^s} \right] \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$
	$k = m, \ldots, M-1$
$\tilde{Z}_1^k$	$L \left[\frac{1}{\omega_{k+1}^s} \begin{pmatrix} 2\lvert\sigma\rvert \\ \lvert\rho\rvert + 1 \\ \lvert\beta\rvert \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_3^{k+1} + \Sigma_1^{k+1} \\ \Sigma_1^{k+1} + \Sigma_2^{k+1} \end{pmatrix} \right] + L \left[\frac{1}{\omega_{k-1}^s} \begin{pmatrix} 2\lvert\sigma\rvert \\ \lvert\rho\rvert + 1 \\ \lvert\beta\rvert \end{pmatrix} + \begin{pmatrix} 0 \\ \Sigma_3^{k-1} + \Sigma_1^{k-1} \\ \Sigma_1^{k-1} + \Sigma_2^{k-1} \end{pmatrix} \right]$
$\tilde{Z}_2^k$	$L \left[\frac{\alpha_{k+1}^2}{\omega_{k+1}^s} + \frac{\alpha_{k-1}^2}{\omega_{k-1}^s} \right] \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$
	$k = M$
$\tilde{Z}_1^M$	$L \left[(1 + (\frac{M}{M-1})^s) \begin{pmatrix} 2\lvert\sigma\rvert \\ \lvert\rho\rvert + 1 \\ \lvert\beta\rvert \end{pmatrix} + (1 + (\frac{M}{M-1})^s) \begin{pmatrix} 0 \\ \Sigma_3^{M-1} + \Sigma_1^{M-1} \\ \Sigma_1^{M-1} + \Sigma_2^{M-1} \end{pmatrix} \right]$
$\tilde{Z}_2^M$	$L \left[(1 + (\frac{M}{M-1})^s) \alpha_{M-1}^2 \right] \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$

Table 4.7: Formulas for $\tilde{Z}_k^l$

By definition of A and $A^\dagger$ there is a δ such that for all $k \geq 0$

$$|((Id - AA^\dagger)rv)_k \preceq r\delta \in \mathbb{R}^3.$$

We now define for $l = 1,2$ vectors $V_l = (\tilde{Z}_l^0, \dots, \tilde{Z}_l^{m-1}) \in \mathbb{R}^{3m}$ to obtain for $k = 0, \dots, m-1$

$$\begin{aligned} Z_1^k &= (|A_m|V_1)_k + \delta \\ Z_l^k &= (|A_m|V_j)_k \quad l = 2 \end{aligned} \tag{4.68}$$

and for $k = m, \dots, \bar{M}$

$$Z_l^k = \frac{1}{2k}\tilde{Z}_l^k \quad k = m, \dots, \bar{M} \tag{4.69}$$

and hereby have that for all $k = 0, \dots, \bar{M}$

$$|(DT(\bar{x} + rw)rv)_k| \preceq Z_1^k r + Z_2^k r^2.$$

As the right hand side is independent of w, v we can take the supremum over all $w, v \in B_1(0) \subset X^s$ and obtain

$$\sup_{w,v \in B_1(0)} |(DT(\bar{x} + rw)rv)_k| \preceq Z_1^k r + Z_2^k r^2$$

for all $k = 0, \dots, \bar{M}$.

Finally we set

$$Z_l^M = \frac{1}{2M}\tilde{Z}_l^M \tag{4.70}$$

for $l = 1,2,3$. Then defining $Z_M(r) = Z_1^M r + Z_2^M r^2$ we obtain that

$$\sup_{w_1,w_2} |(DT(\bar{x} + rw)rv)_k| \preceq \frac{1}{\omega_k^s} Z_M(r)$$

for all $k \geq M$. This finalizes the construction of the bounds necessary to define the radii polynomials specified in 3.2.2.

Spline approach

We continue with the solution of the above described initial value problems by using the linear spline approach. We first compare the results and then give details on the discretization of the corresponding operator F together with the deduction of the Y and Z bounds used in the validation procedure.

Numerical results and performance comparison For a given initial value p_1 the operator to solve the corresponding initial value problem for the Lorenz equation (4.49) over the integration interval $[0, L]$ is given by

$$F(u)(t) = p_1 + L \int_0^t g(u(s))ds - u(t) = 0 \qquad \forall t \in [0,1]. \tag{4.71}$$

We consider the initial value problems from Theorem 4.2.2 and solve them via validating zeros of (4.71).

Let us start with more detailed discussion of p_1^1. Recall that this is the initial value lying approximately on the unstable manifold of the positive eye equilibrium $(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho-1)$. For an integration time of $L = 0.63$ we depict the numerical orbit in Figure 4.12. Before we discuss rigorous numerical results obtained via our linear spline based algorithm, we offer the following heuristics concerning the performance we can expect.

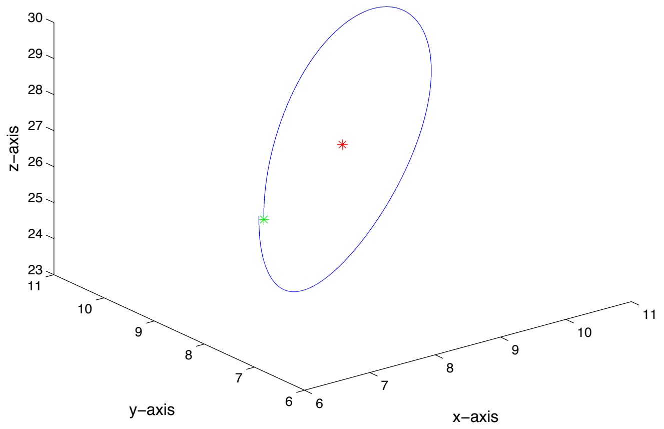


Figure 4.12: Numerical approximate orbit of p_1^1 for $L = 0.63$. p_1^1 is shown in green and the positive eye equilibrium $(\sqrt{\beta(\rho-1)}, \sqrt{\beta(\rho-1)}, \rho-1)$ is depicted in red.

Looking at Figure 4.12 we see a circular shaped motion in phase space caused by the vicinity of p_1^1 to the eye equilibrium that has a complex conjugate pair of unstable eigenvalues. This motivates the following line of thought. Let us for the moment assume we try to approximate a planar circular motion with radius R parametrized over $[0,1]$ by using linear splines, i.e. we wish to approximate

$$\gamma(t) = R \begin{pmatrix} \cos(2\pi t) \\ \sin(2\pi t) \end{pmatrix} \quad t \in [0,1].$$

Looking at this procedure from an elementary perspective, by choosing m equidistant grid points we approximate the circle by a regular m-gon γ_h parametrized over $[0,1]$. For this easy case we can give an explicit lower bound on the C_0 distance between γ and γ_h. In particular we can specify the dependence of this distance on the number of grid points m explicitly. By applying the pythagorean theorem on an arc segment of $\frac{2\pi}{m}$ we get

$$\max_{t\in[0,1]} \|\gamma(t) - \gamma_h(t)\|_\infty \geq \frac{1}{\sqrt{2}} \max_{t\in[0,1]} \|\gamma(t) - \gamma_h(t)\|_2 = \frac{R}{\sqrt{2}} \left(1 - \cos(\frac{\pi}{m})\right).$$

Thus to obtain an accuracy on the order of 10^{-9} as we did for p_1^1 with integration time 1 using the Chebyshev approach we need roughly 50000 grid points. Compare this to 50 modes in the Chebyshev case. Figure 4.13 depicts the evolution of the lower error bound with increasing number of grid points.

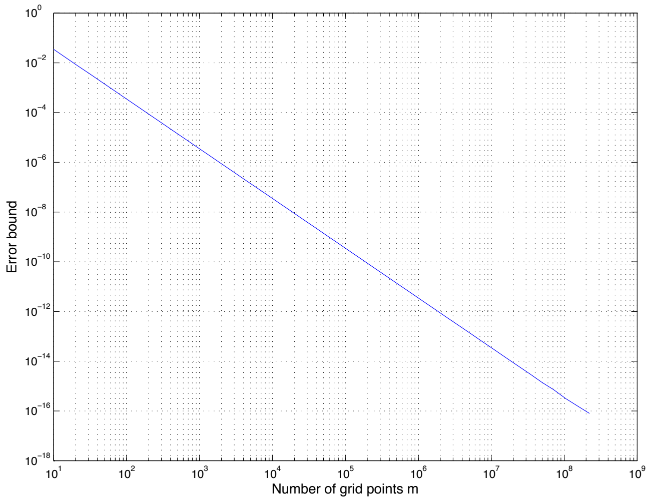


Figure 4.13: Behavior of the lower error bound $\frac{R}{\sqrt{2}}\left(1 - \cos(\frac{\pi}{m})\right)$ in dependence of the number of grid points m.

In Table 4.8 we give some concrete numerical results. We stop at integration time $L = 0.3$ as more than 1200 grid points become computationally challenging. In Figure 4.14 we depict the approximate orbit for

L	0.1	0.15	0.2	0.25	0.3
$m_{p_1^1}$	600	700	1000	1100	1200
$\bar{r}_{p_1^1}$	2.25×10^{-4}	6.83×10^{-4}	5.19×10^{-4}	9.87×10^{-4}	1.95×10^{-3}

Table 4.8: Given p_1^1 and for a fixed L, these are the corresponding number of equidistant grid points $m_{p_1^1}$ and the radius $\bar{r}_{p_1^1}$ around the approximate solution $\bar{u}$ in $C_0(\mathbb{R}, \mathbb{R}^3)$ for which the spline based radii polynomials approach was successful.

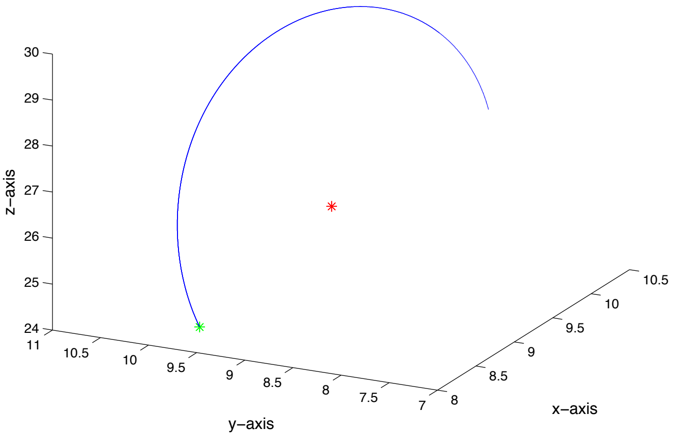


Figure 4.14: Approximate orbit of p_1^1 for integration time $L = 0.3$.

$L = 0.3$. The computations are carried out by *splineIVP1.m* at [28].

Let us go on with considering p_1^2 and p_1^3. We recall that p_1^2 lies on the linear approximation to the unstable manifold of the origin and p_1^3 was chosen at random in the box $[-10, 10] \times [-10, 10] \times [-10, 10]$. Table 4.9 shows the rigorous numerical results and Figure 4.15 shows the approximate orbits for an integration time $L = 0.2$. The computations are carried out by *splineIVPj.m* with $j = 2, 3$ at [28]. In total we see the following points:

- The Chebyshev approach yields higher accuracy for the same integration time.
- We are able to verify solutions over longer integration time intervals by using the Chebyshev approach.
- If we compare the number of grid points in the Spline approach to the dimension of the Galerkin projection in the Chebyshev case, the

L	0.1	0.15	0.2
$m_{p_1^2}$	600	700	1000
$m_{p_1^3}$	600	700	1200
$\bar{r}_{p_1^2}$	9.68×10^{-5}	3.24×10^{-4}	5.05×10^{-4}
$\bar{r}_{p_1^3}$	2.23×10^{-3}	3.27×10^{-2}	1.68×10^{-2}

Table 4.9: Given $p_1^{2,3}$ and for a fixed L, these are the corresponding number of equidistant grid points $m_{p_1^{2,3}}$ and the radius $\bar{r}_{p_1^{2,3}}$ around the approximate solution $\bar{u}$ in $C_0(\mathbb{R}, \mathbb{R}^3)$ for which the spline based radii polynomials approach was successful.

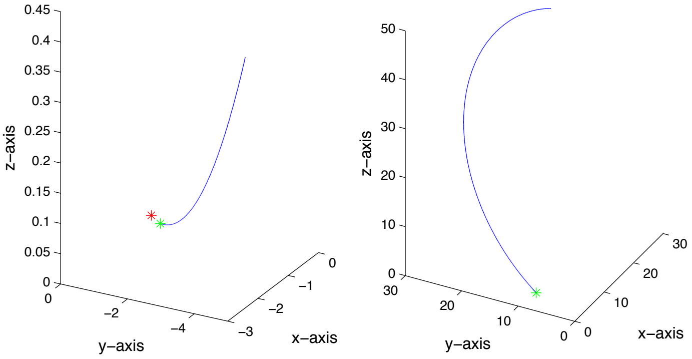


Figure 4.15: Approximate orbit of p_1^2 for integration time $L = 0.2$ on the left hand side and of p_1^3 and the same integration time on the right hand side.

Galerkin dimension is smaller.

Derivation of the Y and Z bounds In order to derive the Y-bounds from (3.38) and (3.39) and the Z-bounds specified in (3.41) and (3.42) necessary to construct the radii polynomials from Definition 3.2.1 we first need to give formulas for F_m and F_∞ from (3.33). We give a dimension independent derivation, i.e. we assume the dimension to be given by a general $d \geq 1$.

Recalling (4.71) and $x = u$ we obtain

$$F_m(x) = (\Pi_m)^d \left(p_1 + L \int_0^t g(u(\tau))d\tau - u(t) \right)$$

and

$$F_\infty(x) = (I - \Pi_m)^d \left(p_1 + L \int_0^t g(u(\tau))d\tau - u(t) \right).$$

More explicitly we define

$$F^m(u_h) = f^m(u_h) - u_h \in \mathbb{R}^{d(m+1)},$$

where $f^m = (f_1^m, \dots, f_{m+1}^m)$ and $f_i^m(u_h) \in \mathbb{R}^d$ is given for $i = 1, \dots, m+1$ component-wise by

$$[f_i^m(u_h)]_j = (p_1)_j + L \int_0^{t_{i-1}} g_j(u_h(\tau))d\tau,$$

where $j = 1, \dots, d$.

We wish to construct the bounds $Y_1, \dots, Y_{m+1}$ and Y_∞ as specified in (3.38) and (3.39) as well as $Z_1, \dots, Z_{m+1}(r)$ and $Z_\infty(r)$ given by (3.41) and (3.42). Assume that the four assumptions (**RP**) of Section 3.2.1 are satisfied, in particular we assume an approximate solution $\bar{u}_h$ such that $\|F^m(\bar{u}_h)\|_{X_m} \leq \epsilon$, for a given *small* $\epsilon > 0$. We can choose $Y_1, \dots, Y_{m+1}$ such that

$$|(A_m F^m(\bar{u}_h))_i| \preceq Y_i \tag{4.72}$$

by evaluating the right hand side rigorously using interval arithmetic. Let us now turn to Y_∞.

Lemma 4.2.1 *Let $\bar{x} = \bar{u}_h$. If we define*

$$Y_\infty \geq \max_{j=1,\dots,d} \max_{i=1,\dots,m} \frac{(t_i - t_{i-1})^2}{8} \sup_{t \in [t_{i-1}, t_i]} \left| \frac{d^2}{dt^2} L \int_0^t g_j(\bar{u}_h(\tau))d\tau \right|$$

then, recalling (3.37), *one has that*

$$\|\Pi_\infty y\|_{X_\infty} = \|F_\infty(\bar{x})\|_{X_\infty} \leq Y_\infty.$$

Proof 4.2.1 *By definition we get*

$$\begin{aligned} \|F_\infty(\bar{x})\|_{X_\infty} &= \|(I - \Pi_m)^d (p_1 + L \int_0^t g(\bar{u}_h(\tau))d\tau - \bar{u}_h)\|_{C^0} \\ &= \|(I - \Pi_m)^d L \int_0^t g(\bar{u}_h(\tau))d\tau\|_{C_0}. \end{aligned}$$

We now apply Theorem 2.6 from [50] to obtain

$$\|F_\infty(\bar{x})\|_{X_\infty} \leq \max_{j=1,\dots,d} \max_{i=1,\dots,m} \frac{(t_i - t_{i-1})^2}{8} \sup_{t \in [t_{i-1}, t_i]} \left| \frac{d^2}{dt^2} L \int_0^t g_j(\bar{u}_h(\tau))d\tau \right|$$

The results follows immediately.

We remark that the quantities involved in Lemma 4.2.1 can be computed rigorously using interval arithmetic.

Next we compute the polynomial bounds $Z_1(r), \dots, Z_{m+1}(r)$ and $Z_\infty(r)$. Recalling (3.40), $z(\xi_1, \xi_2) = DT(\bar{x} + \xi_1)\xi_2$, where T is defined in (3.36) and $\xi_1, \xi_2 \in B(r, \omega)$. In particular we can write $\xi_1 = rw$ and $\xi_2 = rv$, where $w, v \in B(1, \omega)$. Let us start with $Z_1(r), \dots, Z_{m+1}(r)$. In analogy to (3.69) we realize that we have

$$\begin{aligned}\Pi_m z(\xi_1, \xi_2) &= (I - A_m DF_m(\bar{x} + \xi_1))\xi_2 \\ &= (I - A_m DF_m(\bar{x}))(\Pi_m)^d \xi_2 + (A_m DF_m(\bar{x})(\Pi_m)^d \xi_2 - A_m DF_m(\bar{x} + \xi_1)\xi_2).\end{aligned} \tag{4.73}$$

By setting

$$\eta(s) = F_m(\bar{x} + \xi_1 + sv) - F_m(\bar{x} + s(\Pi_m)^d v) \in \mathbb{R}^{d(m+1)},$$

(4.73) can be rewritten as

$$\Pi_m z(\xi_1, \xi_2) = (I - A_m DF_m(\bar{x}))(\Pi_m)^d \xi_2 - A_m \eta'(0) r.$$

Hence we have to estimate $(\eta'(0))_i \in \mathbb{R}^3$ for $i = 1, \dots, m+1$. Choose an index $i \in \{1, \dots, m+1\}$. We obtain

$$\begin{aligned}(\eta(s))_i &= L \int_0^{t_{i-1}} \left[g(\bar{u}_h(\tau) + rw(\tau) + sv(\tau)) - g(\bar{u}_h(\tau) + s(\Pi_m)^d v(\tau)) \right] d\tau \\ &\overset{\text{def}}{=} \beta_i(s).\end{aligned}$$

Also, we have that

$$\begin{aligned}\beta_i'(0) &= \frac{d}{ds} \left[L \int_0^{t_{i-1}} \left[g(\bar{u}_h(\tau) + rw(\tau) + sv(\tau)) - g(\bar{u}_h(\tau) + s(\Pi_m)^d v(\tau)) \right] d\tau \right] \Big|_{s=0} \\ &= L \int_0^{t_{i-1}} \left[Dg(\bar{u}_h(\tau) + rw(\tau)) v(\tau) - Dg(\bar{u}_h(\tau))(\Pi_m)^d v(\tau) \right] d\tau.\end{aligned}$$

We assume for the moment that we can write

$$Dg(\bar{u}_h + rw)v - Dg(\bar{u}_h)(\Pi_m)^d v = \sum_{n=1}^{D} \gamma_n r^{n-1} \tag{4.74}$$

for vector functions $\gamma_n = \gamma_n(\bar{u}_h(\tau), w(\tau), v(\tau)) \in \mathbb{R}^d$ and a suitable $D \in \mathbb{N}$. In case the analytic vector field g is polynomial, D will be the degree of polynomial. We will elaborate on how to compute the expansion (4.74) for the Lorenz equations in the next paragraph. Under this assumption we obtain bounds $\Gamma_n^i \in \mathbb{R}^d$ such that

$$\int_0^{t_{i-1}} |\gamma_n(\tau)| \, d\tau \preceq \Gamma_n^i, \quad \text{for } i = 1, \dots, m+1.$$

The bounds $\Gamma_n^i \in \mathbb{R}^d$ depend on the form of the vector field. We refer to the next paragraph for a specific derivation in the case of the Lorenz equations. Hence, we obtain that

$$\left|\beta_i'(0)\right| \preceq L \sum_{n=1}^{D} \Gamma_n^i r^{n-1}. \tag{4.75}$$

If we define $\tilde{V}_n \in \mathbb{R}^{d(m+1)}$ as

$$\tilde{V}_n = (\Gamma_n^1, \ldots, \Gamma_n^{m+1})$$

we arrive at

$$\left|\eta'(0)r\right| \preceq \sum_{n=1}^{D} \tilde{V}_n r^n.$$

Now using that $\|v\|_\infty \leq 1$, we obtain

$$\begin{aligned} |\Pi_m z(\xi_1,\xi_2)| &\preceq \left|(I - A_m DF^m(\bar{x}))\Pi_m vr - A_m \eta'(0)r\right| \\ &\preceq \|I - A_m DF^m\|_\infty r + \|A_m\|_\infty \sum_{n=1}^{D} \tilde{V}_n r^n. \end{aligned}$$

Define

$$\begin{aligned} V_1 &= \|I - A_m DF^m\|_\infty \mathbf{1}_{d(m+1)} + \|A_m\|_\infty \tilde{V}_1, \\ V_n &= \|A_m\|_\infty \tilde{V}_n, \quad \text{for } n \neq 1, \end{aligned}$$

where $\mathbf{1}_{d(m+1)} = (1, \ldots, 1) \in \mathbb{R}^{d(m+1)}$. Then if we set

$$Z_i(r) = \sum_{n=1}^{D} (V_n)_i r^n \in \mathbb{R}^d$$

for $i = 1, \ldots, m+1$, the inequality

$$\sup_{x_1,x_2 \in B(r,\omega)} |(\Pi_m z(\xi_1,\xi_2))_i| \preceq Z_i(r)$$

is fulfilled. By assumption **RP4.** we can assume that there is a small ϵ_I such that

$$\| \mathrm{I} - A_m DF^m\|_\infty \leq \epsilon_I.$$

The smaller ϵ_I the better is the chance that $V_1 - \mathbf{1_d} \preceq 0$ which a necessary condition in order to find a suitable $\bar{r}$ in order to apply Theorem 3.2.1.

Concerning the bound $Z_\infty(r)$ we have to compute the following:

$$\begin{aligned} &\|\Pi_\infty z(\xi_1,\xi_2)\|_{X_\infty} = \\ &\left\|(I - \Pi_m)^d \left(L \int_0^t Dg(\bar{u}_h(\tau) + rw(\tau))rv(\tau)d\tau\right)\right\|_\infty. \end{aligned}$$

In order to estimate this we use the next result.

Lemma 4.2.2 *Let ξ_1, ξ_2 be specified as above. If we define*

$$Z_\infty(r) \geq \left[\max_{j=1,\ldots,d} \max_{i=1,\ldots,m} \frac{t_i - t_{i-1}}{2} \sup_{t\in[t_{i-1},t_i]} |LDg_j(\bar{u}_h(t) + rw(t))v(t)| \right] r,$$

then $\|\Pi_\infty z(x_1, x_2)\|_{X_\infty} \leq Z_\infty(r)$.

Proof 4.2.2 *By definition we obtain*

$$\begin{aligned}&\|\Pi_\infty z(\xi_1, \xi_2)\|_{X_\infty} = \\ &\|(I - \Pi_m)^d (L \int_0^t Dg(\bar{u}_h(\tau) + rw(s))rv(\tau)d\tau)\|_\infty \leq \\ &\left[\max_{j=1,\ldots,d} \max_{i=1,\ldots,m} \frac{t_i - t_{i-1}}{2} \sup_{t\in[t_{i-1},t_i]} \left|LDg_j(\bar{u}_h(t) + rw(t))v(t)\right| \right] r\end{aligned}$$

where we used a result from [50] for the inequality. The assertion now follows.

This completes the construction of the bounds. We complement the discussion by giving concrete formulas for γ_n and Γ_n^i $(i = 1, \ldots, m+1)$ defined in (4.74) and (4.75) in the context of the Lorenz equations.

Formulas for the Lorenz equations First notice that

$$Dg(x, y, z) = \begin{pmatrix} -\sigma & \sigma & 0 \\ \rho - z & -1 & -x \\ y & x & -\beta \end{pmatrix}. \tag{4.76}$$

Let us start with the computation of the vector functions $\gamma_n(\bar{u}_h, w, v)$ $(n = 1, \ldots, D)$ defined in (4.74). Recall that we seek an expression of the form

$$Dg(\bar{u}_h(s) + rw(s))v(s) - Dg(\bar{u}_h(s))(\Pi_m)^d v(s) = \sum_{n=1}^{D} \gamma_n r^{n-1},$$

where $w, v \in B(1, \omega)$ as defined in (3.35). This in particular implies that

$$\begin{aligned} \|w\|_\infty &\leq 1 + \omega & \|v\|_\infty &\leq 1 + \omega \\ \|w - (\Pi_m)^3 w\|_\infty &\leq \omega & \|v - (\Pi_m)^3 v\|_\infty &\leq \omega. \end{aligned} \tag{4.77}$$

Denoting $\bar{u}_h = ([\bar{u}_h]_1, [\bar{u}_h]_2, [\bar{u}_h]_3)$, $w = (w_1, w_2, w_3)$, $v = (v_1, v_2, v_3)$ and applying (4.76) we can write (4.74) as follows.

$$
\begin{aligned}
&Dg(\bar{u}_h(s) + rw(s))v(s) - Dg(\bar{u}_h(s))(\Pi_m)^3 v(s) \\
&= \begin{pmatrix} -\sigma(v_1 - \Pi_m v_1) + \sigma(v_2 - \Pi_m v_2) \\ \rho(v_1 - \Pi_m v_1) - [\bar{u}_h]_3(v_1 - \Pi_m v_1) - (v_2 - \Pi_m v_2) - [\bar{u}_h]_1(v_3 - \Pi_m[\tilde{x}_2]_3) \\ [\bar{u}_h]_2(v_1 - \Pi_m v_1) + [\bar{u}_h]_1(v_2 - \Pi_m v_2) - \beta(v_3 - \Pi_m v_3) \end{pmatrix} \\
&+ r \begin{pmatrix} 0 \\ w_3 v_1 - w_1 v_3 \\ 2w_2 v_1 \end{pmatrix} \\
&\overset{\text{def}}{=} \gamma_1(\bar{u}_h, w, v) + r\gamma_2(\bar{u}_h, w, v).
\end{aligned}
$$

In particular $D = 2$ in this case. Using (4.77) we can compute $\Gamma^i_{1,2} \in \mathbb{R}^3$ $(i = 1, \ldots, m+1)$. For $i = 1, \ldots, m$

$$
\begin{aligned}
&\int_{t_{i-1}}^{t_i} |\gamma_1(\bar{u}_h(s), w(s), v(s))| ds \preceq \\
&\begin{pmatrix} 2|\sigma| \\ \max_{t \in [t_{i-1}, t_i]} \{|\rho - [\bar{u}_h]_3(t)| + 1 + |[\bar{u}_h]_1(t)|\} \\ \max_{t \in [t_{i-1}, t_i]} \{|[\bar{u}_h]_2(t)| + |[\bar{u}_h]_1(t)| + |\beta|\} \end{pmatrix} \omega(t_i - t_{i-1}) \overset{\text{def}}{=} \tilde{\Gamma}^i_1.
\end{aligned}
\tag{4.78}
$$

Similarly

$$
\int_{t_{i-1}}^{t_i} |\gamma_2(\bar{u}_h(s), w(s), v(s))| ds \preceq \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} (t_i - t_{i-1})(1 + \omega)^2 \overset{\text{def}}{=} \tilde{\Gamma}^i_2. \tag{4.79}
$$

Now set for $n = 1, 2$ and $i = 1, \ldots, m+1$

$$
\Gamma^i_n = \sum_{k=1}^{i-1} \tilde{\Gamma}^k_n.
$$

4.2.3 Connections using the boundary value approach

We finish by presenting an application of the spline based boundary value approach in the context of the generic first order system of the Lorenz equations. We obtain the following result.

Theorem 4.2.3 *Let the parameters $\beta = \frac{8}{3}$ and $\sigma = -2.2$ in the Lorenz system* (4.49) *be fixed. Define the parameter set for ρ by*

$$
U = \{\rho : \rho = 3.2 + k0.01 \ \textit{with} \ k = 0, \ldots, 100\}.
$$

Then there for every $\rho \in U$ exists a tranverse connecting orbit from the origin to the secondary equilibrium q_2.

The proof is based on Theorem 3.2.1. We next describe its ingredients and report more closely on the results.

Formulation of the operator F and discretization Recall that the operator F is defined on $B_1 \times V_{\nu_u} \times C([0,1],\mathbb{R}^d)$ by

$$F(\alpha,\phi,u)(t) = \begin{pmatrix} P(\phi) - \left(Q(\Theta_\nu(\alpha)) + L\int_0^1 g[u(\tau)]\,d\tau\right) \\ Q(\Theta_\nu(\alpha)) + L\int_0^t g[u(\tau)]\,d\tau - u(t) \end{pmatrix}. \tag{4.80}$$

Setting $x = (\alpha,\phi,u)$, we obtain for F_m and F_∞ from (3.33)

$$F_m(x) = \begin{pmatrix} P(\phi) - \left(Q(\Theta_\nu(\alpha)) + L\int_0^1 g[u(\tau)]\,d\tau\right) \\ (\Pi_m)^d\left(Q(\Theta_\nu(\alpha)) + L\int_0^t g[u(\tau)]d\tau - u(t)\right) \end{pmatrix}$$

and

$$F_\infty(x) = \begin{pmatrix} 0 \\ (I-\Pi_m)^d\left(P(\Theta_\nu(\alpha)) + L\int_0^t g[u(\tau)]d\tau - u(t)\right) \end{pmatrix}.$$

More concretely we assume that we are given approximate parametrizations

$$P_N(\phi) = \sum_{k=0}^N a_k^s \phi^k \quad \text{and} \quad Q_M(\varphi) = \sum_{k=0}^M a_k^u \varphi^k.$$

Then we define the operator $F^{m,N,M} : B_1 \times V_{\nu_u} \times \mathbb{R}^{d(m+1)} \to \mathbb{R}^d \times \mathbb{R}^{d(m+1)}$ by

$$F^{m,N,M}(\alpha,\phi,u_h) = \begin{pmatrix} P_N(\phi) - \left(Q_M(\Theta_\nu(\alpha)) + L\int_0^1 g(u_h(\tau))d\tau\right) \\ f^{m,M}(\alpha,u_h) - u_h \end{pmatrix},$$

where m is the number of grid points in the grid Δ and $f^{m,M} = (f_1^{m,M},\ldots,f_{m+1}^{m,M})$ and $f_i^{m,M}(\alpha,u_h) \in \mathbb{R}^d$ is given by

$$f_i^{m,M}(\alpha,u_h) = Q_M(\Theta_\nu(\alpha)) + L\int_0^{t_{i-1}} g(u_h(\tau))d\tau.$$

Then we can use the splitting

$$\begin{aligned} F^m(\alpha,\phi,u_h) &= F^{m,N,M}(\alpha,\phi,u_h) + \begin{pmatrix} (P(\phi)-P_N(\phi)) + (Q_M(\Theta_\nu(\alpha)) - Q(\Theta_\nu(\alpha))) \\ Q_M(\Theta_\nu(\alpha)) - Q(\Theta_\nu(\alpha)) \\ \vdots \\ Q_M(\Theta_\nu(\alpha)) - Q(\Theta_\nu(\alpha)) \end{pmatrix} \\ &= F^{m,N,M}(\alpha,\phi,u_h) + E(\alpha,\phi), \end{aligned}$$

where $E(\alpha, \phi) := F^m(\alpha, \phi, u_h) - F^{m,N,M}(\alpha, \phi, u_h)$ is independent of u_h. In order to derive the Y-bounds from (3.38) and (3.39) and the Z-bounds specified in (3.41) and (3.42) we can treat $F^{m,N,M}$ similar as in Section 4.2.2 and E can be dealt with via the rigorous error analysis together with Cauchy bounds presented in Section 2.1.2 and similar to Section 3.1.2. Rather than giving further details we present some concrete results.

Numerical results and proof of Theorem 4.2.3 We first give some details on the validation for $\rho = 3.2$. Choose $M = 35$, $N = 30$, $\nu_s = 1.75$, $\nu_u = 1.5$, $\beta = 8/3$, $\sigma = -2.2$, and $\rho = 3.2$. For these parameters we again have two dimensional saddles at the equilibria with complex conjugate eigenvalues. We computed validated bounds for the local unstable and stable manifolds of $\delta_u = 1.48 \times 10^{-13}$ and $\delta_s = 2.75 \times 10^{-14}$ and find (by graphical inspection) that the local stable and unstable manifolds do not intersect in phase space. We then define the operator F in (4.80) with $L = 0.5$ and discretize $C_0\left([0,1], \mathbb{R}^3\right)$ using piecewise linear splines with 500 uniformly spaced grid points. We run a classical Newton iteration scheme and obtain an approximate orbit with non-rigorous defect of 9×10^{-16}.

The validation is carried out by using the program *performancerho3p2.m*. We obtain the existence of a unique solution of F about the approximate numerical solution in a 5.11×10^{-5} neighborhood of the approximation. Using the radii polynomials method we also obtain isolation in a neighborhood whose radius is not more than 0.041. Transversality follows as discussed in Section 3.3.2. The proof takes 44 seconds. The results are illustrated in Figure 4.16. (Note we have fixed the phase condition in the stable rather than the unstable parameter space but this makes no difference to the argument). The figure clearly illustrates that the local stable and unstable manifolds do not intersect in phase space, and shows both the spline approximation of the long-connection and the asymptotic orbit segments obtained by applying the linear flow to the boundary points in parameter space. Table 4.10 tabulates performance results for the same proof at $\rho = 3.2$ for several different parameterization orders and grid discretizations.

To proof Theorem 4.2.3 we implemented a simple continuation scheme. This implementation can be found in the program *continuationrho3p2.m*. The results are reported in Table 4.11.

N	M	Grid	δ_u	δ_s	$\bar{r}$	Proof Time
20	25	125	3.09×10^{-11}	4.34×10^{-14}	$[0.00127, 0.04144]$	14 (sec)
20	25	250	3.09×10^{-11}	4.34×10^{-14}	$[0.00023, 0.04145]$	20 (sec)
20	25	500	3.09×10^{-11}	4.34×10^{-14}	$[0.00005, 0.04145]$	36 (sec)
20	25	1000	3.09×10^{-11}	4.34×10^{-14}	$[0.00001, 0.04146]$	1.4 (min)
20	25	2000	3.09×10^{-11}	4.34×10^{-14}	$[2.93 \times 10^{-6}, 0.04146]$	15 (min)

Table 4.10: Performance data for seven proofs of conneciting orbitss for Lorenz with $\beta = 8/3$, $\sigma = -2.2$ and $\rho = 3.2$.

ρ	δ_u	δ_s	$\bar{r}$	Proof Time
3.2	1.48×10^{-13}	2.75×10^{-14}	$[0.00005, 0.04145]$	44 sec
3.3	1.88×10^{-13}	3.33×10^{-14}	$[0.00006, 0.04298]$	45 sec
3.4	3.33×10^{-13}	4.20×10^{-14}	$[0.00006, 0.04101]$	44 sec
3.5	4.81×10^{-13}	5.40×10^{-14}	$[0.00007, 0.03992]$	44 sec
3.6	1.27×10^{-12}	6.52×10^{-14}	$[0.00007, 0.03843]$	44 sec
3.7	5.81×10^{-13}	7.20×10^{-14}	$[0.00007, 0.03286]$	44 sec
3.8	3.73×10^{-11}	1.01×10^{-13}	$[0.00007, 0.02995]$	44 sec
3.9	3.21×10^{-10}	1.21×10^{-13}	$[0.00008, 0.02703]$	44 sec
4.0	3.31×10^{-9}	1.39×10^{-13}	$[0.00008, 0.02395]$	44 sec
4.1	3.86×10^{-8}	2.05×10^{-13}	$[0.00009, 0.02056]$	44 sec
4.2	5.14×10^{-6}	2.52×10^{-13}	$[0.0001, 0.01509]$	44 sec

Table 4.11: Proof of connecting orbits for eight different values of ρ. All manifolds computed to order $M = 25$ and $N = 20$ and spline discretization of 600 grid points.

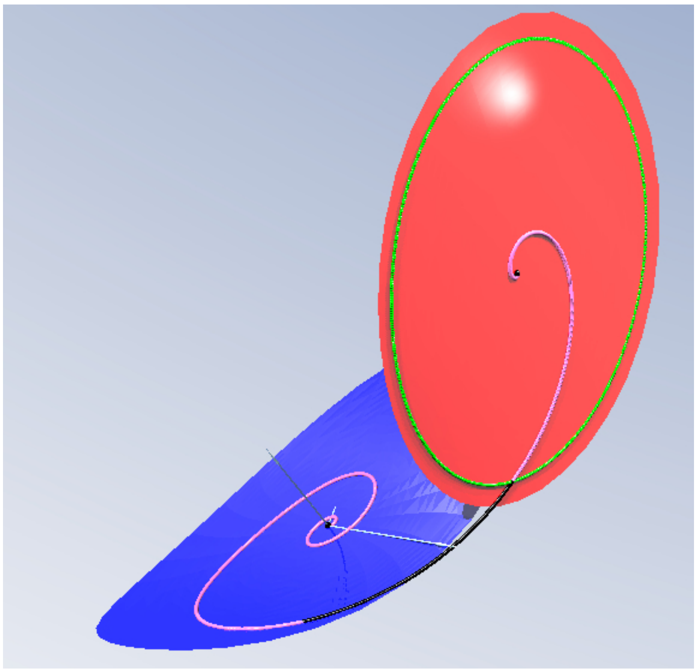


Figure 4.16: Validated *transversal connecting orbit* for Lorenz when $\sigma = -2.2$, $\beta = 8/3$ and $\rho = 3.2$. The image of the phase condition Φ is shown as a green circular arc on the local stable manifold. The zero of 4.80 is shown as a black arc. The pink arcs are the image under the parameterizations of the boundary points in parameter space. The manifolds intersect transversally along the black arc.

Chapter 5

Conclusion

In this thesis we proposed rigorous numerical schemes for the computation of connecting orbits in dynamical systems induced by nonlinear ordinary differential equations. We computed connecting orbits via solving an equivalent zero finding problem $F(x) = 0$.
On the one hand we formulated a finite dimensional equation for generic heteroclinic orbits using a high order approximation of the involved manifolds via the parametrization method, where we used a Newton Kantorovich based approach for the validation and the proof of transversality. We presented some applications in the Lorenz system.
One the other hand we built on the approach from [60] in the spirit of the classical method of projected boundary condition. There the validation is based on the method of radii polynomials and the boundary conditions are formulated via the parametrization method. We considered the following extensions and improvements:

- Formulation of F by using the generic first order formulation of the ODE and obtainment of results about transversality of generic heteroclinics for spline discretization together with applications in the Lorenz system.

- Consideration of an alternative spectral discretization method for the first order formulation based on the Chebyshev expansion together with non-degeneracy results for the derivative of the associated operator. In particular we extended results obtained in [60] on the existence of symmetric homoclinics in the Gray-Scott equations by using our spectral approach.

We finish by giving possible further applications and extensions of this spectral discretization approach.

First, our method could probably be generalized to compute rigorously solutions of higher-order differential equations without re-writing them as first order vector fields. For example, we believe that computing solutions of BVPs associated to the Gray-Scott equations (4.1) could be obtained by integrating twice each equation which could then be solved rigorously by moving to the space of Chebyshev coefficients. The improvement would be twofold. First, the linear part of the equations would grow as $O(k^2)$ (as opposed to $O(k)$ in the BVP-operator (3.50)), hence facilitating the use of a contraction mapping argument based on a Newton-like operator. Second, the size of the finite dimensional projection would be twice smaller. A downside is that we would obtain more complicated formulas of the Chebyshev expansions of the equations resulting from the double integration.

A second extension of the method would be to use a multiple shooting approach to solve the integral operators over long periods of time. Indeed, the theory of the Chebyshev series presented in Section 2.2.1 suggests that integrating over long periods of time (e.g. compute solutions with large scaling factor L) has the disadvantage of bringing the (potentially existing) poles closer to the ρ-ellipse mentioned in Theorem 2.2.2. This implies that the Chebyshev coefficients of the solutions decay to zero at a slow rate. Therefore, an advantage of a multiple shooting approach based on integrating over many short intervals (with corresponding short scaling factor L) would push away the poles, hence bringing a faster decay rate to the Chebyshev coefficients of the solutions. We could then potentially take smaller Galerkin projection dimensions to perform our rigorous computations, thanks to the fast decay rates of the solutions on each sub-intervals. The downside would again be a more complicated formulation of the operators which would need to take care of solving simultaneously many parallel problems.

A third extension of the method would be to combine the rigorous pseudo arc length continuation method of [5, 59] to compute global smooth branches of solutions of BVPs.

A fourth and slightly more challenging improvement consists of modifying the proposed approach to vector fields with nonlinearities that are non polynomial. That would require extending the already existing convolution estimates to the non polynomial case.

A final and most ambitious extension would be to attempt to rigorously compute solutions of spatially periodic PDEs combining a Chebyshev series expansion in time and a Fourier series expansion in space.

Bibliography

[1] Wolf Jürgen Beyn. The numerical computation of connecting orbits in dynamical systems. *IMA Journal of Numerical Analysis*, vol. 10(no. 3):379–405, 1990.

[2] Wolf-Jürgen Beyn. On well-posed problems for connecting orbits in dynamical systems. *Chaotic numerics, eds. Peter E. Kloeden et al. (Contemporary mathematics; 172)*, pages 131–168, 1994.

[3] Wolf-Jürgen Beyn and Jan-Martin Kleinkauf et.al. Numerical analysis of degenerate connecting orbits for maps. *Internat. J. Bifur. Chaos*, (10):3385–3407, 2004.

[4] John P. Boyd. *Chebyshev and Fourier spectral methods*. Dover Publications Inc., Mineola, NY, second edition, 2001.

[5] Maxime Breden, Jean-Philippe Lessard, and Matthieu Vanicat. Global bifurcation diagram of steady states of systems of PDEs via rigorous numerics: a 3-component reaction-diffusion system. *To appear in Acta Applicandae Mathematicae*, 2013.

[6] Nicolas Brisebarre and Mioara Joldeş. Chebyshev interpolation polynomial-based tools for rigorous computing. In *ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, pages 147–154. ACM, New York, 2010.

[7] X. Cabré, E. Fontich, and R. de la Llave. The parameterization method for invariant manifolds. I. Manifolds associated to non-resonant subspaces. *Indiana Univ. Math. J.*, 52(2):283–328, 2003.

[8] X. Cabré, E. Fontich, and R. de la Llave. The parameterization method for invariant manifolds. II. Regularity with respect to parameters. *Indiana Univ. Math. J.*, 52(2):329–360, 2003.

[9] X. Cabré, E. Fontich, and R. de la Llave. The parameterization method for invariant manifolds. III. Overview and applications. *J. Differential Equations*, 218(2):444–515, 2005.

[10] Carmen Chicone. *Ordinary Differential Equations with Applications*. Texts in Applied Mathematics. Springer Verlag, 2nd edition edition, 2006.

[11] C. Conley. *Isolated invariant sets and the Morse index*, volume vol. 38 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1978, iii+89.

[12] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.

[13] S. Day, O. Junge, and K. Mischaikow. A rigorous numerical method for the global analysis of infinite-dimensional discrete dynamical systems. *SIAM J. Appl. Dyn. Syst.*, 3(2):117–160 (electronic), 2004.

[14] Sarah Day, Yasuaki Hiraoka, Konstantin Mischaikow, and Toshi Ogawa. Rigorous numerics for global dynamics: a study of the Swift-Hohenberg equation. *SIAM J. Appl. Dyn. Syst.*, 4(1):1–31 (electronic), 2005.

[15] Sarah Day, Jean-Philippe Lessard, and Konstantin Mischaikow. Validated continuation for equilibria of PDEs. *SIAM J. Numer. Anal.*, 45(4):1398–1424 (electronic), 2007.

[16] M. Dellnitz and A. Hohmann. The computation of unstable manifolds using subdivision and continuation. *Nonlinear Dynamical Systems and Chaos*, PNLDE 19:449–459, 1996.

[17] M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable mani- folds and global attractors. *Numerische Mathematik*, 75:293–317, 1997.

[18] A. Dhooge, W. Govaerts, Yu. A. Kuznetsov, A Mestrom, A.M. Riet, and B. Sautois. *MatCont and CL$_M$atCont: Continuation Toolbox in* MATLAB, December 2006.

[19] E.J. Doedel and M.J. Friedman. Numerical computation of heteroclinic orbits, continuation techniques and bifurcation problems. *Journal of Computational and Applied Mathematics*, 26(no. 1-2):155–170, 1989.

[20] Gregory E. Fasshauer. *Meshfree approximation methods with Matlab*, volume 6 of *Interdisciplinary Mathematical Sciences*. World Scientific Publishing Co. Pte. Ltd., 2007.

[21] Robert Franzosa and Konstantin Mischaikow. Algebraic transition matrices in the Conley index theory. *Trans. Amer. Math. Soc.*, 350(3):889–912, 1998.

[22] E.J. Doedel M.J. Friedman and B.I. Kunin. Successive continuation for locating connecting orbits. *Numerical Algorithms*, vol. 14(no. 1-3):103–124, 1997.

[23] M.J. Friedman and E.J. Doedel. Numerical computation and continuation of invariant manifolds connecting fixed points. *SIAM Journal on Numerical Analysis*, vol. 28(no. 3):789–808, 1991.

[24] Marcio Gameiro and Jean-Philippe Lessard. Analytic estimates and rigorous continuation for equilibria of higher-dimensional PDEs. *J. Differential Equations*, 249(9):2237–2268, 2010.

[25] Marcio Gameiro and Jean-Philippe Lessard. Rigorous numerics for high-dimensional PDEs via one dimensional estimates. To appear in *SIAM J. Numer. Anal.*, 2013.

[26] Peter Grindrod. *Patterns and waves : the theory and applications of reaction-diffusion equations*. Claredon Press, 1991.

[27] J. K. Hale, L. A. Peletier, and W. C. Troy. Exact homoclinic and heteroclinic solutions of the Gray-Scott model for autocatalysis. *SIAM J. Appl. Math.*, 61(1):102–130 (electronic), 2000.

[28] Christian Reinhardt J.P. Lessard, J. Mireles-James. *Thesis Codes*, 2013. `http://www-m3.ma.tum.de/Allgemeines/ChristianReinhardt`.

[29] M. Dellnitz O. Junge and B. Thiere. The numerical detection of connecting orbits. *Discrete and continuous dynamical systems–Series B*, 1(1), 2001.

[30] Oliver Junge et al. *GAIO- Global Analysis of Invariant Objects*. `http://www-m3.ma.tum.de/Allgemeines/OliverJungeSoftware`.

[31] H.B. Keller. *Lectures on numerical methods in bifurcation problems*, 1986.

[32] J. Knobloch and T. Rieß. Lin's method for heteroclinic chains involving periodic orbits. *Nonlinearity*, vol. 23 (2010)(no. 1):23–54.

[33] B.A. Coomes H. Koçak and K.J. Palmer. Transversal connecting orbits from shadowing. *Numerische Mathematik*, vol. 106(no. 3):427–469, 2007.

[34] B. Coomes H. Koçak and K. Palmer. Homoclinic shadowing. *J. Dynam. Differential Equations*, (17):175–215, 2005.

[35] B. Krauskopf and T. Rieß. A Lin's method approach to finding and continuing heteroclinic connections involving periodic orbits. *Nonlinearity*, vol. 21 (2008)(no. 8):1655–1690.

[36] Sil'nikov L.P. On a poincaré-birkhoff problem. *Math.USSR-Sb.*, (3):353–371, 1967.

[37] C. McCord and K. Mischaikow. Connected simple systems, transition matrices, and heteroclinic bifurcations. *Transactions of the American Mathematical Society*, vol. 333 (1992)(no. 1):397–422.

[38] J.D. Mireles-James and K. Mischaikow. Rigorous a posteriori computation of (un)stable manifolds and connecting orbits for analytic maps. *To appear in SIADS*, 2013.

[39] S. Day Y. Hiraoka K. Mischaikow and T. Ogawa. Rigorous numerics for global dynamics: a study of the Swift-Hohenberg equation. *SIAM Journal on Applied Dynamical Systems*, vol. 4(no. 1):1–31 (electronic), 2005.

[40] Marian Mrozek and Piotr Zgliczynski et al. *CAPD- Computer assisted proofs in dynamics*. `http://capd.ii.uj.edu.pl/index.php`.

[41] J.D. Murray. *Mathematical Biology II.* Springer, 1993.

[42] M.T. Nakao. Numerical verification methods for solutions of ordinary and partial differential equations. *Numerical Functional Analysis and Optimization. An International Journal*, vol. 22 (2001)(no. 3-4):321–356.

[43] J.I. Neimark and L.P. Silnikov. A condition for the generation of periodic motions. *Doklady Akademii Nauk SSSR*, vol. 160 (1965):1261–1264.

[44] S. Oishi. Numerical verification method of existence of connecting orbits for continuous dynamical systems. *The Journal of Universal Computer Science*, vol. 4 (1998), no. 2:193–201 (electronic).

[45] J.M. Ortega. The newton-kantorovich theorem. *Amer. Math. Monthly*, 75:658–660, 1968.

[46] Clark Robinson. *Dynamical Systems: Stability, Symbolic Dynamics and Chaos*. CRC Press, second edition edition, 1999.

[47] S. Rump. Verification methods: rigorous results using floating-point arithmetic. *Acta Numer.*, 19:287–449, 2010.

[48] S.M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999. http://www.ti3.tu-harburg.de/rump/.

[49] E Doedel RC Paffenroth AR Champneys TF Fairgrieve YA Kuznetsov BE Oldeman B Sandstede and X Wang. Auto2000: Continuation and bifurcation software for ordinary differential equations (with homcont). Technical report, Concordia University, 2002.

[50] M.H. Schultz. *Spline Analysis*. Prentice Hall, 1973.

[51] Joel Smoller. *Shock Waves and Reaction Diffusion Equations*. Springer, 1994.

[52] Colin Sparow. *The Lorenz Equations: Bifurcations, Chaos and Strange Attractors*, volume 9 of *Applied Mathematical Sciences*. Springer Verlag, New York, 1982.

[53] Gilbert Strang. *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007.

[54] J.B. Swift and P.C. Hohenberg. Hydrodynamic fluctuations at the convective instability. *Phys. Rev. A*, 15(1), 1977.

[55] L. N. Trefethen et al. *Chebfun Version 4.2*. The Chebfun Development Team, 2011. `http://www.maths.ox.ac.uk/chebfun/`.

[56] Lloyd N. Trefethen. *Spectral methods in MATLAB*, volume 10 of *Software, Environments, and Tools*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[57] Lloyd N. Trefethen. *Approximation theory and approximation practice*. Society for Industrial and Applied Mathematics (SIAM), 2012.

[58] Jan Bouwe van den Berg and Jean-Philippe Lessard. Chaotic braided solutions via rigorous numerics: chaos in the Swift-Hohenberg equation. *SIAM J. Appl. Dyn. Syst.*, 7(3):988–1031, 2008.

[59] Jan Bouwe van den Berg, Jean-Philippe Lessard, and Konstantin Mischaikow. Global smooth solution curves using rigorous branch following. *Math. Comp.*, 79(271):1565–1584, 2010.

[60] Jan Bouwe van den Berg, Jason D. Mireles-James, Jean-Philippe Lessard, and Konstantin Mischaikow. Rigorous numerics for symmetric connecting orbits: even homoclinics of the Gray-Scott equation. *SIAM J. Math. Anal.*, 43(4):1557–1594, 2011.

[61] James S. Walker. *A primer on wavelets and their scientific applications*. CRC Press, second edition edition, 2008.

[62] D. Wilczak. Symmetric heteroclinic connections in the Michelson system: a computer assisted proof. *SIAM Journal on Applied Dynamical Systems*, vol. 4 (2005)(no. 3):489–514.

[63] D. Wilczak. Abundances of heteroclinic and homoclinic orbits for the hyperchaotic rössler system. *Discrete Contin. Dyn. Syst. Ser. B 11*, (no. 4):1039–1055, 2009.

[64] D. Wilczak and P. Zgliczyński. Heteroclinic connections between periodic orbits in planar restricted circular three body problem. II. *Communications in Mathematical Physics*, vol. 259, no. 3:561–576, 2005.

[65] Daniel Wilczak and Piotr Zgliczynski. Heteroclinic connections between periodic orbits in planar restricted circular three-body problem—a computer assisted proof. *Comm. Math. Phys.*, 234(1):37–75, 2003.

[66] N. Yamamoto. A numerical verification method for solutions of boundary value problems with local uniqueness by Banach's fixed-point theorem. *SIAM Journal on Numerical Analysis*, vol. 35, no. 5:2004–2013 (electronic), 1998.

[67] Piotr Zgliczyński and Konstantin Mischaikow. Rigorous numerics for partial differential equations: the Kuramoto-Sivashinsky equation. *Found. Comput. Math.*, 1(3):255–288, 2001.