The Lorenz attractor exists
- Revised March 10, 1999 -

Warwick Tucker
Department of Mathematics
Uppsala University
S-751 06 Uppsala
Sweden
warwick@math.uu.se

September 24, 1998

Abstract

We prove that the Lorenz equations support a strange attractor, as conjectured by Edward Lorenz in 1963. We also prove that the attractor is robust, i.e., it persists under small perturbations of the coefficients in the underlying differential equations. Furthermore, the flow admits a unique SRB measure, whose support coincides with the attractor. The proof is based on a combination of normal form theory and rigorous numerics.

Author's note to the revised version

Revision December 8, 1998

The first version (September 24, 1998) of my Ph.D. thesis turned out to contain a mistake in the code dealing with the cone field. After isolating the error, I decided to use a different choice for the field. In the earlier version, the field was taken to be horizontally centered and very wide. The new version uses slimmer cones centered along an approximation of the attractor. This requires some additional functions in the code which have now been added. The main bulk of the code and its underlying mathematics, however, is unchanged. I have also added a statement concerning the existence of a unique SRB measure for the flow.

Revision March 10, 1999

Yet another error in the code was detected. This time it was the expansion estimates that were affected. The faulty algorithm has been corrected, and some global variables have been eliminated. The main bulk of the code and its underlying mathematics is still unchanged, although the code is somewhat more structured now.

Warwick Tucker
Rio de Janeiro
March 10, 1999

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Lennart Carleson for sharing his great depth of knowledge with me. Not only did he give me a great problem to work on (although I had my doubts the first few years), he also introduced me to many of his colleagues who all share the same willingness to discuss mathematics. He has been an excellent supervisor, and I consider myself very fortunate having been one of his students.

The Dynamical Systems group at KTH has been a great source of knowledge and learning for me. I thank all members for the high quality seminars and discussions throughout the years. One member who has meant especially much to my work is Håkan Eliasson, whom I would like to thank for teaching me everything I know about normal forms. Without his outstanding tutorials and down-to-earth approach, I would still be trapped in a small neighbourhood of the origin. I would also like to thank Jürgen Moser who suggested that I should look at the small neighbourhood in the first place.

Both Jacob Palis and Marcelo Viana have made my visits to Rio de Janeiro fruitful and pleasant. The stimulating working atmosphere provided at IMPA has undoubtedly contributed to my development as a mathematician. In particular, the discussions with Marcelo Viana have been invaluable to me, and really helped to clarify so many concepts. I also want to thank Lorenzo Diaz for inviting me to PUC, and for teaching me the essentials[1] of Portuguese - Falou, mermão!

Stefano Luzzatto, whom I first met in Paris, has provided me with invaluable guidance and support for which I thank him. Over the last couple of years, we have become close friends and together we have joyfully experienced some rather unstable events. I would like to take this opportunity to express how much I appreciate working together with him on our joint project.

I am very grateful to the entire staff at the Department of Mathematics of Uppsala University, my home department. As a young and inexperienced Ph.D. student, I found the atmosphere and people there extremely friendly (I still do). Matts Essén deserves a special mention for making it possible for me to participate in several summer schools in Jyväskylä, which were not only outrageous fun, but also mathematically rewarding. Also, I would like to thank my Ph.D. colleagues - and very dear friends - Jonas Avelin, Tobias Ekholm, Jonas Hansbo, Jerk Matero, Björn Selander, and Anders Öberg for making life, mathematically and otherwise, so much more fun.

Finally, I would like to sincerely thank my parents for always believing in me, and for their encouragement and constant support.

Oh yes, a special thanks also to Santa Claus for giving me a pen *identical* to the one that Anders, my stepfather, wrote his thesis in mathematics with. Too bad that he seems to have lost his one...

[1] According to the book *How to be a Carioca* by Priscilla Ann Goslin, no-one will suspect that you are a *gringo* if you stick to the following four phrases: (1) *E aí, mermão, beleza?* (2) *O negócio é o seguinte...* (3) *Num dá pra dá um jeitinho?* (4) *uma estupidamente gelada.* The funny thing is that it actually works!

Contents

1 Introduction

1.1 Chaos theory for pedestrians

In this section we will outline some basic definitions used in the theory of dynamical systems. The presentation is based mainly on [9], [11], and [16].

1.1.1 Hyperbolicity

Consider a C^k diffeomorphism ($k \geq 1$) of a compact manifold to itself, $f\colon M \to M$. The *forward orbit* of a point $p \in M$ under f is the set $\{f^i(p)\}_{i=0}^{\infty}$, where f^i is f composed with itself i times. A point p is a *periodic point of period* k provided $f^k(p) = p$ and $f^i(p) \neq p$ for $0 < i < k$. If p has period one, we call it a *fixed point* of f. We say that p is a *hyperbolic fixed point* of f if $f(p) = p$ and if Df_p has no eigenvalues on the unit circle. According to standard results in spectral theory, there then exists a splitting of the tangent space $T_pM = \mathbb{E}_p^s \oplus \mathbb{E}_p^u$, where the invariant subspaces $\mathbb{E}_p^s$ and $\mathbb{E}_p^u$ correspond to the spectrum inside and outside the unit circle, respectively. This means that we can find constants $0 < \lambda < 1$ and $C > 1$ such that for all $n \in \mathbb{N}$,

$$\|Df_p^{-n}|\mathbb{E}_p^u\| \leq C\lambda^n \quad \text{and} \quad \|Df_p^n|\mathbb{E}_p^s\| \leq C\lambda^n$$

for some Riemannian norm $\|\cdot\|$ on T_pM. The subspaces $\mathbb{E}_p^s$ and $\mathbb{E}_p^u$ are called the *stable* and *unstable subspaces* for the fixed point p. Given a hyperbolic fixed point p, we define its stable and unstable manifolds:

$$W^s(p) = \{x \in M\colon \lim_{i\to\infty} f^i(x) = p\},$$
$$W^u(p) = \{x \in M\colon \lim_{i\to\infty} f^{-i}(x) = p\}.$$

These sets are injectively immersed C^k submanifolds of M, and have the same dimensions as their corresponding linear subspaces.

We can extend these definitions to the case when p is a periodic point of period k simply by replacing f by f^k. Also, we can extend the notion of a hyperbolic fixed point to a whole set.

Consider a compact set $\Lambda \subset M$ which is invariant under f, i.e., $f(\Lambda) = \Lambda$. We say that Λ is a *hyperbolic set* for f if there exists a splitting $T_xM = \mathbb{E}_x^s \oplus \mathbb{E}_x^u$ for each $x \in \Lambda$, such that

1. $\mathbb{E}_x^s$ and $\mathbb{E}_x^u$ vary continuously with x,
2. the splitting is invariant, i.e., $Df_x \cdot \mathbb{E}_x^s = \mathbb{E}_{f(x)}^s$ and $Df_x \cdot \mathbb{E}_x^u = \mathbb{E}_{f(x)}^u$,
3. there are constants $0 < \lambda < 1$ and $C > 1$ such that for all $n \in \mathbb{N}$,

$$\|Df_x^n|\mathbb{E}_x^s\| \leq C\lambda^n \quad \text{and} \quad \|Df_x^{-n}|\mathbb{E}_x^u\| \leq C\lambda^n.$$

1.1.2 Robustness

In practice, it is impossible to explicitly find the invariant set Λ, not to mention the splitting $T_\Lambda M = \mathbb{E}^s \oplus \mathbb{E}^u$, except in the most trivial cases. Fortunately, we shall soon see that hyperbolicity is a robust property, and one can thus make do with pretty crude approximations of both Λ and the subbundles of the splitting. By *robust*, we mean that the defining hypotheses are open in the C^1 topology.

A compact region $N \subset M$ is called a *trapping region* for f provided $f(N) \subset \operatorname{int}(N)$. Given such a set, we can construct the *maximal invariant set* of N:

$$\Lambda = \bigcap_{i=0}^{\infty} f^i(N).$$

It is clear that any other invariant set in N must be a proper subset of Λ. Seeing that the sequence $\{f^i(N)\}_{i=0}^{\infty}$ is nested, we can approximate Λ by considering high iterates of N. Any property valid in an open neighbourhood of Λ will then also hold for $f^k(N)$ if we take k sufficiently large. Seeing that the image of a trapping region also is a trapping region, we may assume that $k = 1$ by taking N close to Λ.

Let $T_N M = \mathbb{F}^s \oplus \mathbb{F}^u$ be a continuous splitting approximating $\mathbb{E}^s \oplus \mathbb{E}^u$. Given $\alpha \geq 0$ we define the *stable* and *unstable cone fields*

$$\mathfrak{C}_x^s(\alpha) = \{v_1 + v_2 \in \mathbb{F}_x^s \oplus \mathbb{F}_x^u \colon |v_2| \leq \alpha |v_1|\},$$
$$\mathfrak{C}_x^u(\alpha) = \{v_1 + v_2 \in \mathbb{F}_x^s \oplus \mathbb{F}_x^u \colon |v_2| \geq \alpha |v_1|\}.$$

The following theorem provides a practical way of proving that a set is hyperbolic:

Theorem 1.1 *Let N be a trapping region for a C^1 diffeomorphism f. Suppose that there exists a continuous splitting $T_N M = \mathbb{F}^s \oplus \mathbb{F}^u$, and that there are constants $\alpha \geq 0$, $C > 0$, and $\sigma > 1$ so that*

$$Df_x^{-1} \cdot \mathfrak{C}_x^s(\alpha) \subset \mathfrak{C}_{f^{-1}(x)}^s(\alpha) \quad \text{and} \quad Df_x \cdot \mathfrak{C}_x^u(\alpha) \subset \mathfrak{C}_{f(x)}^u(\alpha)$$

and

$$\|Df_x^{-n} | \mathfrak{C}_x^s(\alpha)\| \geq C\sigma^n \quad \text{and} \quad \|Df_x^{n} | \mathfrak{C}_x^u(\alpha)\| \geq C\sigma^n$$

for every $x \in N$. Then $\Lambda = \cap_{i=0}^{\infty} f^i(N)$ is hyperbolic for f.

It is clear that the hypotheses of this theorem are open in the C^1 topology, which proves that hyperbolicity is a robust property. In particular, if g is C^1 close to f, then $\Lambda_g = \cap_{i=0}^{\infty} g^i(N)$ is hyperbolic for g.

1.1.3 Strange attractors

We will fix the following notation[2]: a compact, invariant set Λ_f is called *attracting* if there exists an open neighbourhood U of Λ_f such that $\cap_{i=0}^{\infty} f^i(U) = \Lambda_f$. The largest such U is called the basin of attraction for Λ_f, and is denoted $B(\Lambda_f)$. In particular, the maximal invariant set of any trapping region is an attracting set. Even so, it may be the case that most points in $B(\Lambda_f)$ tend to a much smaller subset of Λ_f. As an example, consider the diffeomorphism with a phase portrait as illustrated in Figure 1. Although the whole interval I between the two filled fixed points is

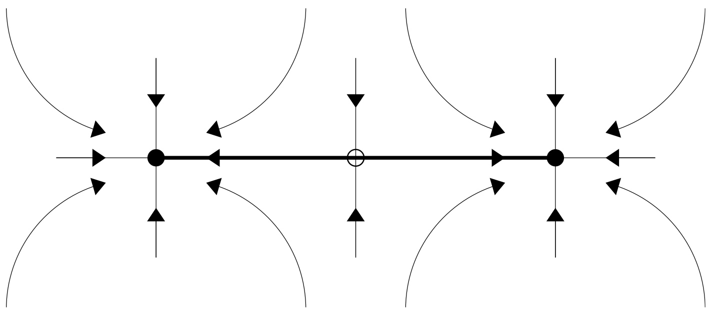

Figure 1: An attracting set which is not an attractor.

attracting with $B(I) = \mathbb{R}^2$, it is clear that most orbits tend to either one of the extreme points of I. In order to avoid this kind of situation we restrict our attention to a subset of the attracting sets. An *attractor* is an attracting set which contains a dense orbit: $\Lambda_f = \overline{\cup_{i=0}^{\infty} f^i(x)}$ for some $x \in \Lambda_f$. This means that Λ_f is minimal in the sense that no proper subset of Λ_f is attracting. Clearly,

[2]The reader should be aware of that there are several different notions of a strange attractor, see [8]. We choose to use very strong (but natural) requirements seeing that the Lorenz attractor satisfies almost all existing definitions of a strange attractor.

the attracting set I in our example is not an attractor whereas the two extreme fixed points are. There is, however, nothing "chaotic" about the asymptotic behaviour of points tending to these attractors, and the situation is therefore dynamically uninteresting.

From this point of view we would like to be able to distinguish attractors exhibiting interesting dynamical properties from those who do not. For this purpose, an attractor is called *strange* if for almost all pairs of different points in $B(\Lambda_f)$, their forward orbits eventually separate by at least a constant δ (only depending on Λ_f). Here, almost all pairs means with probability one in $B(\Lambda_f)\times B(\Lambda_f)$ with respect to Lebesgue measure. These attractors are sometimes called *chaotic* or *sensitive* seeing that, no matter how accurately we measure the initial conditions, we will eventually accumulate an error of size δ.

Sometimes, we can also say something about the speed at which nearby orbits separate. Indeed, if an attractor Λ_f is hyperbolic with a non-trivial unstable tangent bundle, we clearly have exponential divergence of almost all nearby orbits. Such an attractor is called *non-trivial hyperbolic* and, apart from being strange, it is also robust.

1.1.4 Flows and their return maps

We will now describe a useful relation between discrete-time (maps) and continuous-time (flows) dynamical systems. Consider the system of ordinary differential equations

$$\dot{x} = v(x), \tag{1}$$

where $x \in M$ and where the vector field v is a C^r function, $v\colon M \to TM$. Let $\varphi(x,t)$ denote the flow of (1), i.e.,

$$\tfrac{d}{dt}\varphi(x,t) = v(\varphi(x,t)),$$

and suppose that the system (1) has a periodic solution of period $T > 0$, containing the point x_0, i.e., $\varphi(x_0, T+t) = \varphi(x_0,t)$ for all $t \in \mathbb{R}$. Suppose that $\dim(M) = n$, and let Σ be an $(n-1)$-dimensional surface transversal to the vector field at x_0, see Figure 2. By this, we mean that $\langle v(x_0)\rangle \oplus T_{x_0}\Sigma = T_{x_0}M$. Then we can find an open set $U \subset \Sigma$ containing x_0 such that for all $x \in U$, there exists a $\tau(x)$ close to T such that $\varphi(x,\tau(x)) \in \Sigma$.

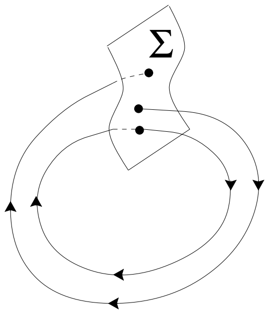

Figure 2: The surface Σ and two trajectories.

The point $\varphi(x,\tau(x))$ is called the *first return* of x, and the map R which associates a point with its first return is called the *return map*: $R(x) = \varphi(x,\tau(x))$. Note that, by construction, we have $\tau(x_0) = T$ and $R(x_0) = x_0$. Thus a fixed point of R corresponds to a periodic orbit of (1), and a periodic point of period k corresponds to a periodic orbit of (1) piercing Σ k times before closing. The following theorem states that the return map is as smooth as the vector field:

Theorem 1.2 *Under the above conditions, and for sufficiently small U, the return map is a C^r diffeomorphism of U onto a subset of Σ.*

This means that the partial derivatives of R are well defined. Once we have the return map and its derivative, we can employ the machinery described in the previous sections: we say that a periodic orbit of (1), γ, is hyperbolic if any member in $\Sigma \cap \gamma$ is a hyperbolic periodic point for return map R. Likewise, we say that a set $\mathcal{A}$, which is flow-invariant, i.e.,

$$\bigcap_{t \geq 0} \varphi(\mathcal{A}, t) = \mathcal{A}$$

is hyperbolic if its intersection with Σ is a hyperbolic set for the R. All definitions concerning attractors can be carried over to flows by substituting f^i, $i \in \mathbb{N}$ for $\varphi(\cdot, t)$, $t \geq 0$.

1.2 Background to the problem

The following non-linear system of differential equations,

$$\begin{aligned} \dot{x}_1 &= -\sigma x_1 + \sigma x_2 \\ \dot{x}_2 &= \varrho x_1 - x_2 - x_1 x_3 \\ \dot{x}_3 &= -\beta x_3 + x_1 x_2, \end{aligned} \tag{2}$$

was introduced 1963 by Edward Lorenz, see [5]. As a crude model of atmospheric dynamics, these equations led Lorenz to the discovery of sensitive dependence of initial conditions - an essential factor of unpredictability in many systems, e.g., meteorology. Numerical simulations for an open neighbourhood of the classical parameter values $\sigma = 10, \beta = 8/3$ and $\varrho = 28$ suggest that almost all points in phase space tend to a strange attractor - *the Lorenz attractor*.

We first note that the system (2) is invariant under the transformation $S(x_1, x_2, x_3) = (-x_1, -x_2, x_3)$. This means that any trajectory that is not itself invariant under S must have a symmetric "twin trajectory".

For $\varrho > 1$, there are three fixed points: the origin and the two "twin points"

$$C^{\pm} = (\pm\sqrt{\beta(\varrho - 1)}, \pm\sqrt{\beta(\varrho - 1)}, \varrho - 1).$$

For the parameter values we are considering, $C^{\pm}$ have a pair of complex eigenvalues with positive real part, and one real, negative eigenvalue. The origin is a saddle point with two negative and one positive eigenvalue satisfying

$$0 < -\lambda_3 < \lambda_1 < -\lambda_2.$$

Thus, the stable manifold of the origin $W^s(0)$ is two-dimensional, and the unstable manifold of the origin $W^u(0)$ is one-dimensional.

It is also worth mentioning that the flow contracts volumes at a significant rate. As the divergence of the vector field is given by

$$\frac{\partial \dot{x}_1}{\partial x_1} + \frac{\partial \dot{x}_2}{\partial x_2} + \frac{\partial \dot{x}_3}{\partial x_3} = -(\sigma + \beta + 1),$$

we see that the volume of a solid at time t can be expressed as

$$V(t) = V(0)e^{-(\sigma+\beta+1)t} \approx V(0)e^{-13.7t},$$

for the classical parameter values. This means that the flow contracts volumes almost by a factor *one million* per time unit, which is quite extreme.

There appears to exist a forward invariant open set U containing the origin but bounded away from $C^{\pm}$. The set U is a double torus (one with two holes), with its holes centered around the two excluded fixed points. If we let φ denote the flow of (2), we can form the maximal invariant set

$$\mathcal{A} = \bigcap_{t \geq 0} \varphi(U, t).$$

Due to the flow being dissipative, the attracting set $\mathcal{A}$ must have zero volume. It must also contain the unstable manifold of the origin $W^u(0)$, which seems to spiral around $C^{\pm}$ in a very complicated, non-periodic fashion, see Figure 3. In particular, $\mathcal{A}$ contains the origin itself, and therefore the flow on $\mathcal{A}$ can not have a hyperbolic structure. The reason is that fixed points of the vector field generate discontinuities for the return maps, and as a consequence, the hyperbolic splitting is not continuous. Apart from this, the attracting set appears to have a strong hyperbolic structure as described below.

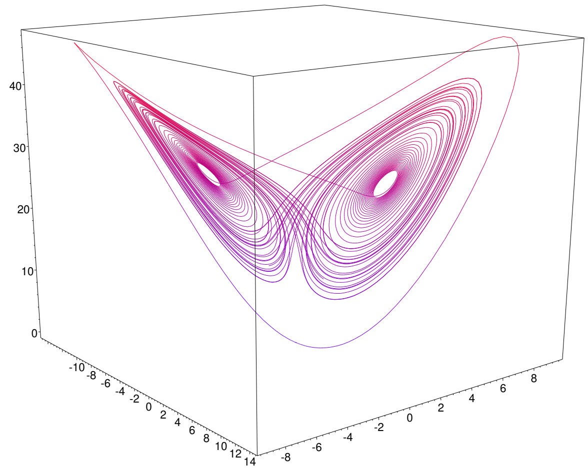

Figure 3: A part of the unstable manifold of the origin.

As it was very difficult to extract rigorous information about the attracting set $\mathcal{A}$ from the differential equations themselves, a *geometric model* of the Lorenz flow was introduced by John Guckenheimer and Robert Williams in the mid-seventies, see [1]. This model has been extensively studied, and it is well understood today, see e.g. [2], [17], [15], [10], [12], [13]. Oddly enough, the original equations introduced by Lorenz have remained a puzzle. A few computer-assisted proofs, however, have recently been announced, see [3] and [7]. These articles deal with subsets of $\mathcal{A}$ which are not attracting, and therefore only concern a set of trajectories having measure zero. Despite this, it has always been widely believed that the flow of the Lorenz equations has the same qualitative behaviour as its geometric model.

The geometric model is made up of two pieces: one piece dealing with all trajectories passing near the origin, and one piece taking care of the global aspects of the flow. We consider a flow with a fixed point at the origin with eigenvalues just as the Lorenz flow. We also assume that there exists a unit rectangle $\Sigma \subset \{x_3 = 1\}$ which is transversal to the flow, such that the induced return map R acts on Σ as illustrated in Figure 4.

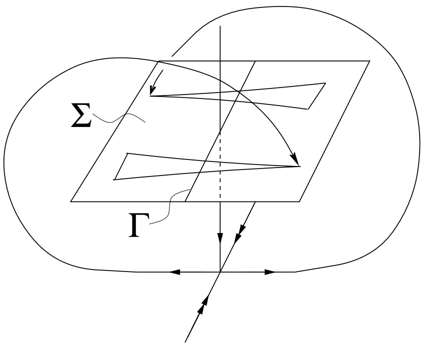

Figure 4: The return map acting on Σ.

Note that R is not defined on the line $\Gamma = \Sigma \cap W^s(0)$: these points tend to the origin, and never return to Σ. We will assume that $R(\Sigma \setminus \Gamma) \subset \Sigma$, to ensure that the flow has an attracting set with a large basin of attraction. We can now decompose the return map: $R = D \circ P$, where D is a diffeomorphism corresponding to the flow outside a unit cube centered at the origin, and P describes the flow inside the cube. By assuming that the flow is linear in the cube, we can explicitly find P:

$$P(x_1, x_2, 1) = \big(\mathrm{sgn}(x_1), x_2|x_1|^{-\lambda_2/\lambda_1}, |x_1|^{-\lambda_3/\lambda_1}\big).$$

Seeing that $-\lambda_3/\lambda_1 < 1 < -\lambda_2/\lambda_1$, we have very strong expansion in the x_1-direction, and an even stronger contraction in the x_2-direction:

$$\lim_{|x_1|\to 0} \frac{\partial P_3}{\partial x_1} = +\infty \qquad \text{and} \qquad \lim_{|x_1|\to 0} \frac{\partial P_2}{\partial x_2} = 0.$$

The model assumes that the flow outside the cube preserves the x_2-direction, i.e., that D takes the horizontal lines $\ell(t) = (\pm 1, t, c)$ into lines $\tilde{\ell}(t) = (\tilde{c}, t, 1)$, $t \in [-1, 1]$. This ensures that the contracting direction is preserved, and it also implies that the first component of the return map is independent of x_2. Therefore, we can write $R = \big(R_1(x_1), R_2(x_1, x_2)\big)$. Further assumptions are that $\frac{\partial R_2}{\partial x_2} \leq \mu < 1$ and $R_1'(x_1) > \sqrt{2}$ for all $x_1, x_2 \in \Sigma$. The return map now has a hyperbolic splitting $\mathbb{E}_x^s \oplus \mathbb{E}_x^u$, with $\mathbb{E}_0^s = \Gamma$, and the *stable leaves* $\tilde{\ell}(t)$ foliate Σ. Since all points on a stable leaf share a common future, we may form an equivalence class of such points. By taking the quotient, we get an interval map f (note that $f = R_1$), which is assumed to satisfy the following conditions:

1. f has a unique singularity at 0 with $f(0^-) = 1$, and $f(0^+) = -1$,
2. $f\colon [-1, 1] \setminus \{0\} \to [-1, 1]$,
3. f is C^1 on $[-1, 1] \setminus \{0\}$, and $f'(x) > \sqrt{2}$ for $x \neq 0$,

This suffices to prove that almost all points in $[-1, 1]$ have dense orbits under f. It is also clear that f exhibits exponential sensitivity. By pulling the information back to the original return map, it is possible to prove that the attracting set of the model flow is a generalized non-trivial hyperbolic attractor (also known as a *singular* hyperbolic attractor).

Before we close this section, let us make some simplifying remarks. By a linear change of variables, the Lorenz equations can be put in their Jordan normal form $\dot{x} = Ax + F(x)$:

$$\begin{aligned} \dot{x}_1 &= \lambda_1 x_1 - k_1(x_1 + x_2)x_3 \\ \dot{x}_2 &= \lambda_2 x_2 + k_1(x_1 + x_2)x_3 \\ \dot{x}_3 &= \lambda_3 x_3 + (x_1 + x_2)(k_2 x_1 + k_3 x_2). \end{aligned} \tag{3}$$

When we want to be very brief, we use the notation $\dot{x} = f(x)$, where $f(x)$ is the right-hand side of (3). Note that the parameters k_1, k_2 and k_3, and the eigenvalues λ_1, λ_2 and λ_3 only depend on the parameters σ, β and ϱ appearing in (2). By inserting the classical parameter values, we get the following approximate system:

$$\begin{aligned} \dot{x}_1 &= 11.8x_1 - 0.29(x_1 + x_2)x_3 \\ \dot{x}_2 &= -22.8x_2 + 0.29(x_1 + x_2)x_3 \\ \dot{x}_3 &= -2.67x_3 + (x_1 + x_2)(2.2x_1 - 1.3x_2). \end{aligned}$$

From now on, we will always refer to (3) as the Lorenz equations.

1.3 The main result

We are now ready to state the sole theorem of this paper:

Main Theorem *For the classical parameter values, the Lorenz equations support a robust strange attractor $\mathcal{A}$. Furthermore, the flow admits a unique SRB measure μ_X with* $\mathrm{supp}(\mu_X) = \mathcal{A}$.

The existence of the SRB measure is equivalent to saying that, for Lebesgue almost all points in the basin of attraction $B(\mathcal{A})$, and for all $h \in \mathrm{C}^0(B(\mathcal{A}), \mathbb{R})$, the time and space averages coincide:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T h(\varphi(x,t))dt = \int h(x)d\mu_X,$$

where μ_X is an φ-invariant probability measure.

In fact, we prove that the attracting set is a singular hyperbolic attractor. Almost all nearby points separate exponentially fast until they end up on opposite sides of the attractor. This means that a tiny blob of initial values soon smears out over the entire attractor.

It is perhaps worth pointing out that the Lorenz attractor does not act quite as the geometric model predicts. The latter can be reduced to an interval map which is everywhere expanding. This is not the case for the Lorenz attractor: there are large regions in Λ that are contracted in all directions under the return map. However, such regions only require a few more iterations before accumulating expansion. This corresponds to the interval map being eventually expanding, and does not lead to any different qualitative long-time behaviour.

Apart from this, the Lorenz attractor is just as the geometric model predicts: it contains the origin, and thus has a very complicated Cantor book structure as described in [17].

2 Outline of the approach

In this section, we will briefly describe the main phases of our approach. Basically, it can be broken down into two main sections: one global part, which involves rigorous numerics, and one local part, which is based on normal form theory. A more detailed description of all necessary steps will be given in subsequent sections.

The novelty of the method of proof lies in that, rather than producing a traditional mathematical proof, we construct an algorithm which, if successfully executed, proves the existence of the strange attractor. This algorithm is put into effect via a C-program. Hence, the only part of the proof that cannot be directly checked is the question whether the compiler and the hardware in the computer do what they are instructed to do. This is, however, a question which we will not further dwell upon. The question concerning errors due to rounding in the computer is discussed later in the text. The source codes and the list of initial data that are used in the proof are available at `http://www.math.uu.se/~warwick/thesis.html`.

2.1 Goals

In all attempts to prove Lorenz's conjecture, the main obstruction has been to gain useful global information about the flow far away from the origin. Locally, the evolution of a trajectory near a fixed point can be examined in detail using standard linearization techniques. Outside a small neighbourhood of the fixed point, however, we are usually completely at a loss. This is exactly the situation that led Guckenheimer and Williams to construct the geometric Lorenz flows, which were *defined* to have exactly those global properties that were not attainable from the original equations.

The global properties we will prove are the following:

- The return map R exists, and it is well defined in the sense of the geometric model.
- There exists a compact subset of the return plane, $N \subset \Sigma$, such that $N \setminus \Gamma$ is *forward invariant* under R, i.e., $R(N \setminus \Gamma) \subset \operatorname{int}(N)$. This ensures that the flow has an attracting set $\mathcal{A}$ with a large basin of attraction. We can then form a cross-section of the attracting set: $\mathcal{A} \cap \Sigma = \cap_{n=0}^{\infty} R^n(N) = \Lambda$. In particular, Λ is an attracting set for R.
- On N, there exists a cone field $\mathfrak{C}$ which is mapped strictly into itself by DR, i.e., for all $x \in N$, $DR(x) \cdot \mathfrak{C}(x) \subset \mathfrak{C}(R(x))$. The cones of $\mathfrak{C}$ are centered along an approximation of Λ, and each cone has an opening of $10°$.
- The tangent vectors in $\mathfrak{C}$ are eventually expanded under the action of DR: there exists $C > 0$ and $\lambda > 1$ such that for all $v \in \mathfrak{C}(x)$, $x \in N$, we have $|DR^n(x)v| \geq C\lambda^n|v|$, $n \geq 0$. In fact, the expansion is strong enough to ensure that R is topologically transitive on Λ. This is equivalent to having a dense orbit, and therefore proves that Λ is an attractor.

In the final section, we will use these properties to prove that Λ (and thus $\mathcal{A}$) carries a singular hyperbolic structure.

2.2 Normal form theory

As mentioned earlier, some regions of the return plane will flow close to the origin. These regions consist of all points in a neighbourhood of the intersection between the stable manifold of the origin and the forward invariant set, see Figure 4.

Once a rectangle has flowed close to the origin, we interrupt the computations, and introduce a close to identity change of coordinates $\Phi(\zeta) = \zeta + \phi(\zeta)$. This change of coordinates will deform the rectangle and its cone field slightly, but in a controllable way. In the new coordinates, the vector field assumes a carefully designed normal form, which is virtually linear (although it is crucial that it need not be completely linear). This permits us to estimate the evolution of the rectangle and its cone field analytically, and thereby avoid the problem of having to use computers. When changing back to the original coordinates, we once again deform the out-going rectangle and its cone field, but still in a controllable fashion.

The change of coordinates is attained by a method developed by H. Poincaré which, at first, seems fairly straight-forward. Basically, we construct ϕ formally by choosing the desired normal form mentioned above. The question of convergence, however, involves a small divisor problem, and is somewhat non-trivial. Furthermore - in order to be able to interrupt the computations before the vector field has become so small that the numerical process breaks down - we need convergence on a relatively large neighbourhood of the origin. This requires explicit knowledge of a large number of coefficients of ϕ, and thus forces us to perform some rather involved induction based proofs.

2.3 Rigorous numerics

First, we present our candidate for the trapping region N. This set consists of two disjoint components, N^- and N^+, each made up of 350 adjacent rectangles belonging to the return plane $x_3 = 27$ $(= \varrho - 1)$. We will call these small rectangles $N_i^{\pm}$, and write

$$N = N^- \cup N^+ = \big(\bigcup_{i=1}^{350} N_i^-\big) \cup \big(\bigcup_{i=1}^{350} N_i^+\big).$$

The two components of N have the same symmetry as the Lorenz equations, i.e., $N_i^+ = S(N_i^-)$, where $S(x_1, x_2, x_3) = (-x_1, -x_2, x_3)$. Thanks to this symmetry, we only have to perform the computations on one component of N. When it is not relevant which component we are considering, we omit the $\pm$ labeling of the small rectangles.

Dealing with one N_i at a time, we compute a pseudo-path that strictly contains the flow of N_i. The pseudo-paths are obtained by introducing several intermediate return planes $\Pi^{(m)}$, which are either x_1x_2-planes or x_2x_3-planes, according to whether $|\dot{x}_1| \le |\dot{x}_3|$ or vise-versa, respectively. The initial rectangle N_i is flowed to the first plane $\Pi^{(1)}$ by using an Euler method with rigorous error estimates. In the plane $\Pi^{(1)}$, we take the rectangular hull of the largest image of N_i, giving us a new starting rectangle $\mathcal{R}^{(1)}$. This rectangle is then flowed to $\Pi^{(2)}$ and so on. If a rectangle $\mathcal{R}^{(m)}$ has grown too large it is partitioned into smaller rectangles, which are then treated separately. This whole procedure is repeated until we return to Σ from above, as illustrated in Figure 4. Due to the contracting forces, the pseudo-return of N_i will consist of many overlapping rectangles $Q_{i,j}$ $j = 1, \dots, k(i)$, whose union strictly contains $R(N_i)$.

The use of rectangles significantly simplifies the computations: when flowing between two intermediate planes $\Pi^{(m)}$, $\Pi^{(m+1)}$, it is generically the corners of $\mathcal{R}^{(m)}$ that yield the largest rectangular hull $\mathcal{R}^{(m+1)} \subset \Pi^{(m+1)}$. This fact allows us to reduce the error analysis to small pieces of $\mathcal{R}^{(m)}$, which greatly reduces the local errors. With only finite precision, however, this property becomes "pseudo-generic", and has to be confirmed at every stage. The exceptional cases are treated slightly different.

Turning to the question concerning the cone field, we define the field by equipping each N_i with an initial cone. Each cone is represented by the two angles α_i^-, α_i^+ its boundary vectors $u^{(0)}, v^{(0)}$ make with the x_1-axis. As mentioned before, we always take $\theta^{(0)} = \alpha_i^+ - \alpha_i^- = \pi/18$. We then use similar techniques as just described: when a rectangle has been flowed from $\Pi^{(m)}$ to $\Pi^{(m+1)}$, we are provided with a box containing the path of the rectangle. The algorithm also gives us upper and lower bounds on the flow time involved. By solving the nine equations governing the partial derivatives of the flow, we obtain rigorous bounds on the evolution of the tangent vectors flowing through the box. By translating the flowed vectors onto the intermediate plane $\Pi^{(m+1)}$, and by selecting (incorporating the errors) the pair of vectors $u^{(m+1)}, v^{(m+1)}$ making the largest angle $\theta^{(m+1)}$, we ensure that the resulting cone contains all images of tangent vectors from the initial cone. At the return, each rectangle $Q_{i,j}$ is thus equipped with a cone represented, as above, by two angles $\beta_{i,j}^-, \beta_{i,j}^+$, $j = 1, \dots, k(i)$.

When computing the minimal expansion in each cone, we start with the widest pair of vectors, $u^{(m)}, v^{(m)}$ at each intermediate plane $\Pi^{(m)}$, as described above. If $\theta^{(m+1)} \leq \theta^{(m)}$, the minimal expansion $\varepsilon^{(m)}$ is attained on the boundary of the cone, i.e., $\varepsilon^{(m)}$ is the smallest growth factor of the images of $u^{(m)}, v^{(m)}$. If $\theta^{(m+1)} > \theta^{(m)}$, however, we must adjust this estimate by a factor which is quadratically close to unity in $\theta^{(m+1)}$. At the return each rectangle $Q_{i,j}$ is equipped with an expansion estimate $\mathcal{E}_{i,j} = \Pi_{m=0}^{n(i,j)} \varepsilon_{i,j}^{(m)}$, and $\mathcal{E}_i = \min_j \mathcal{E}_{i,j}$ gives an estimate for all vectors of the cone associated with N_i.

One major advantage of our numerical method is that we totally eliminate the problem of having to control the global effects of rounding errors due to the computer's internal floating point representation. This is achieved by using a high-dimensional analogue of *interval arithmetic*. Each object Ξ (e.g. a rectangle or a tangent vector) subjected to computation is equipped with a maximal absolute error Δ_Ξ, and can thus be represented as a box $\Xi \pm \Delta_\Xi = [\Xi_1 - \Delta_{\Xi_1}, \Xi_1 + \Delta_{\Xi_1}] \times \cdots \times [\Xi_n - \Delta_{\Xi_n}, \Xi_n + \Delta_{\Xi_n}]$. When following an object from one intermediate plane to another, we compute upper bounds on the images of $\Xi_i + \Delta_{\Xi_i}$, and lower bounds on the images of $\Xi_i - \Delta_{\Xi_i}$, $i = 1, \ldots, n$. This results in a new box $\tilde{\Xi} \pm \Delta_{\tilde{\Xi}}$, which strictly contains the exact image of $\Xi \pm \Delta_\Xi$. To ensure that we have strict inclusion, we use quite rough estimates on the upper and lower bounds. This gives us a margin which is much larger than any error caused by rounding possibly could be. Hence, the rounding errors are taken into account in the computed box $\tilde{\Xi} \pm \Delta_{\tilde{\Xi}}$, and we may continue to the following intermediate plane by restarting the whole process.

As long as we do not flow close to a fixed point, the local return maps are well defined diffeomorphisms, and the computer can handle all calculations. Some rectangles, however, will approach the origin (which is a fixed point), and then the computations must be interrupted as discussed in the previous section.

3 Local theory

3.1 Flattening out the invariant manifolds

In this section we will construct the local change of variables which straightens out the stable and unstable manifolds, and linearizes the flow on these. We will also obtain estimates on the change of variables. For convenience, we will often use the vector notation $\xi = (\xi_1, \xi_2, \xi_3)$ combined with the multi-index notation $\xi^n = \xi_1^{n_1}\xi_2^{n_2}\xi_3^{n_3}$.

In order that the stable and unstable manifolds should coincide with the coordinate planes, it is necessary that these are invariant under the flow. To ensure this, we need a change of variables which, in a small neighbourhood of the origin, transforms the Lorenz equations $\dot{\xi} = A\xi + F(\xi)$ into $\dot{\zeta} = A\zeta + G(\zeta)$, where $G = (G_1, G_2, G_3)$, satisfies the following conditions:

$$G_1(0, \zeta_2, \zeta_3) = 0, \quad \text{and} \quad G_2(\zeta_1, 0, 0) = G_3(\zeta_1, 0, 0) = 0.$$

In these new coordinates, the unstable manifold coincides with the ζ_1-axis, and the stable manifold coincides with the $\zeta_2\zeta_3$-plane. However, this will not linearize the flow on the invariant manifolds. For this, we need to impose the condition

$$G_i \in \mathcal{O}(\zeta_1) \cap \mathcal{O}(\zeta_2, \zeta_3) \qquad (i = 1, 2, 3).$$

This simply means that if a point ζ is close to the ζ_1-axis (the unstable manifold) or the $\zeta_2\zeta_3$-plane (the stable manifold), then the perturbation $G(\zeta)$ is linearly small, i.e.,

$$\min\{|\zeta_1|, |\zeta_2| + |\zeta_3|\} = \mathcal{O}(\varepsilon) \Rightarrow |G_i(\zeta)| = \mathcal{O}(\varepsilon) \qquad (i = 1, 2, 3).$$

However, we will need to do better than this: we actually want to flatten out the invariant manifolds even more. Flatness of order p is given by requiring that $G \in \mathcal{O}^p(\zeta_1) \cap \mathcal{O}^p(\zeta_2, \zeta_3)$, i.e.,

$$\min\{|\zeta_1|, |\zeta_2| + |\zeta_3|\} = \mathcal{O}(\varepsilon) \Rightarrow |G_i(\zeta)| = \mathcal{O}(\varepsilon^p) \qquad (i = 1, 2, 3).$$

To be able to talk about smallness, we need a norm to work with. We will work in a complex neighbourhood of the origin, and use the following notations:

$$|\zeta| = \max\{|\zeta_i| \colon i = 1, 2, 3\} \qquad \|f\|_r = \max\{|f(\zeta)| \colon |\zeta| \le r\}.$$

Our Ansatz is to do the calculations with formal vector-valued polynomials. The following proposition states that we obtain not only a formal change of variables, but that the formal power series actually converges in a fixed neighbourhood of the origin.

Proposition 3.1 *There exists a close to identity change of variables $\xi = \zeta + \phi(\zeta)$ with*

$$\|\phi\|_r \le \frac{r^2}{2} \qquad r \le 1$$

such that the Lorenz equations, $\dot{\xi} = A\xi + F(\xi)$, are transformed into the normal form $\dot{\zeta} = A\zeta + G(\zeta)$, where $G(\zeta) \in \mathcal{O}^{10}(\zeta_1) \cap \mathcal{O}^{10}(\zeta_2, \zeta_3)$, and satisfies

$$\|G\|_r \le 7 \cdot 10^{-9} \frac{r^{20}}{1 - 3r} \qquad r < \frac{1}{3}.$$

Before getting involved in the proof of the proposition, we highlight some important consequences of the statement.

Lemma 3.2 *For any ρ satisfying $0 \le \rho \le \frac{1}{2}$, we have*

$$\|D\phi\|_\rho \le 2\rho.$$

Proof: We use a classical argument in function theory: If $h(z)$ is a regular function of a complex variable z in the disk $|z| \le r$ where it satisfies $|h(z)| \le M$, then for $|z| \le \rho < r$ Cauchy's integral formula gives

$$h'(z) = \frac{1}{2\pi i} \int_\gamma \frac{h(w)}{(w-z)^2} dw,$$

where the path of integration γ can be taken as the circle $|w - z| = r - \rho$. Since the circle lies within $|w| \le r$, this leads to the estimate

$$|h'(z)| \le M/(r-\rho) \qquad |z| \le \rho.$$

Now, given w and ζ such that $|w| = 1$ and $|\zeta| = \rho$, we define

$$h(z) = \phi_i(\zeta_1 + w_1 z, \zeta_2 + w_2 z, \zeta_3 + w_3 z).$$

For $|z| \le r - \rho$, we clearly have $|h(z)| \le \|\phi\|_r$, and so $|h'(z)| \le \|\phi\|_r/(r - \rho - |z|)$. In particular, for $z = 0$, we have $|h'(0)| \le \|\phi\|_r/(r-\rho)$. On the other hand, we have

$$|h'(0)| = \Big| \sum_{j=1}^{3} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta) w_j \Big| = \sum_{j=1}^{3} \Big| \frac{\partial \phi_i}{\partial \zeta_j}(\zeta) \Big|,$$

by taking $w_j = \operatorname{sgn}(\frac{\partial \phi_i}{\partial \zeta_j}(\zeta))$. Since $\|D\phi\|_\rho = \max\{|D\phi(\zeta)w| \colon |\zeta| \le \rho, |w| = 1\}$, we immediately have the following estimates of $\|D\phi\|_\rho$ for $\rho < r \le 1$:

$$\|D\phi\|_\rho \le \frac{\|\phi\|_r}{r-\rho} \le \frac{r^2}{2(r-\rho)}. \tag{4}$$

It is easy to see that the optimal bound is given by substituting r for 2ρ. Inserting this value directly gives the result. □

Lemma 3.3 *For $|\zeta| \le r < \frac{1}{2}$, the change of variables $\xi = \zeta + \phi(\zeta)$ has a well defined inverse $\zeta = \xi + \psi(\xi)$ in the ball $|\xi| \le \tilde{r} = r - \|\phi\|_r$ satisfying*

$$\|\psi\|_{\tilde{r}} \le \|\phi\|_r \qquad \|D\psi\|_{\tilde{r}} \le \frac{\|D\phi\|_r}{1 - \|D\phi\|_r}.$$

Proof: Set $\Phi(\zeta) = \zeta + \phi(\zeta)$, $\Psi(\xi) = \xi + \psi(\xi)$ and let D denote the ball $|\zeta| \le r$. By Proposition 3.1, it is clear that $\Phi(D)$ must contain a ball $\tilde{D}$, centered at the origin, and of radius $\tilde{r} = r - \|\phi\|_r$, see Figure 5. Let $\zeta_1, \zeta_2 \in D$. Then we have

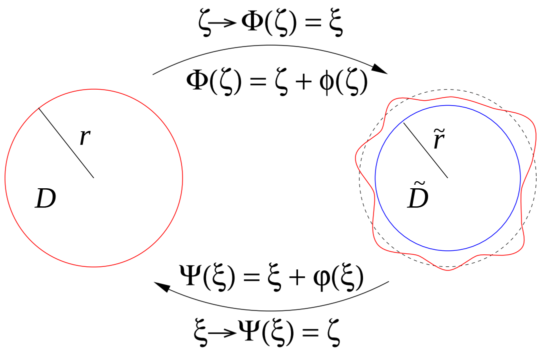

Figure 5: The change of variables and its inverse.

$$|\Phi(\zeta_1) - \Phi(\zeta_2)| \ge |\zeta_1 - \zeta_2| - |\phi(\zeta_1) - \phi(\zeta_2)| \ge (1 - \|D\phi\|_r)|\zeta_1 - \zeta_2|.$$

Hence, if $\|D\phi\|_r < 1$, then Φ is injective on the whole of $\Phi(D)$, and in particular on $\tilde{D}$. By Lemma 3.2, we know that $\|D\phi\|_r \leq 2r$ for all $r \leq \frac{1}{2}$. This proves that the inverse Φ is well defined in the ball $|\xi| \leq \tilde{r} = r - \|\phi\|_r$ for $r < \frac{1}{2}$.

Combining the two coordinate changes, we immediately arrive at

$$\psi(\xi) = -\phi(\xi + \psi(\xi)),$$

which gives the first estimate of the lemma. Differentiating gives

$$D\psi(\xi) = -D\phi(\xi + \psi(\xi))(I + D\psi(\xi)),$$

which, after solving for $D\psi(\xi)$, becomes

$$D\psi(\xi) = -[I + D\phi(\xi + \psi(\xi))]^{-1} D\phi(\xi + \psi(\xi)).$$

The second estimate now follows by using the well known estimate

$$\left\| [I + D\phi]^{-1} \right\|_r \leq \frac{1}{1 - \|D\phi\|_r}.$$

□

Let us conclude this section by explicitly computing the numerical values of the coordinate changes used in the program `lorenz_2.1.c`. We will change coordinates in the cube centered at the origin, and having radius $\frac{1}{5}$.

Lemma 3.4

$$\|\phi\|_{\frac{1}{5}} \leq \frac{1}{50} = 0.02 \qquad \|D\phi\|_{\frac{1}{5}} \leq \frac{2}{5} = 0.4,$$
$$\|\psi\|_{\frac{1}{5}} < 0.02541 \qquad \|D\psi\|_{\frac{1}{5}} < 0.8209.$$

Proof: The first line of the lemma follows immediately from Proposition 3.1 and Lemma 3.2, respectively. The second line requires a little thought. By Lemma 3.3, we must find r^* such that $\frac{1}{5} = r^* - \|\phi\|_{r^*}$. Once we know r^*, we can use the formulae

$$\|\psi\|_{\frac{1}{5}} \leq \|\phi\|_{r^*} \qquad \|D\psi\|_{\frac{1}{5}} \leq \frac{\|D\phi\|_{r^*}}{1 - \|D\phi\|_{r^*}}$$

to obtain the desired estimates. Solving for r^* gives $r^* = 1 - \sqrt{\frac{3}{5}} \approx 0.2254$, and plugging this into the estimates above gives

$$\|\psi\|_{\frac{1}{5}} \leq \tfrac{1}{5}\big(4 - \sqrt{15}\big) \qquad \|D\psi\|_{\frac{1}{5}} \leq \tfrac{2}{7}\big(\sqrt{15} - 1\big).$$

The numerical values (rounded up) of these estimates appear in the statement. □

This lemma tells us several things: when we enter the small cube, we must increase the sides of the rectangle by the amount $2 \cdot 0.02$. We also lose 40 per-cent of the cone field information. When we leave the small cube, we must increase the sides of the outgoing rectangle by the amount $2 \cdot 0.02541$, and now we lose roughly 82 per-cent of the cone field information. This may seem like a great loss of information, but thanks to the strong hyperbolicity near the origin, the cone field will still remain intact.

3.2 Proof of Proposition 3.1

First, we need to know how the vector field (3) is affected by the close to identity change of variables $\xi = \zeta + \phi(\zeta)$. We have the following:

$$\dot{\xi} = A(\zeta + \phi(\zeta)) + F(\zeta + \phi(\zeta)) = A\zeta + A\phi(\zeta) + F(\zeta + \phi(\zeta)). \tag{5}$$

On the other hand, we also have

$$\begin{aligned}\dot{\xi} &= \frac{d}{dt}(\zeta + \phi(\zeta)) = (I + D\phi(\zeta))\dot{\zeta} = (I + D\phi(\zeta))(A\zeta + G(\zeta)) \\ &= A\zeta + D\phi(\zeta)A\zeta + G(\zeta) + D\phi(\zeta)G(\zeta). \end{aligned} \tag{6}$$

Comparing the two right-hand sides of (5) and (6) gives

$$D\phi(\zeta)A\zeta - A\phi(\zeta) = F(\zeta + \phi(\zeta)) - D\phi(\zeta)G(\zeta) - G(\zeta). \tag{7}$$

For shorthand, we will use the following notation

$$L_A\phi(\zeta) = D\phi(\zeta)A\zeta - A\phi(\zeta).$$

The operator L_A is linear, and it acts on the space of formal vector fields. It leaves the spaces of homogeneous vector-valued polynomials of any degree invariant. Looking at (7) on component level, we have

$$L_{A,i}\phi_i(\zeta) = F_i(\zeta + \phi(\zeta)) - \sum_{j=1}^{3} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta)G_j(\zeta) - G_i(\zeta) \qquad (i = 1, 2, 3), \tag{8}$$

where

$$L_{A,i}\phi_i(\zeta) = \sum_{j=1}^{3} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta)\lambda_j\zeta_j - \lambda_i\phi_i(\zeta) \qquad (i = 1, 2, 3).$$

Note that

$$L_{A,i}(a_n\zeta^n) = (n_1\lambda_1 + n_2\lambda_2 + n_3\lambda_3 - \lambda_i)a_{n_1,n_2,n_3}\zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3} = (n\lambda - \lambda_i)a_n\zeta^n.$$

The crux is now to choose ϕ so that we have $G_i \in \mathcal{O}^p(\zeta_1) \cap \mathcal{O}^p(\zeta_2, \zeta_3)$. This means that $G_i(\zeta)$, must not contain elements on the form $\zeta^n = \zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3}$, where the natural numbers n_i satisfy $n_1 < p$ or $n_2 + n_3 < p$. By (8), elements on this form can only come from $F_i(\zeta + \phi(\zeta))$, and any such term can be canceled by an appropriate choice of ϕ_i if the corresponding *divisor* $(n\lambda - \lambda_i)$ does not vanish. Thus the component functions ϕ_i $(i = 1, 2, 3)$, need only consist of the undesired elements just described.

We start by splitting the 3-space of natural numbers into two disjoint sets: $\mathbb{N}^3 = \mathbb{U}_p \cup \mathbb{V}_p$, where

$$\mathbb{V}_p = \{(n_1, n_2, n_3) \in \mathbb{N}^3 \colon n_1 < p \text{ or } n_2 + n_3 < p\}.$$

Next, we define the following filters, which act on formal vector-valued polynomials: Consider $H = (H_1, H_2, H_3)$, where

$$H_i(\zeta) = \sum_n a_{i,n}\zeta^n = \sum_n a_{i,n_1,n_2,n_3}\zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3} \qquad (i = 1, 2, 3).$$

Then we define

$$\{H_i(\zeta)\}_{\mathbb{U}_p} = \sum_{n \in \mathbb{U}_p} a_{i,n}\zeta^n, \quad \text{and} \quad \{H_i(\zeta)\}_{\mathbb{V}_p} = \sum_{n \in \mathbb{V}_p} a_{i,n}\zeta^n.$$

We extend the definition of these filters so that they act not only on components, but also on the whole formal vector-valued polynomial: $H = \{H\}_{\mathbb{U}_p} \oplus \{H\}_{\mathbb{V}_p}$. Note that $\{G\}_{\mathbb{U}_p} = G$, $\{G\}_{\mathbb{V}_p} = 0$, $\{\phi\}_{\mathbb{U}_p} = 0$, $\{\phi\}_{\mathbb{V}_p} = \phi$, and $\{MG\}_{\mathbb{V}_p} = 0$ for any 3×3-matrix M with formal polynomial entries.

By filtering (8), we get

$$L_{A,i}\phi_i(\zeta) = \{F_i(\zeta + \phi(\zeta))\}_{\mathbb{V}_p} \qquad (i = 1, 2, 3), \tag{9}$$

and

$$G_i(\zeta) = \{F_i(\zeta + \phi(\zeta))\}_{\mathbb{U}_p} - \sum_{j=1}^{3} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta)G_j(\zeta) \qquad (i = 1, 2, 3). \tag{10}$$

3.2.1 The change of variables

The recursive scheme (9) can be formally solved by a power series

$$\phi_i(\zeta) = \sum_{|n|=2}^{\infty} a_{i,n}\zeta^n \qquad (i = 1, 2, 3),$$

where the coefficients are determined by inserting this expression into (9). The existence of a solution ϕ is given by comparing both sides of (9) : If $a_{i,n}\zeta^n$ is a term of $\phi_i(\zeta)$ with $|n| = n_1+n_2+n_3$, the comparison gives

$$(n\lambda - \lambda_i)a_{i,n} = \gamma,$$

where γ is a polynomial in the coefficients of the terms in ϕ_1, ϕ_2, ϕ_3 of degree less than $|n|$. Thus the existence is proved if we show that the divisors $(n\lambda - \lambda_i)$ do not vanish. As ϕ does not contain constant or linear terms, and $\{\phi\}_{\mathbb{V}_p} = \phi$, the only divisors we need to consider are on the form $n\lambda - \lambda_i$ $(i = 1, 2, 3)$, where $n \in \mathbb{V}_p$ and $|n| \geq 2$. The following computer-aided lemma proves the existence of a formal series for ϕ.

Lemma 3.5 *For any $p \in [2, 10]$, $n \in \mathbb{V}_p$ and $|n| \geq 2$, the divisors*

$$n\lambda - \lambda_i \qquad (i = 1, 2, 3)$$

are bounded away from zero. Furthermore, after a finite time of fluctuations, there exists a sharp lower bound on the modulus of these divisors:

$$|n\lambda - \lambda_i| \geq \big|(p-1)\lambda_1 + \big(|n| - (p-1)\big)\lambda_3 - \lambda_i\big| \qquad (i = 1, 2, 3).$$

This bound is valid when $|n| \geq \mathcal{N}(p)$, where $\mathcal{N}(p) = \frac{11p+6}{2}$ if p is even, and $\mathcal{N}(p) = \frac{11p+7}{2}$ if p is odd.

Proof: Take $|n|$ large. Since $n \in \mathbb{V}_p$, there are two cases to consider:
(1) $n_1 < p$: This means that n_2+n_3 is large, i.e, the divisor $n\lambda - \lambda_i$ is large and negative. Recalling that the eigenvalues satisfy $0 < -\lambda_3 < \lambda_1 < -\lambda_2$, we clearly minimize the divisor's modulus by choosing $n_1 = p-1$, $n_2 = 0$ and $n_3 = |n| - (p-1)$, which gives

$$|n\lambda - \lambda_i| \geq \big|(p-1)\lambda_1 + \big(|n| - (p-1)\big)\lambda_3 - \lambda_i\big| \qquad (i = 1, 2, 3).$$

(2) $n_2 + n_3 < p$: This means that n_1 is large, i.e, the divisor $n\lambda - \lambda_i$ is large and positive. Its modulus is minimized by choosing $n_1 = |n| - (p-1)$, $n_2 = p-1$ and $n_3 = 0$, which gives

$$|n\lambda - \lambda_i| \geq \big|\big(|n| - (p-1)\big)\lambda_1 + (p-1)\lambda_2 - \lambda_i\big| \qquad (i = 1, 2, 3). \tag{11}$$

Comparing these two candidates for the smallest divisor, we find that case (1) yields the optimal lower bound, as stated in the lemma. This completes the proof of the asymptotic lower bound.

To see when the lower bound becomes valid, we note that the choice $i = 2$ in the right-hand side of (11) gives the smallest divisor for $|n|$ large, i.e.,

$$|n\lambda - \lambda_i| \geq \big|(p-1)\lambda_1 + \big(|n| - (p-1)\big)\lambda_3 - \lambda_2\big| \qquad (i = 1, 2, 3).$$

The expression on the right-hand side has a minimum when its positive part has roughly the same modulus as its negative part. This occurs when

$$(p-1)\lambda_1 + |\lambda_2| \approx \big(|n| - (p-1)\big)|\lambda_3|,$$

or, equivalently, when

$$|n| \approx \frac{1}{|\lambda_3|}\big((p-1)(\lambda_1 + |\lambda_3|) + |\lambda_2|\big).$$

By defining

$$\mathcal{N}(p) = \left\lceil \frac{1}{|\lambda_3|}\big((p-1)(\lambda_1 + |\lambda_3|) + |\lambda_2|\big) \right\rceil,$$

where $\lceil x \rceil = \min\{k\colon k \in \mathbb{N}, k \geq x\}$, we have $\mathcal{N}(p) = \frac{11p+6}{2}$ if p is even, and $\mathcal{N}(p) = \frac{11p+7}{2}$ if p is odd for $p \in [2, 10]$, as stated in the lemma. That this is the appropriate choice for $\mathcal{N}(p)$ when $p \in [2, 10]$ (for $p \geq 11$ it is not) can be checked by explicit calculations.

Finally, to show that the low-order divisors are non-zero, we note that there are only a finite number of them that we have to check. This can be done by explicit calculations carried out on a computer, and gives the desired result. In Table 1, we list the values of $\Omega(k) = \min_{i=1,2,3} \min\{|\lambda n - \lambda_i|\colon |n| = k, n \in \mathbb{V}_{10}\}$. This required the computation of 19,386 divisors. The C-program `smalldiv.c` handles all necessary computations. $\square$

$\Omega(2) \approx 2.6667$	$\Omega(3) \approx 3.4944$	$\Omega(4) \approx 0.8277$	$\Omega(5) \approx 1.8389$
$\Omega(6) \approx 1.1611$	$\Omega(7) \approx 1.5056$	$\Omega(8) \approx 1.0112$	$\Omega(9) \approx 1.1723$
$\Omega(10) \approx 0.6779$	$\Omega(11) \approx 0.1835$	$\Omega(12) \approx 0.3446$	$\Omega(13) \approx 0.1498$
$\Omega(14) \approx 0.0112$	$\Omega(15) \approx 0.4832$	$\Omega(16) \approx 0.9776$	$\Omega(17) \approx 0.8165$
$\Omega(18) \approx 1.1947$	$\Omega(19) \approx 1.1498$	$\Omega(20) \approx 0.8614$	$\Omega(21) \approx 0.3670$
$\Omega(22) \approx 0.5280$	$\Omega(23) \approx 0.6891$	$\Omega(24) \approx 0.1947$	$\Omega(25) \approx 0.3558$
$\Omega(26) \approx 0.1386$	$\Omega(27) \approx 0.6330$	$\Omega(28) \approx 0.4720$	$\Omega(29) \approx 0.9663$
$\Omega(30) \approx 0.8053$	$\Omega(31) \approx 1.2997$	$\Omega(32) \approx 1.3670$	$\Omega(33) \approx 1.6330$
$\Omega(34) \approx 1.0337$	$\Omega(35) \approx 0.5393$	$\Omega(36) \approx 0.7003$	$\Omega(37) \approx 0.2059$
$\Omega(38) \approx 2.7941$	$\Omega(39) \approx 0.1274$	$\Omega(40) \approx 2.5393$	$\Omega(41) \approx 0.4607$
$\Omega(42) \approx 0.9551$	$\Omega(43) \approx 1.7115$	$\Omega(44) \approx 1.2885$	$\Omega(45) \approx 1.3782$
$\Omega(46) \approx 1.6218$	$\Omega(47) \approx 1.0449$	$\Omega(48) \approx 3.7115$	$\Omega(49) \approx 2.4495$
$\Omega(50) \approx 0.2172$	$\Omega(51) \approx 2.7828$	$\Omega(52) \approx 0.1162$	$\Omega(53) \approx 2.5505$
$\Omega(54) \approx 5.2172$	$\Omega(55) \approx 6.6106$	$\Omega(56) \approx 3.9439$	$\Omega(57) \approx 1.2772$
$\Omega(58) \approx 1.3894$	$\Omega(59) \approx 4.0561$	$\Omega(60) \approx 6.7228$	$\Omega(61) \approx 9.3895$

Table 1: The smallest absolute values of low-order divisors for $p = 10$.

Remark: For a given p, the non-vanishing of the divisors is an open condition. Thus the lemma is valid for an open neighbourhood of the classical parameter values of the Lorenz equations.

Remark: The lemma gives an asymptotic estimate on the growth of the modulus of the smallest divisors. For large $|n|$, we have $|n\lambda - \lambda_i| \sim |n||\lambda_3| \approx \frac{8|n|}{3}$.

Now that we know that the formal power series for ϕ defined by (9) exists, we want to show that it also actually converges. We follow [14], and use the methods of majorants. If

$$f(\zeta) = \sum_n a_{n_1,n_2,n_3}\zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3} \qquad g(\zeta) = \sum_n b_{n_1,n_2,n_3}\zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3}$$

are two formal power series, g is said to be a majorant of f, which we denote $f \prec g$, if

$$|a_{n_1,n_2,n_3}| \leq b_{n_1,n_2,n_3}$$

for all the coefficients. Note that the coefficients of g must be real and non-negative, which implies that f must have at least as large radius of convergence as g. We will also use the notation

$$[f(\zeta)]_k = \sum_{|n|=k} a_{n_1,n_2,n_3}\zeta_1^{n_1}\zeta_2^{n_2}\zeta_3^{n_3}.$$

Now suppose that we find a function $\tilde{F}\colon \mathbb{C}^3 \to \mathbb{C}$ such that $F_i \prec \tilde{F}$ $(i = 1, 2, 3)$ and, together with (9), consider the majorant system

$$\tilde{L}_A\varphi_i(\zeta) = \left\{\tilde{F}(\zeta + \varphi(\zeta))\right\}_{\mathbb{V}_p} \qquad (i = 1, 2, 3), \tag{12}$$

where $\tilde{L}_A(\zeta^n) = \Theta(n)\zeta^n$ and $\Theta\colon \mathbb{N}^3 \to \mathbb{R}$ is defined by $\Theta(n) = \min_{i=1,2,3} |n\lambda - \lambda_i|$. This can again be solved formally by a power series

$$\varphi_i(\zeta) = \sum_{|n|=2}^{\infty} b_{i,n}\zeta^n, \tag{13}$$

and it is clear that φ_i is a majorant of ϕ_i. Furthermore, since the right-hand side of (12) is independent of i, we have $\varphi_1 = \varphi_2 = \varphi_3 = \tilde{\varphi}$. If we set $\zeta_1 = \zeta_2 = \zeta_3 = \zeta$, and find a function $\hat{F}\colon \mathbb{C} \to \mathbb{C}$ such that $\tilde{F}(\zeta,\zeta,\zeta) \prec \hat{F}(\zeta)$, we may, together with (12), consider the majorant system

$$\hat{L}_A\psi(\zeta) = \hat{F}(\zeta + \psi(\zeta)), \tag{14}$$

where $\hat{L}_A(\zeta^k) = \Omega(k)\zeta^k$ and $\Omega\colon \mathbb{N} \to \mathbb{R}$ is defined by $\Omega(k) = \min\{\Theta(n)\colon |n| = k, \quad n \in \mathbb{V}_p\}$. Again, this can be solved formally by a power series

$$\psi(\zeta) = \sum_{k=2}^{\infty} c_k\zeta^k, \tag{15}$$

and it is clear that $\tilde{\varphi}(\zeta,\zeta,\zeta) \prec \psi(\zeta)$. Note that this implies that $\|\phi\|_r \le \psi(r)$ in the region of convergence. Thus it suffices to prove the convergence of ψ.

Recall that the non-linear part of the Lorenz equations is given by

$$F(\zeta) = \begin{pmatrix} -k_1\zeta_1\zeta_3 - k_1\zeta_2\zeta_3 \\ k_1\zeta_1\zeta_3 + k_1\zeta_2\zeta_3 \\ k_2\zeta_1^2 + (k_2+k_3)\zeta_1\zeta_2 + k_3\zeta_2^2 \end{pmatrix} \approx \begin{pmatrix} -0.29\zeta_1\zeta_3 - 0.29\zeta_2\zeta_3 \\ 0.29\zeta_1\zeta_3 + 0.29\zeta_2\zeta_3 \\ 2.2\zeta_1^2 + 0.9\zeta_1\zeta_2 - 1.3\zeta_2^2 \end{pmatrix}.$$

Thus if we set

$$\tilde{F}(\zeta) = k_2\zeta_1^2 + (k_2+k_3)\zeta_1\zeta_2 + k_1\zeta_1\zeta_3 - k_3\zeta_2^2 + k_1\zeta_2\zeta_3,$$

we clearly have $F_i \prec \tilde{F}$ $(i = 1,2,3)$. Using the exceptional fact that F only contains quadratic terms, we can find a particularly simple majorant of $\tilde{F}(\zeta,\zeta,\zeta)$:

$$\tilde{F}(\zeta,\zeta,\zeta) = k_2\zeta^2 + (k_2+k_3)\zeta^2 + k_1\zeta^2 - k_3\zeta^2 + k_2\zeta^2 \prec 5\zeta^2 = \hat{F}(\zeta).$$

Combining this with (14) and (15), we get the following recursive scheme for the coefficients of $\psi(r)$:

$$c_m r^m = \frac{5}{\Omega(m)}\left[\left(r + \sum_{k=2}^{m-1} c_k r^k\right)^2\right]_m \qquad (m = 2,3,\ldots). \tag{16}$$

Unfortunately, by Lemma 3.5, we have no uniform estimates on $\Omega(m)$ for $m < \mathcal{N}(p)$. This is due to low-order resonances between the eigenvalues, and therefore we must calculate these divisors explicitly. We did this already in the proof of Lemma 3.5, and found that the smallest modulus of such a divisor was $|\lambda_1 + 13\lambda_3 - \lambda_2| \approx 0.0112$. This divisor appears already for $p = 2$, and it remains minimizing for all $p \in [2,10]$. Using this a an crude estimate on $\Omega(m)$ for $m < \mathcal{N}(p)$ results in very poor bounds on the coefficients of ψ, and thus gives a very small radius of convergence. This problem can be avoided by postponing the use of the recursive scheme (16) until $m \ge \mathcal{N}(p)$. We can achieve this if we can estimate the leading coefficients of ψ by other means. This will produce a smaller function (which we still call ψ) than had we used only (16), but the important fact is that it will still majorize the ϕ_i:s.

The way we will proceed is to explicitly compute $a_{i,n}$ $(i = 1,2,3)$ for $|n| = 2,\ldots,n_1$ by using (9). This is possible by Lemma 3.5. Then, if we define the leading coefficients of φ by $b_n = \max_{i=1,2,3} |a_{i,n}|$ for $|n| = 2,\ldots,n_1$, we clearly have $[\phi_i(\zeta)]_j \prec [\varphi(\zeta)]_j$ $(i = 1,2,3)$ for $j = 2,\ldots,n_1$. Continuing, we may define the leading coefficients of ψ by

$$c_j = \sum_{|n|=j} b_n = \sum_{|n|=j} \max_{i=1,2,3} |a_{i,n}| \qquad (j = 2,\ldots,n_1).$$

Naturally, we then have $[\varphi(r,r,r)]_j = [\psi(r)]_j$ $(= c_j)$ for $j = 2, ..., n_1$. Now assume that we fix a positive integer n_0 and find two positive constants C and M such that the following induction assumption holds: $c_j \leq CM^j$ for $j = n_0+1, ..., 2n_0, ..., n_1$, where $n_1 \geq \mathcal{N}(p)$. We then want to prove that $c_j \leq CM^j$ for all $j > n_1$. This we do by induction: by defining $c_1 = 1$, the recursive scheme (16), gives

$$c_{n_1+1}r^{n_1+1} = \frac{5}{\Omega(n_1+1)}\left[\left(\sum_{k=1}^{n_1} c_k r^k\right)^2\right]_{n_1+1} = \frac{5}{\Omega(n_1+1)}\left(\sum_{k=1}^{n_1} c_k c_{n_1+1-k}\right) r^{n_1+1},$$

so the coefficients of ψ satisfy

$$\begin{aligned} c_{n_1+1} &= \frac{5}{\Omega(n_1+1)}\left(\sum_{k=1}^{n_0} c_k c_{n_1+1-k} + \sum_{k=n_0+1}^{n_1-n_0} c_k c_{n_1+1-k} + \sum_{k=n_1+1-n_0}^{n_1} c_k c_{n_1+1-k}\right) \\ &= \frac{5}{\Omega(n_1+1)}\left(2\sum_{k=1}^{n_0} c_k c_{n_1+1-k} + \sum_{k=n_0+1}^{n_1-n_0} c_k c_{n_1+1-k}\right). \end{aligned}$$

Although we know nothing about the n_0 first coefficients, we know that $c_j \leq CM^j$ for $j = n_0+1, \ldots, n_1$ by the induction hypothesis. Using this gives

$$\begin{aligned} c_{n_1+1} &\leq \frac{5}{\Omega(n_1+1)}\left(2\sum_{k=1}^{n_0} c_k CM^{n_1+1-k} + \sum_{k=n_0+1}^{n_1-n_0} C^2 M^{n_1+1}\right) \\ &= \frac{5}{\Omega(n_1+1)}\left(2\sum_{k=1}^{n_0} c_k M^{-k} + (n_1 - 2n_0)C\right) CM^{n_1+1}. \end{aligned}$$

Since we chose n_1 large enough for the lower bound of $\Omega(n_1+1)$ (see Lemma 3.5) to be valid, we have

$$c_{n_1+1} \leq \frac{5\left(2\sum_{k=1}^{n_0} c_k M^{-k} + (n_1-2n_0)C\right)}{|(p-1)\lambda_1 + (n_1+1-(p-1))\lambda_3 - \lambda_2|} CM^{n_1+1}, \tag{17}$$

which, for n_1 large, gives the asymptotic estimate

$$c_{n_1+1} \sim \frac{5C}{|\lambda_3|} CM^{n_1+1}.$$

From this it follows that, in order to prove the induction step, we must choose $C \leq \frac{|\lambda_3|}{5} \approx \frac{8}{15}$ ($C \leq \frac{1}{2}$ works nicely for an open neighbourhood of the classical parameter values). Then, if we take n_1 sufficiently large, we will have completed the induction step.

The following computer-aided lemma[3] proves the induction hypothesis needed in the above argument, and gives an estimate on the sum appearing in (17).

Lemma 3.6 *For $p \in [2, 10]$, we have the estimates*

$$c_j < 5 \cdot 10^{-6}\left(\frac{5}{9}\right)^j \qquad (j = 11, ..., 70) \quad \textit{and} \quad \sum_{j=1}^{10} c_j \left(\frac{5}{9}\right)^{-j} < 3.54.$$

Furthermore, $\mathcal{N}(p) < 70$ for $p \in [2, 10]$.

Proof: A computer program that computes the coefficients $a_{i,n}$, and then calculates the sums $c_j = \sum_{|n|=j} \max_{i=1,2,3} |a_{i,n}|$ was constructed, see `coeff.c`. Thanks to the simple form of the

[3] The reader may be interested in knowing that verifying this induction hypothesis requires knowledge of the first 186,576 coefficients of ϕ.

non-linear terms appearing in the Lorenz equations, the program only needs to handle two sums and two products. If we define $\Phi(\zeta) = \zeta + \phi(\zeta)$, the two sums we need to compute are of the type:

$$S_1 = [\Phi_1(\zeta) + \Phi_2(\zeta)]_n \text{ and } S_2 = [k_2\Phi_1(\zeta) + k_3\Phi_2(\zeta)]_n,$$

and the two products are of the type:

$$P_1 = [(\Phi_1(\zeta) + \Phi_2(\zeta))\Phi_3(\zeta)]_n \text{ and } P_2 = [(\Phi_1(\zeta) + \Phi_2(\zeta))(k_2\Phi_1(\zeta) + k_3\Phi_2(\zeta))]_n.$$

The exact implementations can be found at (B1) and (B2), respectively. The results are presented in Table 2, and it is simple to check that the coefficients c_j satisfy the conditions in the lemma. Since, for $p \in [2, 10]$, $\mathcal{N}(p) \leq 58$, the final statement in the lemma is verified. □

$c_2 <$ 4.373e-01	$c_3 <$ 4.321e-02	$c_4 <$ 4.928e-03
$c_5 <$ 5.703e-04	$c_6 <$ 7.196e-05	$c_7 <$ 1.095e-05
$c_8 <$ 2.250e-06	$c_9 <$ 3.762e-07	$c_{10} <$ 7.072e-08
$c_{11} <$ 7.459e-09	$c_{12} <$ 1.091e-09	$c_{13} <$ 1.221e-10
$c_{14} <$ 1.541e-11	$c_{15} <$ 1.824e-12	$c_{16} <$ 2.524e-13
$c_{17} <$ 4.315e-14	$c_{18} <$ 7.176e-15	$c_{19} <$ 1.133e-15
$c_{20} <$ 1.245e-16	$c_{21} <$ 1.432e-17	$c_{22} <$ 1.672e-18
$c_{23} <$ 1.302e-19	$c_{24} <$ 1.480e-20	$c_{25} <$ 1.473e-21
$c_{26} <$ 1.353e-22	$c_{27} <$ 1.618e-23	$c_{28} <$ 1.517e-24
$c_{29} <$ 2.453e-25	$c_{30} <$ 2.116e-26	$c_{31} <$ 2.951e-27
$c_{32} <$ 3.418e-28	$c_{33} <$ 3.512e-29	$c_{34} <$ 3.177e-30
$c_{35} <$ 3.107e-31	$c_{36} <$ 2.775e-32	$c_{37} <$ 2.241e-33
$c_{38} <$ 1.702e-34	$c_{39} <$ 1.227e-35	$c_{40} <$ 9.371e-37
$c_{41} <$ 6.902e-38	$c_{42} <$ 5.015e-39	$c_{43} <$ 3.376e-40
$c_{44} <$ 2.711e-41	$c_{45} <$ 2.264e-42	$c_{46} <$ 2.108e-43
$c_{47} <$ 1.935e-44	$c_{48} <$ 1.868e-45	$c_{49} <$ 1.733e-46
$c_{50} <$ 1.511e-47	$c_{51} <$ 1.217e-48	$c_{52} <$ 9.282e-50
$c_{53} <$ 6.713e-51	$c_{54} <$ 4.710e-52	$c_{55} <$ 3.299e-53
$c_{56} <$ 2.311e-54	$c_{57} <$ 1.542e-55	$c_{58} <$ 1.052e-56
$c_{59} <$ 7.835e-58	$c_{60} <$ 5.670e-59	$c_{61} <$ 4.451e-60
$c_{62} <$ 3.834e-61	$c_{63} <$ 3.421e-62	$c_{64} <$ 3.180e-63
$c_{65} <$ 2.873e-64	$c_{66} <$ 2.498e-65	$c_{67} <$ 2.121e-66
$c_{68} <$ 1.760e-67	$c_{69} <$ 1.419e-68	$c_{70} <$ 1.109e-69

Table 2: The leading coefficients of ψ

If we assign the constants n_0, C and M the same values as in this lemma, it is plain to see that the expression appearing in (17) is decreasing in n_1, for $n_1 \geq 70$. Thus, if c_{71} satisfies the desired estimate, all coefficients of higher order will too. Considering the worst case, $p = 10$, we have

$$\begin{aligned} c_{71} &\leq \frac{5\left(2 \cdot 3.54 + 5(70-20) \cdot 10^{-6}\right)}{|(10-1)\lambda_1 + (71 - (10-1))\lambda_3 - \lambda_2|} CM^{71} \\ &\leq \frac{5\left(7.08 + 2.5 \cdot 10^{-4}\right)}{|9\lambda_1 + 62\lambda_3 - \lambda_2|} CM^{71} \leq 0.982 CM^{71}, \end{aligned}$$

which completes the induction. Hence, for $r < 9/5$, we arrive at

$$
\begin{aligned}
\psi(r) &\leq \sum_{j=2}^{10} c_j r^j + \sum_{j=11}^{\infty} C M^j r^j \leq r^2 \sum_{j=2}^{10} c_j r^{j-2} + 5 \cdot 10^{-6} \sum_{j=11}^{\infty} \Big(\frac{5r}{9}\Big)^j \\
&\leq r^2 \sum_{j=2}^{10} c_j r^{j-2} + 5 \cdot 10^{-6} \Big(\frac{5r}{9}\Big)^2 \sum_{j=9}^{\infty} \Big(\frac{5r}{9}\Big)^j \\
&\leq r^2 \sum_{j=2}^{10} c_j r^{j-2} + 5 \cdot 10^{-6} \Big(\frac{5}{9}\Big)^2 r^2 \frac{(5r/9)^9}{1 - 5r/9} \\
&\leq r^2 \sum_{j=2}^{10} c_j r^{j-2} + 1.4 \cdot 10^{-5} \frac{r^2}{9 - 5r}.
\end{aligned}
$$

As we will restrict our attention to the case $r < 1$, we can replace the estimate on $\psi(r)$ by

$$
\psi(r) \leq \Big(\sum_{j=2}^{10} c_j + \frac{1.4 \cdot 10^{-5}}{9 - 5} \Big) r^2 \leq \frac{r^2}{2}.
$$

This completes the proof of the first part of Proposition 3.1.

3.2.2 The normal form

Now, we turn our attention to the existence and size of the normal form $G(\zeta)$. Recall that G is defined recursively by

$$
G_i(\zeta) = \{F_i(\zeta + \phi(\zeta))\}_{\mathbb{U}_p} - \sum_{j=1}^{3} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta) G_j(\zeta) \qquad (i = 1, 2, 3). \tag{18}
$$

The existence of a formal solution is immediate as we have no divisors whatsoever. Turning to the question of convergence, we use similar majorization techniques as in the previous section. We arrive at the estimate $\|G\|_r \leq \hat{G}(r)$, by considering the following majorant systems together with (18):

$$
\tilde{G}(\zeta) = \Big\{ \tilde{F}(\zeta + \varphi(\zeta)) \Big\}_{\mathbb{U}_p} + \sum_{j=1}^{3} \frac{\partial \tilde{\varphi}}{\partial \zeta_j}(\zeta) \tilde{G}_j(\zeta)
$$

(recall that $\varphi = (\tilde{\varphi}, \tilde{\varphi}, \tilde{\varphi})$ and compare with (12)), and

$$
\begin{aligned}
\Big[\hat{G}(r)\Big]_n &= \Big[\hat{F}(r + \psi(r)) + 3\psi'(r)\hat{G}(r)\Big]_n \\
&= \Big[5\big(r + \psi(r)\big)^2 + 3\psi'(r)\hat{G}(r)\Big]_n \qquad (n \geq 2p),
\end{aligned} \tag{19}
$$

compare with (14). By the last equation, it is clear that $\hat{G}$ does not contain terms of degree less than $2p$, and that its lowest order term is given by

$$
\Big[\hat{G}(r)\Big]_{2p} = \Big[5\big(r + \psi(r)\big)^2\Big]_{2p}. \tag{20}
$$

As in the previous section, the recursive equation for $\hat{G}$ can be solved formally by a power series

$$
\hat{G}(r) = \sum_{k=20}^{\infty} \hat{g}_k r^k,
$$

and we will prove, by induction, that the coefficients of $\hat{G}$ satisfy $\hat{g}_n \le DK^n$, where we can take the constants $D = 2 \cdot 10^{-18}$ and $K = 3$. This will then immediately give

$$\|G\|_r \le \hat{G}(r) \le 2 \cdot 10^{-18} \sum_{k=20}^{\infty} (3r)^k \le 2 \cdot 10^{-18} \frac{(3r)^{20}}{1-3r} \le 7 \cdot 10^{-9} \frac{r^{20}}{1-3r},$$

which will complete the proof of Proposition 3.1.

Starting with (20), and using the computed numerical values of $\{c_j\}_1^{19}$, we have (recall that we defined $c_1 = 1$)

$$\hat{g}_{20} r^{20} = \left[\hat{G}(r)\right]_{20} \le \left[5(r + \psi(r))^2\right]_{20} = 5 \sum_{j=1}^{19} c_j c_{20-j} r^{20} \le 3 \cdot 10^{-13} r^{20}$$

which clearly satisfies our induction hypothesis, as $3 \cdot 10^{-13} < 2 \cdot 10^{-18} \cdot 3^{20} = DK^{20}$. We now proceed with the induction step: assume that we have proved that $\hat{g}_j \le DK^k$ for $j = 20, ..., n$. Then, by (19), we have

$$\hat{g}_{n+1} = 5 \sum_{j=1}^{n} c_j c_{n+1-j} + 3 \sum_{j=2}^{n-18} j c_j \hat{g}_{n+2-j} = S_1 + S_2,$$

where

$$\begin{aligned}
S_1 &\le 5 \left(2 \sum_{j=1}^{10} c_j M^{-j} + (n-20)C \right) C M^{n+1} \\
&\le \left[5 \left(2 \sum_{j=1}^{10} c_j M^{-j} + (n-20)C \right) \frac{C}{D} \left(\frac{M}{K} \right)^{n+1} \right] DK^{n+1} \\
&\le \left[5 \left(7.08 + (n-20) \cdot 5 \cdot 10^{-6} \right) \frac{5 \cdot 10^{-6}}{2 \cdot 10^{-18}} \left(\frac{5}{27} \right)^{n+1} \right] DK^{n+1} \\
&\le 0.0369 DK^{n+1},
\end{aligned}$$

and

$$\begin{aligned}
S_2 &\le 3D \left(\sum_{j=2}^{10} j c_j K^{n+2-j} + C \sum_{j=11}^{n-18} j M^j K^{n+2-j} \right) \\
&\le \left[3 \left(\sum_{j=2}^{10} j c_j K^{1-j} + CM \sum_{j=11}^{\infty} j \left(\frac{M}{K} \right)^{j-1} \right) \right] DK^{n+1} \\
&\le \left[3 \left(0.31 + \frac{25}{9} \cdot 10^{-6} \cdot \left(\frac{5}{27} \right)^{10} \frac{11 - 50/27}{(1 - 5/27)^2} \right) \right] DK^{n+1} \\
&\le 3 \left(0.31 + 2 \cdot 10^{-12} \right) DK^{n+1}.
\end{aligned}$$

Summing up, gives the following estimate

$$\hat{g}_{n+1} \le 0.97 DK^{n+1},$$

which proves the induction step. This completes the final part of Proposition 3.1.

3.3 The dynamics inside the cube

In this section, we will show that the normal form flow, $\psi(\zeta, t)$, i.e., the solution to the equations $\dot{\zeta} = A\zeta + G(\zeta)$, where G is as in Proposition 3.1, acts much like its linear counterpart used in the geometric model of the Lorenz flow. The geometric model uses the linearity near the origin to obtain estimates on trajectories passing near the origin. These estimates, however, are not valid for the original Lorenz flow without an analog of the change of variables described in Proposition 3.1. Finding a change of variables which completely linearizes the Lorenz equations in a neighbourhood of the origin poses two major difficulties:

- As we must remove all non-linear terms of the vector field, we will encounter all possible divisors, having modulus $|n\lambda - \lambda_i|$, where $n \in \mathbb{N}^3$ and $|n| \geq 2$. These are of course not bounded away from zero, so unless we impose a Diophantine condition[4] on the eigenvalues, the linearizing change of variables will not converge. Even if we manage to get a positive radius of convergence, it is likely that the radius is too small to be of any practical use.

- Although the set of all eigenvalues satisfying any Diophantine condition has full measure, the set of resonant eigenvalues[5] is everywhere dense. As resonant eigenvalues produce vanishing divisors, we must exclude a dense set of parameters to avoid this situation. In doing so, we lose the robustness of our statements.

This is why we bring the Lorenz equations into a carefully selected normal form instead of the linear one. The price we have to pay is that it is a little more difficult to gain estimates on the normal form flow. On the other hand, once we have these estimates, we may effectively link them to the original Lorenz flow by using Proposition 3.1. As mentioned in the outline, we will make the change of coordinates when a trajectory hits the lid of a small cube centered around the origin. Inside the cube, we use our estimates on the normal form flow to find the exit point of the trajectory. We then change back to the original coordinates and carry on with the numerical computations.

3.4 The linear flow

In this brief section, we will extract the properties of the linear flow that really are used in the geometric model of the Lorenz flow. Let $\phi(\zeta, t)$ denote the flow of the linearized Lorenz equations $\dot{\zeta} = A\zeta$. These can be solved explicitly: $\phi(\zeta, t) = e^{At}\zeta$, i.e., $\phi_i(\zeta, t) = e^{\lambda_i t}\zeta_i$ $(i = 1, 2, 3)$. Consider a small cube centered at the origin, with radius r, and take a trajectory starting from the interior of the cube, say at the point ζ, with $|\zeta_1| \neq 0$. Since the eigenvalues satisfy $0 < -\lambda_3 < \lambda_1 < -\lambda_2$, it is plain to see that the trajectory exits the cube when $\phi_1(\zeta, t) = \mathrm{sgn}(\zeta_1)r$. Solving for the exit-time $\tau_e(\zeta)$ gives $\tau_e(\zeta) = \frac{1}{\lambda_1}\log\frac{r}{|\zeta_1|}$. Note that $\lim_{\zeta_1 \to 0}\tau_e(\zeta) = \infty$. This is one of the reasons why numeric calculations break down near the origin. Inserting τ_e in the other coordinate functions gives the location of the exit

$$\phi(\zeta, \tau_e(\zeta)) = \left(\mathrm{sgn}(\zeta_1)r, \zeta_2\left(|\zeta_1|/r\right)^{-\frac{\lambda_2}{\lambda_1}}, r\left(|\zeta_1|/r\right)^{-\frac{\lambda_3}{\lambda_1}}\right). \tag{21}$$

Since $-\lambda_2/\lambda_1 \approx 1.93$ and $-\lambda_3/\lambda_1 \approx 0.225$, a line segment lying in the ζ_2 direction will be strongly contracted, whereas a line segment lying in the ζ_1 direction will have expanded on its exit. An explicit calculation gives

$$\frac{\phi_2(\zeta, t)}{\phi_3(\zeta, t)} = \frac{\zeta_2}{r}e^{(\lambda_2 - \lambda_3)t}.$$

Therefore, since $\lambda_2 - \lambda_3 < 0$, the lid of the cube will exit as two cusp shaped regions (in Figure 6 one of the cusps is illustrated). In the cube, the $\zeta_1\zeta_3$-plane acts as a separatrix, and all trajectories approach this exponentially fast.

[4]We say that the eigenvalues $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ satisfy a Diophantine condition of type (κ, τ) if there exists positive κ, τ such that for $i = 1, 2, 3$, we have $|n\lambda - \lambda_i| \geq \kappa|n|^{-\tau}$ for all $n \in \mathbb{N}^3$ with $|n| \geq 2$.

[5]We say that the eigenvalues $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ are resonant if there exists an $i \in \{1, 2, 3\}$ such that $|n\lambda - \lambda_i| = 0$ for some $n \in \mathbb{N}^3$ with $|n| \geq 2$.

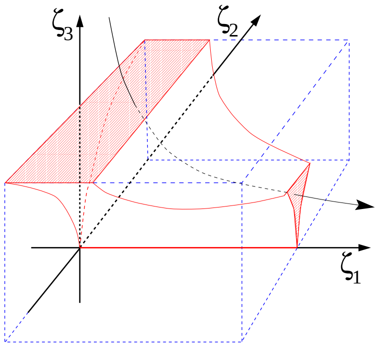

Figure 6: The linear flow inside the cube for $\zeta_1 > 0$.

We are also interested in the evolution of tangent vectors following a trajectory inside the cube. Since $\phi(\zeta,t) = e^{At}\zeta$, it is clear that $D\phi(\zeta,t) = e^{At}$. Hence we have

$$D\phi(\zeta,\tau_e(\zeta)) = \begin{pmatrix} (|\zeta_1|/r)^{-1} & 0 & 0 \\ 0 & (|\zeta_1|/r)^{-\frac{\lambda_2}{\lambda_1}} & 0 \\ 0 & 0 & (|\zeta_1|/r)^{-\frac{\lambda_3}{\lambda_1}} \end{pmatrix}.$$

Any three-dimensional cone field centered around $e_1 = (1,0,0)$ is taken into itself under $D\phi$. This is due to the fearsome hyperbolicity (expansion and contraction) $D\phi$ exhibits near the origin. This property and (21), which gives rise to the cusps, are the two features used in the geometric models.

3.5 The normal form flow

In this section, we will show that the normal form flow and the linear flow have very similar behaviour. Although we use a cube having radius $\frac{1}{5}$ in the computer program, we do all estimates for a radius of $\frac{1}{4}$. This allows for a slight deformation of the cube as we change coordinates, see Lemma 3.4. We will need the following estimate on G:

Lemma 3.7 *In the cube $\{\zeta\colon |\zeta| \le \frac{1}{4}\}$, we have*

$$|G(\zeta)| \le 3\cdot 10^{-8}|\zeta_1|^{10}\max\{|\zeta_2|^{10}, |\zeta_3|^{10}\}.$$

Proof: Recall that we arranged for $G(\zeta) \in \mathcal{O}^{10}(\zeta_1) \cap \mathcal{O}^{10}(\zeta_2,\zeta_3)$. This means that for any term $g_{i,n}\zeta^n$ of G_i, there exist $\tilde{n} \in \mathbb{N}^3$ and $k \in [0,10]$ such that we can factor the term as

$$g_{i,n}\zeta^n = g_{i,n}\zeta_1^{10}\zeta_2^{10-k}\zeta_3^k\zeta^{\tilde{n}}. \tag{22}$$

The estimate on G in Proposition 3.1 implies that $|G(\zeta)| \le 3\cdot 10^{-8}|\zeta|^{20}$ for $|\zeta| \le \frac{1}{4}$, but we also have some additional information: since the estimate was obtained by using majorants on two levels ($\mathbb{C}^3 \to \mathbb{C}$ and $\mathbb{C} \to \mathbb{R}$), it is also valid for the majorants themselves. The estimate is naturally valid for all smaller majorants of G and, in particular, for the smallest majorants of G. Let

$$G_i(\zeta) = \sum_{\substack{n_1 \ge 10 \\ n_2+n_3 \ge 10}} g_{i,n}\zeta^n, \quad m(\zeta) = \sum_{\substack{n_1 \ge 10 \\ n_2+n_3 \ge 10}} m_n\zeta^n, \quad \text{and} \quad M(r) = \sum_{k\ge 20} M_k r^k,$$

where $m_n = \max_{i=1,2,3}|g_{i,n}|$, and $M_k = [m(r,r,r)]_k$. Then clearly $G_i \prec m \prec M$, and the functions m and M are the smallest majorants of G on their corresponding levels. Now, set $\omega = |\zeta_1|\max\{|\zeta_2|,|\zeta_3|\}$, and suppose that $|\zeta| \le \frac{1}{4} = r$. We have

$$\begin{aligned}
|G_i(\zeta)| &= \Big| \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} g_{i,n}\zeta^n \Big| \leq \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} |g_{i,n}||\zeta_1|^{n_1}|\zeta_2|^{n_2}|\zeta_3|^{n_3} \\
&\leq \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} m_n|\zeta_1|^{n_1}|\zeta_2|^{n_2}|\zeta_3|^{n_3} \leq \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} m_n|\zeta_1|^{n_1}\big(\max\{|\zeta_2|,|\zeta_3|\}\big)^{n_2+n_3} \\
&\leq |\zeta_1|^{10}\big(\max\{|\zeta_2|,|\zeta_3|\}\big)^{10} \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} m_n|\zeta_1|^{n_1-10}\big(\max\{|\zeta_2|,|\zeta_3|\}\big)^{n_2+n_3-10} \\
&\leq \omega^{10} \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} m_n\big(\max\{|\zeta_1|,|\zeta_2|,|\zeta_3|\}\big)^{n_1+n_2+n_3-20} \\
&= \omega^{10} \sum_{\substack{n_1 \geq 10 \\ n_2+n_3 \geq 10}} m_n|\zeta|^{|n|-20} \leq \omega^{10} \sum_{k \geq 20} \Big(\sum_{|n|=k} m_n\Big) r^{k-20} = \omega^{10} \sum_{k \geq 20} M_k r^{k-20} \\
&= \omega^{10} r^{-20} \sum_{k \geq 20} M_k r^k \leq \omega^{10} r^{-20} \cdot 3 \cdot 10^{-8} r^{20} = 3 \cdot 10^{-8}\omega^{10},
\end{aligned}$$

which completes the proof. □

3.5.1 C^0-properties

In this section we will examine how the orbits of the linear and normal form flows differ. Using Lemma 3.7, we shall first prove that the modulus of the ζ_1-component of the normal form flow is monotonically increasing.

Lemma 3.8 *In the cube $\{\zeta \colon |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, we have*

$$(\lambda_1 - \kappa)\psi_1(\zeta, t) \leq \dot{\psi}_1(\zeta, t) \leq (\lambda_1 + \kappa)\psi_1(\zeta, t)$$

and

$$|\zeta_1|e^{(\lambda_1 - \kappa)t} \leq |\psi_1(\zeta, t)| \leq |\zeta_1|e^{(\lambda_1 + \kappa)t},$$

where $\kappa = 2 \cdot 10^{-19}$.

Proof: We just have to check the differential equation for the ζ_1-component of the normal form flow: $\dot{\psi}_1(\zeta, t) = \lambda_1\psi_1(\zeta, t) + G_1(\psi(\zeta, t))$. By Lemma 3.7,

$$\begin{aligned}
|\dot{\psi}_1(\zeta, t) - \lambda_1\psi_1(\zeta, t)| &= |G_1(\psi(\zeta, t))| \\
&\leq 3 \cdot 10^{-8}|\psi_1(\zeta, t)|^{10} \max\{|\psi_2(\zeta, t)|^{10}, |\psi_3(\zeta, t)|^{10}\} \\
&\leq 3 \cdot 10^{-8} \cdot 4^{-10}|\psi_1(\zeta, t)|^{10} \leq 3 \cdot 10^{-8} \cdot 4^{-19}|\psi_1(\zeta, t)|.
\end{aligned}$$

Setting $\kappa = 2 \cdot 10^{-19} > 3 \cdot 10^{-8} \cdot 4^{-19}$ completes the proof. □

Next, we prove that the ζ_3-component of the normal form flow dominates the ζ_2-component.

Lemma 3.9 *For all trajectories starting from the lid of the cube $\{\zeta \colon |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, we have $\psi_3(\zeta, t) \geq |\psi_2(\zeta, t)|$. These estimates hold throughout the cube.*

Proof: The differential equations for ψ_2 and ψ_3 can be expressed as differential inequalities:

$$\begin{aligned}
|\dot{\psi}_2(\zeta, t) - \lambda_2\psi_2(\zeta, t)| &= |G_2(\psi(\zeta, t))| \leq 3 \cdot 10^{-8} \cdot 4^{-10} \max\{|\psi_2(\zeta, t)|^{10}, |\psi_3(\zeta, t)|^{10}\} \\
|\dot{\psi}_3(\zeta, t) - \lambda_3\psi_3(\zeta, t)| &= |G_3(\psi(\zeta, t))| \leq 3 \cdot 10^{-8} \cdot 4^{-10} \max\{|\psi_2(\zeta, t)|^{10}, |\psi_3(\zeta, t)|^{10}\}.
\end{aligned}$$

Initially, we have $\frac{1}{4} = \psi_3(\zeta, 0) \geq |\psi_2(\zeta, 0)|$, and by the differential inequalities it is clear that, if $\psi_3(\zeta, 0) = |\psi_2(\zeta, 0)|$, then $|\psi_2(\zeta, t)|$ decreases faster than $\psi_3(\zeta, t)$, since $\lambda_2 < \lambda_3 < 0$. So suppose that after some time t^* we have $\psi_3(\zeta, t^*) = |\psi_2(\zeta, t^*)|$. Then we can rewrite the differential inequalities as

$$|\dot{\psi}_2(t^*, \zeta) - \lambda_2\psi_2(t^*, \zeta)| \leq 3 \cdot 10^{-8} \cdot 4^{-10}|\psi_2(t^*, \zeta)|^{10} \leq \kappa|\psi_2(t^*, \zeta)|$$
$$|\dot{\psi}_3(t^*, \zeta) - \lambda_3\psi_3(t^*, \zeta)| \leq 3 \cdot 10^{-8} \cdot 4^{-10}|\psi_3(t^*, \zeta)|^{10} \leq \kappa|\psi_3(t^*, \zeta)|,$$

and, again, $|\psi_2(t^*, \zeta)|$ decreases faster than $\psi_3(t^*, \zeta)$. Hence $\psi_3(\zeta, t) \geq |\psi_2(\zeta, t)|$ throughout the whole cube. □

It is now easy to show that the ζ_3-component of the flow is monotonically decreasing:

Lemma 3.10 *For all trajectories starting from the lid of the cube $\{\zeta : |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, we have*

$$(\lambda_3 - \kappa)\psi_3(\zeta, t) \leq \dot{\psi}_3(\zeta, t) \leq (\lambda_3 + \kappa)\psi_3(\zeta, t)$$

and

$$re^{(\lambda_3 - \kappa)t} \leq \psi_3(\zeta, t) \leq re^{(\lambda_3 + \kappa)t},$$

where $\kappa = 2 \cdot 10^{-19}$.

Proof: Using Lemma 3.9, we just copy the proof of Lemma 3.8 by changing the roles of $\psi_1(\zeta, t)$ and $\psi_3(\zeta, t)$. □

Now, given an initial point in the lid of the cube $\{\zeta : |\zeta| \leq r\}$ we know that $|\psi_1(\zeta, t)|$ is increasing, whereas $\psi_3(\zeta, t)$ is decreasing and majorizing $|\psi_2(\zeta, t)|$. Since $\lambda_1 > |\lambda_3|$, we also know that $|\psi_1(\zeta, t)|$ increases faster than $\psi_3(\zeta, t)$ decreases. Therefore, it is clear that the orbit exits the cube through one of the sides $\{\zeta : |\zeta_1| = r, |\zeta_2|, |\zeta_3| \leq r\}$. A trivial calculation enables us to control the exit-time of the normal form flow:

Lemma 3.11 *For all trajectories starting from the lid of the cube $\{\zeta : |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, the normal form flow exits the cube at time $\tau_e(\zeta)$, where*

$$\frac{1}{\lambda_1 + \kappa} \log \frac{r}{|\zeta_1|} \leq \tau_e(\zeta) \leq \frac{1}{\lambda_1 - \kappa} \log \frac{r}{|\zeta_1|},$$

and $\kappa = 2 \cdot 10^{-19}$.

Proof: We just have to solve $|\psi_1(\zeta, \tau_e)| = r$ for τ_e. Using Lemma 3.8, we get

$$|\zeta_1|e^{(\lambda_1 - \kappa)\tau_e} \leq r \leq |\zeta_1|e^{(\lambda_1 + \kappa)\tau_e},$$

which immediately gives the desired result. □

Note that we have $\lim_{\zeta_1 \to 0} \tau_e(\zeta) = \infty$, just as in the linear case.

Turning to the ζ_2-component of the normal form flow, we have the following differential inequality

$$|\dot{\psi}_2(\zeta, t) - \lambda_2\psi_2(\zeta, t)| \leq 3 \cdot 10^{-8} \cdot r^{10}|\psi_3(\zeta, t)|^{10}$$

for all trajectories starting from the lid of the cube $\{\zeta : |\zeta| \leq r\}$. This does not imply that $|\psi_2(\zeta, t)|$ decreases exponentially, so there is a slight discrepancy between the normal form and linear flows in this sense. However, just as in the linear case, there exists a surface acting as a separatrix. It is a slight deformation of the $\zeta_1\zeta_3$-plane, and all orbits tend to this separatrix exponentially fast. The important property of both the normal form and linear flows is that the quotient of the ζ_2- and ζ_3-components tends to zero exponentially fast with respect to the exit-time. This gives rise to the nice cusp shaped image of the cube's lid, as illustrated in Figure 6.

Lemma 3.12 *For all trajectories starting from the lid of the cube $\{\zeta : |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, we have*

$$\big(\zeta_2 - \kappa r(1 - e^{-3t})\big)e^{\lambda_2 t} \leq \psi_2(\zeta, t) \leq \big(\zeta_2 + \kappa r(1 - e^{-3t})\big)e^{\lambda_2 t},$$

where $\kappa = 2 \cdot 10^{-19}$. These estimates hold throughout the cube.

Proof: Using the second part of Lemma 3.10, we can rewrite the differential inequality for $\dot{\psi}_2(\zeta, t)$ as

$$|\dot{\psi}_2(\zeta,t) - \lambda_2\psi_2(\zeta,t)| \leq 3 \cdot 10^{-8} \cdot 4^{-19} r e^{10(\lambda_3+\kappa)t} \leq \kappa r e^{10(\lambda_3+\kappa)t}.$$

A straight-forward calculation yields the following bounds on $\psi_2(\zeta, t)$:

$$|\psi_2(\zeta,t) - \zeta_2 e^{\lambda_2 t}| \leq \frac{\kappa r\big(e^{\lambda_2 t} - e^{10(\lambda_3+\kappa)t}\big)}{\big||\lambda_2| + 10(\lambda_3+\kappa)\big|} \leq \kappa r(1 - e^{-3t})e^{\lambda_2 t},$$

using the fact that $10(\lambda_3 + \kappa) - \lambda_2 < -3$. □

Combining this lemma with the second part of Lemma 3.10 gives

$$\left|\frac{\psi_2(\zeta,t)}{\psi_3(\zeta,t)}\right| \leq \frac{|\zeta_2| + \kappa r(1 - e^{-3t})}{r e^{(\lambda_3-\kappa)t}} e^{\lambda_2 t} \leq (1+\kappa)e^{(\lambda_2-\lambda_3+\kappa)t},$$

which proves that the lid of the cube will exit as two cusp shaped region.

Remark: Had we flattened out the invariant manifolds to an order $p < 9$, the situation would be slightly different: We would then have $p(\lambda_3 + \kappa) - \lambda_2 > 0$, which would result in a blunter cusp.

3.5.2 C^1-properties

We will now prove that, in our small cube, the normal form flow expands and contracts tangent vectors at almost the same rate as the linear flow does. The first variational equations for the normal form flow are

$$\frac{d}{dt}D\psi(\zeta,t) = AD\psi(\zeta,t) + DG(\psi(\zeta,t))D\psi(\zeta,t), \qquad D\psi(\zeta,0) = I,$$

and the solution is formally given by

$$D\psi(\zeta,t) = e^{At}D\psi(\zeta,0) + \int_0^t e^{A(t-s)}DG(\psi(\zeta,s))D\psi(\zeta,s)ds.$$

We will prove the following lemma, which states that $D\psi$ is virtually linear:

Lemma 3.13 *For all trajectories starting from the lid of the cube $\{\zeta\colon |\zeta| \leq r\}$, where $r \leq \frac{1}{4}$, we have*

$$\left|\frac{\partial\psi_i}{\partial\zeta_j}(\zeta,t) - \frac{\partial\phi_i}{\partial\zeta_j}(\zeta,t)\right| \leq \kappa e^{[9(\lambda_3+\kappa)+\lambda_j]t} \qquad (i,j = 1,2,3),$$

where $\kappa = 2 \cdot 10^{-19}$. These estimates hold throughout the cube.

Remark: The key point is that $9(\lambda_3 + \kappa) + \lambda_j$ is negative. This means that the error decreases as the exit-time increases, i.e., as we take $|\zeta_1|$ small. This is expected seeing how we constructed G: the perturbation of the linearized equations is small enough to completely dampen out the bad effects due to having to spend a long time near the origin.

Proof: Since $D\psi(\zeta, 0) = D\phi(\zeta, 0) = I$, the inequality is trivially true for small t, say for $t \in [0, t^*]$. We will prove that we can take $t^* = \tau_e(\zeta)$, which will prove the lemma. For $t \in [0, t^*]$, we have

$$\begin{aligned} |D\psi(\zeta,t) - D\phi(\zeta,t)| &= \left|\int_0^t e^{A(t-s)}DG(\psi(\zeta,s))D\psi(\zeta,s)ds\right| \qquad (23)\\ &\leq \int_0^t \left|e^{A(t-s)}DG(\psi(\zeta,s))D\phi(\zeta,s)\right| ds \\ &+ \int_0^t \left|e^{A(t-s)}DG(\psi(\zeta,s))\big(D\psi(\zeta,s) - D\phi(\zeta,s)\big)\right| ds. \end{aligned}$$

By using the facts that A is diagonal, and $D\phi(\zeta, s) = e^{As}$, a simple calculation gives that

$$\left|\left\{e^{A(t-s)}DG(\zeta)e^{As}\right\}_{ij}\right| = e^{\lambda_i(t-s)}e^{\lambda_j s}\left|\frac{\partial G_i}{\partial \zeta_j}(\zeta)\right| \qquad (i,j=1,2,3).$$

If we set $E(\zeta,t) = D\psi(\zeta,t) - D\phi(\zeta,t)$, we can treat each matrix entry separately: For $i,j=1,2,3$, we have

$$\begin{aligned}|E_{ij}(\zeta,t)| \le \int_0^t & e^{\lambda_i(t-s)}e^{\lambda_j s}\left|\frac{\partial G_i}{\partial \zeta_j}(\psi(\zeta,s))\right| ds \\ &+ \int_0^t e^{\lambda_i(t-s)}\sum_{k=1}^{3}\left|\frac{\partial G_i}{\partial \zeta_k}(\psi(\zeta,s))E_{kj}(\zeta,s)\right| ds.\end{aligned}$$

To carry on further, it is clear that we need some estimates on the partial derivatives of G.

Lemma 3.14 *In the cube* $\{\zeta\colon |\zeta| \le \frac{1}{4}\}$, *we have*

$$\left|\frac{\partial G_i}{\partial \zeta_j}(\zeta)\right| \le 9\cdot 10^{-8}|\zeta_1|^9\max\{|\zeta_2|^9, |\zeta_3|^9\}.$$

Proof: By Proposition 3.1, the functions G_i $(i=1,2,3)$ can be extended to a ball in $\mathbb{C}^3$ such that they are analytic in each of the three variables ζ_i $(i=1,2,3)$. Furthermore, for $|\zeta| \le \frac{3}{10}$, the functions satisfy

$$|G_i(\zeta)| \le \frac{7\cdot 10^{-9}|\zeta|^{20}}{1-\frac{9}{10}} = 7\cdot 10^{-8}|\zeta|^{20},$$

and the argument used in the proof of Lemma 3.2 shows that in the complex ball $|\zeta| \le \frac{1}{4}$, we have the estimate

$$\left|\frac{\partial G_i}{\partial \zeta_j}(\zeta)\right| \le \frac{7\cdot 10^{-8}|\zeta|^{20}}{\frac{3}{10}-\frac{1}{4}} = 1.4\cdot 10^{-6}|\zeta|^{20} < 9\cdot 10^{-8}|\zeta|^{18} \qquad (i,j=1,2,3). \tag{24}$$

Since we arranged for $G(\zeta) \in \mathcal{O}^{10}(\zeta_1) \cap \mathcal{O}^{10}(\zeta_2,\zeta_3)$, we know that terms of the partial derivatives of G_i belong to $\mathcal{O}^9(\zeta_1) \cap \mathcal{O}^9(\zeta_2,\zeta_3)$. Thus, for any term $g_{i,n}\zeta^n$ of a partial derivative of G_i, there exists $\tilde{n} \in \mathbb{N}^3$ and $k \in [0,9]$ such that we can factor the term as

$$g_{i,n}\zeta^n = g_{i,n}\zeta_1^9\zeta_2^{9-k}\zeta_3^k\zeta^{\tilde{n}}. \tag{25}$$

Applying the methods used in the proof of Lemma 3.7 combined with the estimate (24) gives the desired result. □

From the previous section, we know that $|\psi_2(\zeta,t)| \le \psi_3(\zeta,t)$ for all trajectories starting from the lid of the cube. Using this with Lemmas 3.8 and 3.10, gives

$$\left|\frac{\partial G_i}{\partial \zeta_j}(\psi(\zeta,t))\right| \le 9\cdot 10^{-8}|\psi_1(\zeta,t)\psi_3(\zeta,t)|^9 \le 4\cdot 10^{-13}|\zeta_1|^9 e^{9(\lambda_1+\lambda_3+2\kappa)t}.$$

We are now prepared to continue our estimates. We will use the facts that $|E_{kj}(\zeta,s)| \le \kappa e^{[9(\lambda_3+\kappa)+\lambda_j]s} \le \kappa e^{\lambda_j s}$ and $|\zeta_1 e^{(\lambda_1+\kappa)t}| \le r \le \frac{1}{4}$. For $i,j = 1,2,3$, we have

$$
\begin{aligned}
\left|E_{ij}(\zeta,t)\right| \le & \int_0^t e^{\lambda_i(t-s)} e^{\lambda_j s} \left|\frac{\partial G_i}{\partial \zeta_j}(\psi(\zeta,s))\right| ds \\
& + \int_0^t e^{\lambda_i(t-s)} \sum_{k=1}^{3} \left|\frac{\partial G_i}{\partial \zeta_k}(\psi(\zeta,s))\right| \kappa e^{\lambda_j s} ds \\
\le & \; 4\cdot 10^{-13}|\zeta_1|^9(1+3\kappa)e^{\lambda_i t} \int_0^t e^{[9(\lambda_1+\lambda_3+2\kappa)+\lambda_j-\lambda_i]s} ds \\
= & \; 4\cdot 10^{-13}|\zeta_1|^9(1+3\kappa)e^{\lambda_i t} \left[\frac{e^{[9(\lambda_1+\lambda_3+2\kappa)+\lambda_j-\lambda_i]s}}{9(\lambda_1+\lambda_3+2\kappa)+\lambda_j-\lambda_i}\right]_0^t \\
\le & \; \frac{4\cdot 10^{-13}|\zeta_1|^9(1+3\kappa)}{9(\lambda_1+\lambda_3)+\lambda_2-\lambda_1} e^{[9(\lambda_1+\lambda_3+2\kappa)+\lambda_j]t} \\
\le & \; \frac{4\cdot 10^{-13}(1+3\kappa)}{47} |\zeta_1 e^{(\lambda_1+\kappa)t}|^9 e^{[9(\lambda_3+\kappa)+\lambda_j]t} \\
\le & \; \frac{4\cdot 10^{-13}(1+3\kappa)}{47\cdot 4^9} e^{[9(\lambda_3+\kappa)+\lambda_j]t} \le \kappa e^{[9(\lambda_3+\kappa)+\lambda_j]t},
\end{aligned}
$$

for $\kappa = 2\cdot 10^{-19}$. This completes the proof, since all estimates are valid in the cube. $\square$

4 Rigorous numerics - details

In this section we will derive the mathematics used in the main computer program `lorenz_2.1.c` described in the outline. Its core consists of five functions, `Stage0`, … ,`Stage4`, which are tailor-made to suit certain regions of phase-space. To be more specific, we split each orbit under consideration into five segments, depending on the relative sizes of $|f_1(x)|$ and $|f_3(x)|$, see Figure 7 (recall that $f(x)$ is right-hand side of the Lorenz equations (3)). In the even-numbered sectors, $|f_3(x)|$

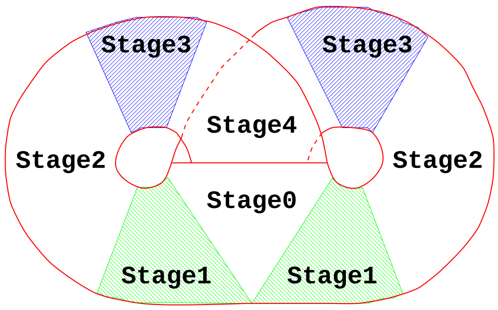

Figure 7: The domains of the five `Stage`-functions.

dominates $|f_1(x)|$, and trajectories are either traveling upwards or downwards. Hence x_1x_2-planes are transversal to the flow in these regions, and may be used as intermediate Poincaré planes. In the odd-numbered sectors, x_2x_3-planes are used instead.

In the following sections, we will describe how to follow a sub-rectangle of the forward invariant set taken along the flow. This is done by controlling the error growth of the rectangle and its associated cone field when passing from one intermediate Poincaré plane to another. We will also describe how to switch from x_1x_2-planes to x_2x_3-planes and vice-versa. The computer program composes these elementary algorithms, and thus produces a global return map as desired.

4.1 Surface propagation

Consider a rectangle[6] $\mathcal{R}^{(k)} = [x_1^-, x_1^+] \times [x_2^-, x_2^+] \times \{x_3^{(k)}\}$, and suppose that it is transversal to the flow of the Lorenz equations. This means that there is no point $x \in \mathcal{R}^{(k)}$ such that $f_3(x) = 0$, and hence the flow is either traveling downwards or upwards through $\mathcal{R}^{(k)}$. For the moment, we assume that it is traveling downwards. By continuity, the flow will pass through a plane $\Sigma^{(k+1)} = \{x : x_3 = x_3^{(k+1)}\}$ situated slightly beneath $\mathcal{R}^{(k)}$. The question we now pose is the following: how do we estimate the rectangular hull $\mathcal{R}^{(k+1)}$ of the flow of $\mathcal{R}^{(k)}$ passing through $\Sigma^{(k+1)}$? The problem is illustrated in Figure 8.

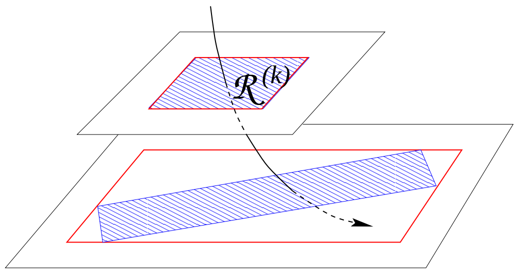

Figure 8: Finding the rectangular hull of a propagated surface.

[6]We always demand that the sides of a rectangle be parallel to the coordinate axes.

If we denote the flow by $\varphi(x,t)$, we may define a local Poincaré map $\Pi\colon \mathcal{R}^{(k)} \to \Sigma^{(k+1)}$ by

$$\Pi(x) = \big(\Pi_1(x), \Pi_2(x)\big) = \big(\varphi_1(x,t(x)), \varphi_2(x,t(x))\big),$$

where $t(x)$ is the solution to $\varphi_3(x,t(x)) = x_3^{(k+1)}$. We will sometimes view $x_3^{(k+1)}$ and $x_3^{(k)}$ as being fixed, and consider the Poincaré map to be a function of two variables.

The first problem now is to find bounds on the various flow times. For all $x \in \mathcal{R}^{(k)}$, we want to find $t = t(x)$ such that $\varphi_3(x,t) = x_3^{(k+1)}$. By the mean value theorem, we know that for $i = 1,2,3$, we have

$$\varphi_i(x,t) = x_i + t f_i(\varphi(x,\tau_i)), \tag{26}$$

for some $\tau_i \in [0,t]$. If we define $h = x_3^{(k)} - x_3^{(k+1)}$, and solve the third equation of (26) for t, we get

$$t = \frac{\varphi_3(x,t) - x_3^{(k)}}{f_3(\varphi(x,\tau_3))} = \frac{-h}{f_3(\varphi(x,\tau_3))}.$$

Inserting this into the two remaining equations of (26), we can solve for the components of the Poincaré map:

$$\Pi_i(x) = x_i - h\frac{f_i(\varphi(x,\tau_i))}{f_3(\varphi(x,\tau_3))} = x_i - h\frac{f_i(p_i)}{f_3(p_3)} \qquad (i = 1,2),$$

where p_i are points situated somewhere along the trajectory connecting x and $\Pi(x)$. This means that if we can find a bounding box B containing all trajectories between $\mathcal{R}^{(k)}$ and $\Sigma^{(k+1)}$, we can estimate the image of $\mathcal{R}^{(k)}$ under Π by Taylor-expanding the vector field in B. We make an (educated) guess of the size of B by approximating the image of the center of $\mathcal{R}^{(k)}$ under Π. We then compute the approximate drift in the x_1- and x_2-variables:

$$\delta x_i = h\left|\frac{f_i(x)}{f_3(x)}\right| \approx |x_i - \Pi_i(x)| \qquad (i = 1,2).$$

Due to proximity, all other trajectories starting from $\mathcal{R}^{(k)}$ will have roughly the same drift which we can bound by slightly increasing the δx_i's. Still calling the increased drifts $\delta x_1, \delta x_2$, we define the bounding box by

$$B = [x_1^- - \delta x_1, x_1^+ + \delta x_1] \times [x_2^- - \delta x_2, x_2^+ + \delta x_2] \times [x_3^{(k+1)}, x_3^{(k)}].$$

For simplicity, we will use the notation $B = (\bar{x}_1, \bar{x}_2, \bar{x}_3) \pm (\Delta x_1, \Delta x_2, \Delta x_3)$.

In the program, we have fixed $\Delta x_3 \equiv \frac{1}{2000}$, whereas for $i = 1,2$, Δx_i are bounded by `MaxSize` which varies between 0.010 and 0.070. By choosing Δx_3 small we get small flow times, which is good for computational reasons. At the same time, we do not want to have to compute the propagation of too many rectangles. Therefore, we take Δx_i as large as possible for $i = 1,2$.

4.1.1 The generic case

Generically, the trajectories maximizing the rectangular hull of $\mathcal{R}^{(k+1)}$ originate from the corners of $\mathcal{R}^{(k)}$ as illustrated in Figure 8. There are, however, rare occasions when this is not the case. The exact conditions for this special situation to occur will be given later. We will then also show that in these cases the trajectories must at least originate from certain parts of the boundary of $\mathcal{R}^{(k)}$.

Assuming the generic case for now, the maximizing trajectories are confined to small sub-boxes $B_1, \dots, B_4$ with lids centered on the corners of $\mathcal{R}^{(k)}$, and extending down to $\Sigma^{(k+1)}$, see Figure 9. We index the four boxes as we do with quadrants. Thus the two coloured boxes in the figure are given by $B_4 = (x_1^+, x_2^-, \bar{x}_3) \pm (\delta x_1, \delta x_2, \Delta x_3)$ and $B_1 = (x_1^+, x_2^+, \bar{x}_3) \pm (\delta x_1, \delta x_2, \Delta x_3)$, and the two remaining boxes are given by switching x_1^+ for x_1^-.

We organize our computations as follows:

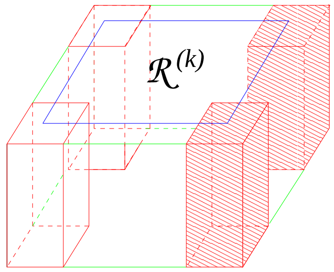

Figure 9: The generic case.

- In B_1 we compute $y_1^1 \geq \max_{B_1} \Pi_1(x_1^+, x_2^+)$ and $y_2^1 \geq \max_{B_1} \Pi_2(x_1^+, x_2^+)$,
- In B_2 we compute $y_1^2 \leq \min_{B_2} \Pi_1(x_1^-, x_2^+)$ and $y_2^2 \geq \max_{B_2} \Pi_2(x_1^-, x_2^+)$,
- In B_3 we compute $y_1^3 \leq \min_{B_3} \Pi_1(x_1^-, x_2^-)$ and $y_2^3 \leq \min_{B_3} \Pi_2(x_1^-, x_2^-)$,
- In B_4 we compute $y_1^4 \geq \max_{B_4} \Pi_1(x_1^+, x_2^-)$ and $y_2^4 \leq \min_{B_4} \Pi_2(x_1^+, x_2^-)$,

If we set $y^i = (y_1^i, y_2^i, x_3^{(k+1)})$, for $i = 1, 2, 3, 4$, it is clear that the rectangular hull of these four points contains the flow of $\mathcal{R}^{(k)}$ passing through $\Sigma^{(k+1)}$.
As an example, we will show how to compute y_1^1 and y_2^1. The six remaining coordinates are treated analogously. We will make the assumption that the trajectory starting at $(x_1^+, x_2^+, x_3^{(k)})$ is contained in B_1. If we set

$$f_3^- = \min_{B_1} f_3(x) \quad \text{and} \quad f_3^+ = \max_{B_1} f_3(x)$$

we have the following estimates on the flow times:

$$t^- = \frac{-h}{f_3^-} \leq t(x) \leq \frac{-h}{f_3^+} = t^+.$$

For $i = 1, 2$, we analyze the situation a bit more carefully: if we set $t = t(x)$ we have

$$\Pi_i(x) = x_i + t f_i(x) + \frac{t^2}{2} \dot{f}_i(q_i),$$

where $\dot{f}_i = \frac{d}{dt} f_i$, and q_i is a point on the trajectory from x to $\Pi(x)$. Set

$$\Delta \dot{f}_i = \sum_{j=1}^{2} \delta x_j \max_{B_1} \left| \frac{\partial \dot{f}_i}{\partial x_j}(x) \right| + \Delta x_3 \max_{B_1} \left| \frac{\partial \dot{f}_i}{\partial x_3}(x) \right|. \tag{27}$$

Then for $i = 1, 2$, we have the estimates

$$\Pi_i(x) \leq A_i^- = x_i + t^- f_i(x) + \frac{(t^+)^2}{2}(|\dot{f}_i(x_c)| + \Delta \dot{f}_i) \quad \text{if} \quad f_i(x) \leq 0,$$
$$\Pi_i(x) \leq A_i^+ = x_i + t^+ f_i(x) + \frac{(t^+)^2}{2}(|\dot{f}_i(x_c)| + \Delta \dot{f}_i) \quad \text{if} \quad f_i(x) \geq 0.$$

If we define $y_1^1 = \max\{A_1^-, A_1^+\}$ and $y_2^1 = \max\{A_2^-, A_2^+\}$, we clearly have $y_1^1 \geq \max_{B_1} \Pi_1(x_1^+, x_2^+)$ and $y_2^1 \geq \max_{B_1} \Pi_2(x_1^+, x_2^+)$ as desired. To conclude our example, we need to verify that our assumption (that the trajectory was confined to B_1) was true. This is done by checking that $|x_1^+ - y_1^1| \leq \delta x_1$ and $|x_2^+ - y_2^1| \leq \delta x_2$.

For practical reasons, we do not compute $\Delta\dot{f}_i$ by using (27). Instead we estimate the partial derivatives once and for all in a much larger region containing the entire attractor[7]:

$$\Delta\dot{f}_i = \sum_{j=1}^{3} \delta x_j \max_A \left|\frac{\partial \dot{f}_i}{\partial x_j}(x)\right| + \Delta x_3 \max_A \left|\frac{\partial \dot{f}_i}{\partial x_3}(x)\right|,$$

where $A = [-15, 15] \times [-10, 10] \times [-1, 50]$. Seeing that all partial derivatives appearing in $\Delta\dot{f}_i$ are polynomials of degree at most two, it is straight-forward to gain global estimates on them. We list the estimates on all nine partial derivatives:

$$\begin{array}{lll}
\max_A \left|\frac{\partial \dot{f}_1}{\partial x_1}(x)\right| < 2201 & \max_A \left|\frac{\partial \dot{f}_1}{\partial x_2}(x)\right| < 310 & \max_A \left|\frac{\partial \dot{f}_1}{\partial x_3}(x)\right| < 131 \\
\max_A \left|\frac{\partial \dot{f}_2}{\partial x_1}(x)\right| < 2159 & \max_A \left|\frac{\partial \dot{f}_2}{\partial x_2}(x)\right| < 634 & \max_A \left|\frac{\partial \dot{f}_2}{\partial x_3}(x)\right| < 161 \\
\max_A \left|\frac{\partial \dot{f}_3}{\partial x_1}(x)\right| < 1799 & \max_A \left|\frac{\partial \dot{f}_3}{\partial x_2}(x)\right| < 1716 & \max_A \left|\frac{\partial \dot{f}_3}{\partial x_3}(x)\right| < 633.
\end{array}$$

These are defined as constants in the beginning of the code.

4.1.2 Conditions for the generic case

In the previous section, we claimed that, generically, the optimizing trajectories originate from the corners of $\mathcal{R}^{(k)}$. To prove this, we consider the partial derivatives of Π:

$$\frac{\partial \Pi_i}{\partial x_j}(x) = \delta_{ij} - h\frac{(f_i)'_{x_j}(p_i)f_3(p_3) - f_i(p_i)(f_3)'_{x_j}(p_3)}{(f_3(p_3))^2} \qquad (i, j = 1, 2),$$

where δ_{ij} is the Kronecker delta, and $p_i = p_i(x)$ are points on the trajectory of x, as before. Due to the smallness of h, we always have $(\Pi_i)'_{x_i} > 0$. This implies that $\max / \min \Pi_i(\mathcal{R}^{(k)})$ is always attained somewhere along the line $\max / \min\{x_i : x \in \mathcal{R}^{(k)}\}$.

Once again, let us consider an example: if we want to maximize $\Pi_1(x)$ on $\mathcal{R}^{(k)}$, we now know that we only have to examine trajectories starting from the line

$$\ell(t) = (\bar{x}_1 + \Delta x_1, t, x_3^{(k)}), \qquad \text{where} \qquad t \in [x_2^-, x_2^+].$$

As before, we make the assumption that all these trajectories stay inside the box $B_{1,4} \subset B$, defined by

$$B_{1,4} = (x_1^+, \bar{x}_2, \bar{x}_3) \pm (\delta x_1, \Delta x_2, \Delta x_3),$$

see Figure 10.

We can then proceed just as before, with the exception that we now Taylor-expand in the box $B_{1,4}$. Generally we can do even better: if we can show that $(\Pi_1)'_{x_2}$ is non-zero in $B_{1,4}$, then we only have to consider the two smaller boxes B_1 and B_4 defined in the previous section. We must, however, prove that $(\Pi_1)'_{x_2}$ does not vanish in $B_{1,4}$. Thus the condition we have to check is

$$(f_1)'_{x_2}(p_1)f_3(p_3) - f_1(p_1)(f_3)'_{x_2}(p_3) \neq 0 \quad \text{ for all } p_1, p_3 \in B_{1,4}.$$

In order to verify this condition, we Taylor-expand the functions around the center of $B_{1,4}$. First,

[7] The fact that A contains the attractor is verified by the program itself.

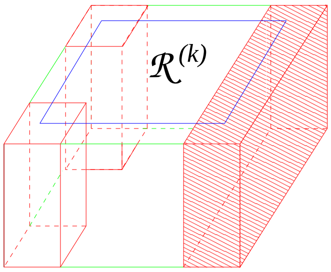

Figure 10: The non-generic case.

we compute the maximal errors:

$$\varepsilon_1 = \delta x_1 \max_{B_{1,4}} \left|\frac{\partial (f_1)'_{x_2}}{\partial x_1}(x)\right| + \sum_{i=2}^{3} \Delta x_i \max_{B_{1,4}} \left|\frac{\partial (f_1)'_{x_2}}{\partial x_i}(x)\right|,$$

$$\varepsilon_2 = \delta x_1 \max_{B_{1,4}} \left|\frac{\partial f_3}{\partial x_1}(x)\right| + \sum_{i=2}^{3} \Delta x_i \max_{B_{1,4}} \left|\frac{\partial f_3}{\partial x_i}(x)\right|$$

$$\varepsilon_3 = \delta x_1 \max_{B_{1,4}} \left|\frac{\partial f_1}{\partial x_1}(x)\right| + \sum_{i=2}^{3} \Delta x_i \max_{B_{1,4}} \left|\frac{\partial f_1}{\partial x_i}(x)\right|,$$

$$\varepsilon_4 = \delta x_1 \max_{B_{1,4}} \left|\frac{\partial (f_3)'_{x_2}}{\partial x_1}(x)\right| + \sum_{i=2}^{3} \Delta x_i \max_{B_{1,4}} \left|\frac{\partial (f_3)'_{x_2}}{\partial x_i}(x)\right|.$$

Next, we note that for all $p_1, p_3 \in B_{1,4}$, we have

$$\begin{aligned}
&(f_1)'_{x_2}(p_1) f_3(p_3) - f_1(p_1)(f_3)'_{x_2}(p_3) \\
&\in \big((\overline{f_1})'_{x_2} \pm \varepsilon_1\big)\big(\bar{f_3} \pm \varepsilon_2\big) - \big(\bar{f_1} \pm \varepsilon_3\big)\big((\overline{f_3})'_{x_2} \pm \epsilon_4\big) \\
&\subset (\overline{f_1})'_{x_2}\bar{f_3} - \bar{f_1}(\overline{f_3})'_{x_2} \pm \big(\varepsilon_1|\bar{f_3}| + \varepsilon_2|(\overline{f_1})'_{x_2}| + \varepsilon_3|(\overline{f_3})'_{x_2}| + \varepsilon_4|\bar{f_1}| + \varepsilon_1\epsilon_2 + \varepsilon_3\varepsilon_4\big),
\end{aligned}$$

where $\bar{f_i}$ and $(\overline{f_i})'_{x_2}$ are f_i and $(f_i)'_{x_2}$ evaluated at the center of $B_{1,4}$, respectively. This means that if

$$\big|(\overline{f_1})'_{x_2}\bar{f_3} - \bar{f_1}(\overline{f_3})'_{x_2}\big| > \varepsilon_1|\bar{f_3}| + \varepsilon_2|(\overline{f_1})'_{x_2}| + \varepsilon_3|(\overline{f_3})'_{x_2}| + \varepsilon_4|\bar{f_1}| + \varepsilon_1\epsilon_2 + \varepsilon_3\varepsilon_4,$$

then $(\Pi_1)'_{x_2}$ does not vanish in $B_{1,4}$, and we can do the computations restricted to B_1 and B_4. If this is not the case, we compute on the whole of $B_{1,4}$. Of course, we can obtain estimates for the remaining coordinates of $\mathcal{R}^{(k+1)}$ analogously.

The conditions for the generic case to occur are checked by calling the C-functions `dNewXdy` etc. which are declared at the beginning of the code.

4.2 Surface transition

Once we have found rigorous bounds on $\mathcal{R}^{(k+1)}$, we can restart the whole procedure with $\mathcal{R}^{(k+1)}$ as the initial rectangle. This can be repeated as long as f_3 has modulus greater than Δf_3. We then rely on that, for trajectories in a neighbourhood of the Lorenz attractor, either $|f_3(x)|$ or $|f_1(x)|$ is reasonably large. The only exception to this rule is when a trajectory comes close to the fixed point situated at the origin. As we will not be performing any numerical computations near the origin, we may disregard this particular situation. Thus, when $|f_3(x)|$ has become small, we switch to x_2x_3-planes by a transition procedure described below. All methods described in the previous section have direct analogues for this situation, and we can continue to follow the surface flowing in the general x_1-direction. By switching between these two different cases, we can follow the surface a whole lap up to its complete return to Σ.

Let us continue our example from the previous section. Suppose that we have followed the initial rectangle $\mathcal{R}^{(0)} \subset \Sigma$ by using the surface propagation algorithm. Let $\mathcal{R}^{(k)}$ denote its image after k steps, and suppose that we, at stage n, have $|f_3(x)| \leq f_1(x)$. This means that the flow is turning to the right. We then retrace our steps and consider the set of rectangles $\mathcal{R}^{(k)}$, $k = 0, \ldots, n$. If we use the notion $\mathcal{R}^{(k)} = [x_1^-(k), x_1^+(k)] \times [x_2^-(k), x_2^+(k)] \times \{x_3^{(k)}\}$, and denote our new initial x_2x_3-rectangle by

$$\hat{\mathcal{R}} = \{\hat{x}_1\} \times [\hat{x}_2^-, \hat{x}_2^+] \times [\hat{x}_3^-, \hat{x}_3^+],$$

we have the following defining relations, see Figure 11:

$$\begin{aligned}
\hat{x}_3^- &= x_3(n), \\
\hat{x}_3^+ &= x_1^+(n_0), \qquad \text{where } n_0 = \max_{k \in [0,n]} \{k \colon x_1^+(k) < x_1^-(n)\}, \\
\hat{x}_2^- &= \min\{x_2^-(k) \colon k \in [n - n_0, n]\}, \\
\hat{x}_2^+ &= \max\{x_2^-(k) \colon k \in [n - n_0, n]\}, \\
\hat{x}_1 &= x_1^+(n_0).
\end{aligned}$$

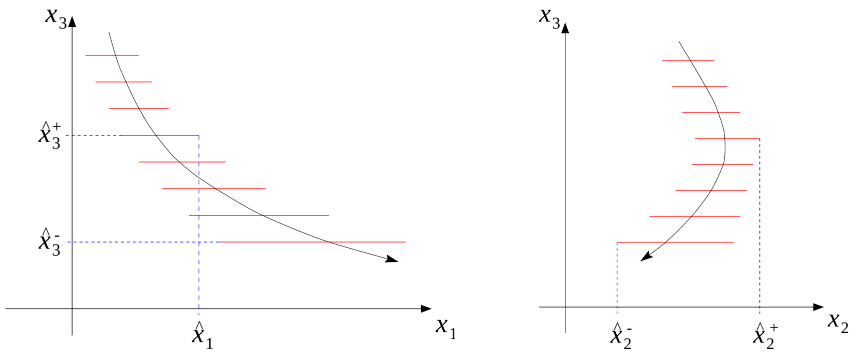

Figure 11: Geometric view of the transition coordinates $\hat{x}_1, \hat{x}_2^-, \hat{x}_2^+, \hat{x}_3^-$, and $\hat{x}_3^+$.

These definitions ensure that all trajectories passing through the rectangles $\mathcal{R}^{(k)}$, will also pass through $\hat{\mathcal{R}}$. This means that we keep our bounds rigorous. We can then continue the surface propagation via x_2x_3-planes using the same techniques as described in section 4.1. The C-functions `Stage0`, ... ,`Stage3` all end by a code segment taking care of the surface transition.

4.3 Partitioning

There is one more problem that we need to consider: as the initial rectangle propagates along the flow, it tends to expand. A lot. This expansion acts in different directions, depending on the local "pull" of the attractor. When the rectangle has reached a critical size, we therefore partition it into smaller rectangles. Failing to do so would not only result in very bad error estimates, which would prevent the algorithms from converging, but we would also end up with expansion in both directions on the return, which clearly rules out the possibility of verifying the existence of a forward invariant set.

When a rectangle has expanded enough, we partition as follows: if one of the side-lengths is less than half of the length of the remaining side, we split the rectangle in two pieces, leaving the shorter side intact. Otherwise, we split the rectangle into four equally sized pieces. The elements of the partition are then treated separately as initial rectangles starting somewhere along the attractor. The splitting is handled by the C-function `Partition`.

The concept of partitioning along the flow is probably the most fundamental idea in the entire program. Although the rectangles produced by our algorithms generally expand in both directions, we can force their returns to be confined to a thin strip by partitioning often. Each partitioned rectangle feels the pull in the attractor's normal direction (that is why we call it an attractor), and is therefore forced to center itself along the attractor. This results in a considerable overlapping, and is precisely why we can see contraction in one direction at the return, see Figure 12.

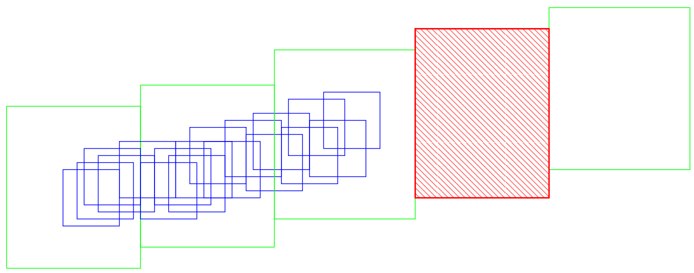

Figure 12: The images (small rectangles) of the initial rectangle (marked) belong to the interior of the forward invariant set (large rectangles).

4.4 Vector propagation

Now that we can rigorously compute the propagation of a rectangle traveling along the flow, we want to compute the evolution of the tangent vectors associated to the rectangle. To be more specific, we have the following problem: given two consecutive rectangles $\mathcal{R}^{(k)}$ and $\mathcal{R}^{(k+1)}$, their bounding box B defined in a previous section, and an initial tangent vector $v^{(k)}$, how do we compute the outgoing vector $v^{(k+1)}$ associated to $\mathcal{R}^{(k+1)}$?

First, we need the differential equations for the Jacobian components. These are attained simply by differentiating the equations for the flow, $\frac{d}{dt}\varphi_i(x,t) = f_i(\varphi(x,t))$ $(i = 1,2,3)$, and changing the order of differentiation. This gives

$$\frac{d}{dt}\frac{\partial \varphi_i}{\partial x_j}(x,t) = \sum_{k=1}^{3} \frac{\partial f_i}{\partial x_k}(\varphi(x,t))\frac{\partial \varphi_k}{\partial x_j}(x,t) \qquad (i,j = 1,2,3),$$

or, in matrix form, $\frac{d}{dt}D\varphi(x,t) = Df(\varphi(x,t))D\varphi(x,t)$.

Next, we proceed much as in the previous sections: we make the assumption that the Jacobian matrix will be confined to a 9-dimensional box $B_{D\varphi}$, we compute absolute error bounds $E_{D\varphi}$ by Taylor-expanding in $B_{D\varphi}$ and B, we calculate the propagation of the Jacobian elements using an Euler method, and finally, we make sure that we really stayed in the box $B_{D\varphi}$.

As just mentioned, the Jacobian and its error define a box in $\mathbb{R}^9$, which we will call $B_{D\varphi}$. Seeing that we always restart our calculations of the Jacobian at each step, a good guess is naturally to take $E_{D\varphi} \approx t^+|Df(\bar{x})|$, where t^+ is the maximal flow time between $\mathcal{R}^{(k)}$ and $\mathcal{R}^{(k+1)}$. Doing so means that we can write $B_{D\varphi} = I \pm E_{D\varphi}$.

We also need bounds on Df in B. On component level, Taylor expanding in B gives

$$(E_{Df})_{ij} = \sum_{k=1}^{3} \Delta x_k \max_B \left|\frac{\partial^2 f_i}{\partial x_j \partial x_k}(x)\right| \qquad (i,j = 1,2,3),$$

which means that, restricted to B, we have $Df(x) \in Df(\bar{x}) \pm E_{Df}$.

Combining our estimates on Df and the Jacobian elements, we now have

$$\tfrac{d}{dt}D\varphi(x,t) \in (Df(\bar{x}) \pm E_{Df})(I \pm E_{D\varphi}),$$

which is valid for all $x \in B$ under our assumption on the box $B_{D\varphi}$. It is convenient to rewrite this differential inclusion as

$$\tfrac{d}{dt}D\varphi(x,t) \in Df(\bar{x}) \pm E_{\Delta},$$

where $E_\Delta = E_{Df}(I + E_{D\varphi}) + |Df(\bar{x})|E_{D\varphi}$. Suppose all flow times from $\mathcal{R}^{(k)}$ to $\mathcal{R}^{(k+1)}$ belong to $[t^-, t^+]$, and let $D\bar{f}$ denote $Df(\bar{x})$. If we define

$$J^+ = \max\left\{t^-(D\bar{f} - E_\Delta), t^-(D\bar{f} + E_\Delta), t^+(D\bar{f} - E_\Delta), t^+(D\bar{f} + E_\Delta)\right\}$$

and

$$J^- = \min\left\{t^-(D\bar{f} - E_\Delta), t^-(D\bar{f} + E_\Delta), t^+(D\bar{f} - E_\Delta), t^+(D\bar{f} + E_\Delta)\right\},$$

it is clear that the Jacobian matrix associated to $\mathcal{R}^{(k+1)}$ belongs to $D\tilde{\varphi} \pm E_{D\tilde{\varphi}}$, where

$$D\tilde{\varphi} = I + \tfrac{1}{2}(J^+ + J^-) \quad \text{and} \quad E_{D\tilde{\varphi}} = \tfrac{1}{2}(J^+ - J^-).$$

By the same argument as in the previous section, these estimates are rigorous if $D\tilde{\varphi} \pm E_{D\tilde{\varphi}} \subset B_{D\varphi}$. If this is the case, we can formally represent the vector associated with $\mathcal{R}^{(k+1)}$ by $v^{(k+1)} = \big(D\tilde{\varphi} \pm E_{D\tilde{\varphi}}\big)v^{(k)}$, i.e.,

$$v^{(k+1)} = D\tilde{\varphi}v^{(k)} \quad \text{and} \quad E_v^{(k+1)} = E_{D\tilde{\varphi}}|v^{(k)}|.$$

These computations are performed at the beginning of the C-functions `Jacobian` and `ConeField`.

4.5 Optimizing the cone field

Now that we know how a single tangent vector evolves along a trajectory, we are prepared to consider the cone field. Initially, we take two tangent vectors $u^{(0)} = (1, u_2^{(0)}, 0)$ and $v^{(0)} = (1, v_2^{(0)}, 0)$ with $u_2^{(0)} > v_2^{(0)}$. These are chosen to define a cone centered around an approximation of the attracting set $\Lambda \subset \Sigma$. The angle between $u^{(0)}$ and $v^{(0)}$ is taken to be $10°$. We then take the two vectors along the flow, just as described in the previous section. At each step, i.e., after flowing from $\mathcal{R}^{(k)}$ to $\mathcal{R}^{(k+1)}$, we select the widest resulting cone as described below.

Since the flow is three-dimensional, the vectors $u^{(k)}, v^{(k)}$ will pick up third components flowing from $\mathcal{R}^{(k)}$ to $\mathcal{R}^{(k+1)}$: $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$. Furthermore, these vectors carry errors E_u and E_v. In order to simplify matters, we eliminate the third components by translating the vectors along the flow down to the tangent plane of $\mathcal{R}^{(k+1)}$.

To see how this is done, consider the local Poincaré map $\Pi^{(k)}\colon \mathcal{R}^{(k)} \to \mathcal{R}^{(k+1)}$ defined by

$$\Pi^{(k)}(x) = \varphi(x, \tau(x)),$$

where $\tau(x)$ is time it takes for the trajectory starting at the point $x \in \mathcal{R}^{(k)}$ to hit the rectangle $\mathcal{R}^{(k+1)}$. Considering the partial derivatives of $\Pi^{(k)}$, we have

$$\begin{aligned}
\frac{\partial \Pi_i^{(k)}}{\partial x_j}(x) &= \frac{\partial}{\partial x_j}\left[\varphi_i(x, \tau(x))\right] = \frac{\partial \varphi_i}{\partial x_j}\big(x, \tau(x)\big) + \frac{\partial \tau}{\partial x_j}(x)\frac{d\varphi_i}{dt}\big(x, \tau(x)\big) \\
&= \frac{\partial \varphi_i}{\partial x_j}\big(x, \tau(x)\big) + \frac{\partial \tau}{\partial x_j}(x) f_i\big(\varphi(x, \tau(x))\big) \\
&= \frac{\partial \varphi_i}{\partial x_j}\big(x, \tau(x)\big) + \frac{\partial \tau}{\partial x_j}(x) f_i\big(\Pi^{(k)}(x)\big) \qquad (i, j = 1, 2, 3). \qquad (28)
\end{aligned}$$

The partial derivatives of $\tau(x)$ are obtained by noting that either $\Pi_1^{(k)}(x)$ or $\Pi_3^{(k)}(x)$ is constant. Continuing our example (i.e. assuming that we still are flowing between x_1x_2-planes), we have

$$0 = \frac{\partial \Pi_3^{(k)}}{\partial x_j}(x) = \frac{\partial \varphi_3}{\partial x_j}\big(x, \tau(x)\big) + \frac{\partial \tau}{\partial x_j}(x) f_3\big(\Pi^{(k)}(x)\big) \qquad (j = 1, 2, 3),$$

and solving for $\frac{\partial \tau}{\partial x_j}$ yields

$$\frac{\partial \tau}{\partial x_j}(x) = -\big[f_3\big(\Pi^{(k)}(x)\big)\big]^{-1}\frac{\partial \varphi_3}{\partial x_j}\big(x, \tau(x)\big) \qquad (j = 1, 2, 3).$$

Inserting this expression into (28) gives

$$\frac{\partial \Pi_i^{(k)}}{\partial x_j}(x) = \frac{\partial \varphi_i}{\partial x_j}\big(x,\tau(x)\big) - \frac{\partial \varphi_3}{\partial x_j}\big(x,\tau(x)\big)\frac{f_i\big(\Pi^{(k)}(x)\big)}{f_3\big(\Pi^{(k)}(x)\big)} \qquad (i,j=1,2,3).$$

Note that the components with $i=3$ vanish, just as we desired. Ignoring the error vectors for the moment, this yields two new vectors $\mathcal{T}(u)$ and $\mathcal{T}(v)$, with components

$$\mathcal{T}_i(u) = u_i - u_3\frac{f_i(x)}{f_3(x)} \qquad \mathcal{T}_i(v) = v_i - v_3\frac{f_i(x)}{f_3(x)} \qquad (i=1,2). \tag{29}$$

Seeing that we do not know the exact location of the base point x, we have to Taylor-expand the components of the vector field appearing in (29). We must also take the errors E_u and E_v into account. Using the same notation as in the previous sections, we can compute E_{f_i/f_3} such that

$$\frac{f_i(x)}{f_3(x)} \in \frac{\bar{f}_i}{\bar{f}_3} \pm E_{f_i/f_3}$$

in the bounding box B. We then have for $i=1,2$

$$\begin{aligned}\mathcal{T}_i(u) \;&\in\; u_i \pm E_{u_i} - \Big(u_3 \pm E_{u_3}\Big)\Big(\frac{\bar{f}_i}{\bar{f}_3} \pm E_{f_i/f_3}\Big)\\ &\subseteq\; u_i - u_3\frac{\bar{f}_i}{\bar{f}_3} \pm \Big(E_{u_i} + \Big|\frac{\bar{f}_i}{\bar{f}_3}\Big|E_{u_3} + E_{f_i/f_3}(|u_3| + E_{u_3})\Big)\\ &\stackrel{\text{def}}{=}\; [\tilde{u}_i^-, \tilde{u}_i^+].\end{aligned}$$

Of course, we just swap u for v to get the interval containing $\mathcal{T}(v)$.

Once we have computed these estimates on $\mathcal{T}(u)$ and $\mathcal{T}(v)$, it is easy to find the widest cone spanned by $\mathcal{T}(u)$ and $\mathcal{T}(v)$. We define $u^{(k+1)}$ and $v^{(k+1)}$ to be the boundary vectors of this cone, scaled so that $u_1^{(k+1)} = v_1^{(k+1)} = 1$. The second components are given by

$$u_2^{(k+1)} = \begin{cases}\tilde{u}_2^+/\tilde{u}_1^-, & \text{if } \tilde{u}_1^+ > 0;\\ \tilde{u}_2^+/\tilde{u}_1^+, & \text{otherwise,}\end{cases} \qquad v_2^{(k+1)} = \begin{cases}\tilde{v}_2^-/\tilde{v}_1^+, & \text{if } \tilde{v}_1^- > 0;\\ \tilde{v}_2^-/\tilde{v}_1^-, & \text{otherwise.}\end{cases}$$

This whole procedure can now be repeated until we complete a return to Σ. The computations are performed in the C-function `ConeField`.

Due to the partitioning, an initial rectangle $\mathcal{R}^{(0)}$ returns to Σ as several small rectangles $\{\mathcal{R}_i\}_{i=1}^n$. By the procedure just described, each one of these rectangles is equipped with a cone spanned by $u^{(n_i)}$ and $v^{(n_i)}$. By our construction, it is clear that the image of the initial cone is contained within the final one.

4.6 Vector transition

When switching from x_1x_2-planes to x_2x_3-planes or vice versa, we must find rigorous estimates on the tangent vectors. Using the notation introduced in section 4.2, we set for $i=1,2,3$,

$$u_i = \max\{u_i^{(k)} : k \in [n-n_0, n]\} \quad v_i = \min\{v_i^{(k)} : k \in [n-n_0, n]\}.$$

The C-functions `Stage0`, ... ,`Stage3` all end by a code segment taking care of the vector transition.

4.7 Expansion estimates

Given an initial rectangle $\mathcal{R}^{(0)}$ with an unstable cone $\mathfrak{C}^u(\mathcal{R}^{(0)})$, we also need to know how much vectors in the cone are expanded under DR. We thus want to find a lower bound of

$$\min\Big\{|DR(x)\cdot w| \colon w \in \mathfrak{C}^u(\mathcal{R}^{(0)}), |w| = 1, x \in \mathcal{R}^{(0)}\Big\},$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^2$. To achieve this, we proceed much as in the previous section. After having passed from $\mathcal{R}^{(k)} \to \mathcal{R}^{(k+1)}$, and translating the vectors down to $\mathcal{R}^{(k+1)}$, we produced two new families of vectors $\mathcal{T}(u)$ and $\mathcal{T}(v)$. Now, instead of selecting the widest pair of vectors, we choose the shortest ones, which we call $\hat{u}^{(k+1)}$ and $\hat{v}^{(k+1)}$. Still assuming that we are flowing between x_1x_2-planes, the components of $\hat{u}^{(k+1)}$ are given by $\hat{u}_1^{(k+1)} = \tilde{u}_1^-$ and

$$\hat{u}_2^{(k+1)} = \begin{cases} 0, & \text{if } \tilde{u}_2^- \tilde{u}_2^+ < 0; \\ \tilde{u}_2^-, & \text{if } |\tilde{u}_2^-| < |\tilde{u}_2^+|; \\ \tilde{u}_2^+, & \text{if } |\tilde{u}_2^-| \geq |\tilde{u}_2^+|. \end{cases}$$

$\hat{v}^{(k+1)}$ is given by exactly the same formulae.
Now, define $w^{(k)}(t) = \frac{u^{(k)}}{|u^{(k)}|} \sin t + \frac{v^{(k)}}{|v^{(k)}|} \cos t$, $t \in [0, \frac{\pi}{2}]$, and let

$$\varepsilon_p^{(k+1)} = \min \left\{ \frac{|\hat{u}_2^{(k+1)}|}{|u_2^{(k)}|}, \frac{|\hat{v}_2^{(k+1)}|}{|v_2^{(k)}|} \right\}$$

be a preliminary estimate on the smallest expansion of vectors $w^{(k)}(t)$ flowing from $\mathcal{R}^{(k)}$ to $\mathcal{R}^{(k+1)}$. Initially, we take $\varepsilon_p^{(0)} = 1$. Also, let $\theta^{(k)}$ denote the angle between the boundary vectors $u^{(k)}$ and $v^{(k)}$. If $\theta^{(k)} \geq \theta^{(k+1)}$, the smallest expansion of $w^{(k)}(t)$ is attained for either $t = 0$ or $t = \frac{\pi}{2}$, so $\varepsilon^{(k+1)}$ is a correct lower bound. If, however, $\theta^{(k)} < \theta^{(k+1)}$, there is an interior vector giving the smallest bound. We then must correct the preliminary expansion estimate by the factor:

$$f = \sqrt{\frac{1 + \cos\theta^{(k+1)}}{1 + \cos\theta^{(k)}}}.$$

To see where this extra factor comes from, we first note that

$$|w^{(k)}(t)|^2 = \left(\tfrac{|u^{(k)}|}{|u^{(k)}|}\right)^2 \sin^2 t + \left(\tfrac{|v^{(k)}|}{|v^{(k)}|}\right)^2 \cos^2 t + \frac{2\langle u^{(k)}, v^{(k)}\rangle}{|u^{(k)}||v^{(k)}|} \sin t \cos t = 1 + 2 \sin t \cos t \cos\theta^{(k)},$$

and

$$\begin{aligned} |\hat{w}^{(k+1)}(t)|^2 &= \left(\tfrac{|\hat{u}^{(k+1)}|}{|u^{(k)}|}\right)^2 \sin^2 t + \left(\tfrac{|\hat{v}^{(k+1)}|}{|v^{(k)}|}\right)^2 \cos^2 t + \frac{2\langle \hat{u}^{(k+1)}, \hat{v}^{(k+1)}\rangle}{|u^{(k)}||v^{(k)}|} \sin t \cos t \\ &\leq \left(\varepsilon_p^{(k+1)}\right)^2 \left(1 + 2 \sin t \cos t \cos\theta^{(k+1)}\right). \end{aligned}$$

Therefore, we clearly have

$$\left(\frac{|\hat{w}^{(k+1)}(t)|}{|w^{(k)}(t)|}\right)^2 \leq \left(\varepsilon_p^{(k+1)}\right)^2 \frac{1 + 2\sin t \cos t \cos\theta^{(k+1)}}{1 + 2 \sin t \cos t \cos\theta^{(k)}} \leq \left(\varepsilon_p^{(k+1)}\right)^2 \frac{1 + \cos\theta^{(k+1)}}{1 + \cos\theta^{(k)}},$$

seeing that $\theta^{(k)} < \theta^{(k+1)}$. If $\theta^{(k)} \geq \theta^{(k+1)}$, we set $f = 1$. Before continuing to the next plane, we multiply the expansion estimate with the previous one, $\varepsilon^{(k+1)} = \varepsilon_p^{(k+1)} f \varepsilon^{(k)}$.
As mentioned earlier, an initial rectangle $\mathcal{R}^{(0)}$ returns to Σ as several small rectangles $\{\mathcal{R}_i\}_{i=1}^n$. By the procedure just described, each one of these rectangles is equipped with an expansion estimate $\varepsilon_i = \Pi_{k=0}^{n(i)} \varepsilon_i^{(k)}$. By selecting $\mathcal{E} = \min_i \varepsilon_i$, it is clear that any vector of the initial cone associated with $\mathcal{R}^{(0)}$ is expanded by at least the factor $\mathcal{E}$. The C-function `ConeField` performs all computations described in this section.

4.8 Hitting the cube

Let us finally deal with the trajectories passing close to the fixed point at the origin. The sole purpose of Chapter 3 was to allow us to interrupt the numeric computations whenever a rectangle

$\mathcal{R}^{(i)}$ comes close to the origin. To be more precise, we center a cube of radius $\frac{1}{5}$ at the origin, and we interrupt the computations if a rectangle hits the lid of the cube. At this stage, the rectangles are small compared to the lid of the cube, and therefore we need not interrupt the computations for partial hits. If, however, the rectangle is completely within the boundary of the lid, we change to the normal form coordinates. This will distort the rectangle and its tangent vectors as described in Lemma 3.4 and the discussion following it. Once we have taken the distortion into account in the worst possible way, we may assume that the flow is totally linear in the cube. Indeed, the maximal error this (erroneous) assumption yields is of the same size as κ, and can thus be considered as taken into account via the distortion[8]. Thus we can explicitly compute where the rectangle exits the cube, and also how the tangent vectors are affected.

There are two different ways that a rectangle can pass through the cube. If the distorted rectangle intersects the stable manifold of the origin (which is the x_2x_3-plane), it is split along the line of intersection, and exits the cube as two cusps. Otherwise the rectangle flows out in one piece, see Figure 13.

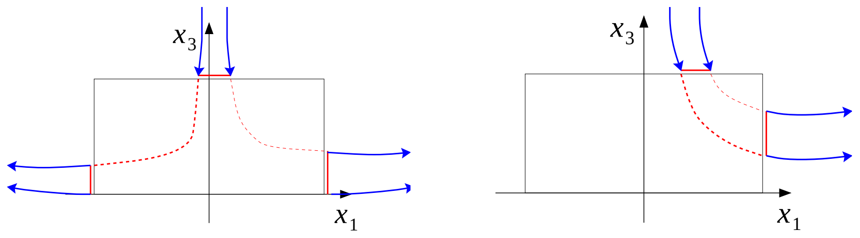

Figure 13: (a) Splitting along the stable manifold (b) No splitting.

Now we are ready to switch back to the original coordinates. Once again, the computed exit rectangle(s) and the associated tangent vectors will be distorted as described in Lemma 3.4. After having assumed the worst possible case, we can resume the numeric computations.

All computations described in this section are performed in `Stage0`.

[8]Recall that κ was used in the estimates on the normal flow. Its value is $2 \cdot 10^{-19}$.

5 Results of the computations

In this section, we summarize the outcome of the program `lorenz_2.1.c`, and see what implications the resulting data has.

5.1 The initial data

First, let us recall our candidate for the trapping region N. This set consists of two disjoint components, N^- and N^+, each made up of 350 small rectangles belonging to the return plane $x_3 = 27$ (note that the plane $x_3 = \varrho - 1$ contains the two fixed points $C^\pm$). We will call these small rectangles $N_i^\pm$, and write

$$N = N^- \cup N^+ = \Big(\bigcup_{i=1}^{350} N_i^-\Big) \cup \Big(\bigcup_{i=1}^{350} N_i^+\Big).$$

The two components of N have the same symmetry as the Lorenz equations, i.e., $N_i^+ = S(N_i^-)$, where $S(x_1, x_2, x_3) = (-x_1, -x_2, x_3)$. Thanks to this symmetry, we only have to perform the computations on one component of N. When it is not relevant which component we are considering, we sometimes omit the $\pm$ labeling of the small rectangles. Each initial rectangle N_i comes with a core angle α_i which indicates the central direction of the associated cone. By selecting the size of the cone openings $\Delta\alpha$, we define the cone field over N, see below. The file `INPUT_FILE_I` contains the coordinates of all rectangles constituting N^+, and their associated core angles.

As the previous file name suggests, there is also a file named `INPUT_FILE_II`. This contains the coordinates and core angles of a partition of N_{121} into 96 smaller rectangles. The rectangle N_{121} is special in the sense that it is the only rectangle of `INPUT_FILE_I` that is split over the stable manifold, and thus subjected to quite extreme hyperbolic forces. We use the partition to get a better resolution of orbits passing near the fixed point.

The program is controlled by a small file named `INIT`. In this file, we may select various options controlling the program and its output, all of which will be explained as we go along.

As mentioned earlier, all source codes and input files are available from the following web site: `http://www.math.uu.se/~warwick/thesis.html`.

5.2 Forward invariance

The program computes C^0 and C^1 information about the return of the rectangle, and saves the output to `LogFile`. It also prints the output to the screen. One of the parameters in `INIT` makes the output more or less detailed.

The C^0 information gives us rigorous bounds on the entire orbit of the initial rectangle. In particular, given any N_i, the program produces a set of overlapping rectangles $\{Q_{i,j}\}_{j=1}^{k(i)}$ whose union strictly contains the return of N_i:

$$R(N_i) \subset \bigcup_{j=1}^{k(i)} Q_{i,j}.$$

By verifying that $\cup_{j=1}^{k(i)} Q_{i,j}$ is strictly contained in N for each $i = 1, \dots, 350$, we conclude that $R(N \setminus \Gamma) \subset \mathrm{int}(N)$, and thus that the return map is well defined (in the sense of the geometric model) on the whole trapping region N.

The program also finds the smallest block in $\mathbb{R}^3$ containing the entire orbit of an initial rectangle. Thus the union of these blocks contains the attracting set of the flow $\mathcal{A}$. An error message is returned if any such block is not contained in

$$A = [-15, 15] \times [-10, 10] \times [-1, 50].$$

As the program does not return an error message for any $\{N_i\}_{i=1}^{350}$, we verify the assumption used in our estimates in Section 4.1.1.

Turning to the cone field, we select a cone opening $\Delta\alpha$ in `INIT`. This is always taken to be $10°$ unless specified otherwise. Thus each initial rectangle N_i comes with a cone $\mathfrak{C}_i = \mathfrak{C}(\alpha_i^-, \alpha_i^+)$, where $\alpha_i^- = \alpha_i - \Delta\alpha/2$, $\alpha_i^+ = \alpha_i + \Delta\alpha/2$, and

$$\mathfrak{C}(\alpha^-, \alpha^+) = \{v \in \mathbb{R}^2 \colon v\angle(0,1) \bmod 180 \in [\alpha^-, \alpha^+]\}.$$

Given an initial rectangle N_i and its associated cone $\mathfrak{C}(\alpha_i^-, \alpha_i^+)$, the C^1 information provides us with $k(i)$ new cones $\mathfrak{C}(\beta_{i,j}^-, \beta_{i,j}^+)$ associated with the returns $Q_{i,j}$, $j = 1, \dots, k(i)$. The program verifies that if $Q_{i,j} \cap N_k \neq \emptyset$, then

$$\alpha_k^- < \beta_{i,j}^- < \beta_{i,j}^+ < \alpha_k^+,$$

which in other words means

$$Q_{i,j} \cap N_k \neq \emptyset \implies \mathfrak{C}(\beta_{i,j}^-, \beta_{i,j}^+) \subset \mathfrak{C}(\alpha_k^-, \alpha_k^+).$$

By verifying this inclusion for each $i = 1, \dots, 350$, we prove the existence of a forward invariant cone field.
All of this was achieved by running the program with the following initial file `INIT`:

```
1               (0 = INPUT_FILE_II [#=96], 1 = INPUT_FILE_I [#=350])
1  350          (InitialN, FinalN)
10.0            (Cone opening in degrees)
1               (0 = sparse output, 1 = more output).
```

5.3 Expansion estimates

We now turn to the question of expansion. Much to our surprise, we found regions in N which were contracted in all directions under R. In view of the geometric model, which is everywhere expanding in the unstable cones, this was not anticipated. We prove, however, that all tangent vectors within the cone field are *eventually* expanded under DR. More precisely, given any orbit $x_0, x_1, \dots$, where $x_j = R^j(x_0)$, we can divide it into non-overlapping pieces $[x_0, \dots, x_{k_0}], [x_{k_0+1}, \dots, x_{k_1}], \dots$ where all but the first piece accumulate an expansion factor greater than 2. The factor 2 is relevant when proving transitivity of the attracting set, see below. We also show that $k_{i+1} - k_i \leq 29$, which gives a very crude lower estimate of the positive Lyapunov exponent of R:

$$\liminf_{n\to\infty} \frac{1}{n} \log(\min\{|DR^n(x)v| \colon v \in \mathfrak{C}_i, |v| = 1, x \in N_i\}) \geq \sqrt[29]{2} > 1.024.$$

5.3.1 Getting oriented

Before we prove these facts, let us pause for a moment, and consider Figure 15. We will use this illustration as a dynamic road map. First, note that Λ consists of two main branches (upper and lower). Next, note that each branch is split in two by the line $\Gamma = W^s(0) \cap \Sigma$. This gives us four naturally defined regions, and we will refer to these as $A_1, \dots, A_4$, where the labeling is analogous to that of quadrants. The dynamics of R is now easy to describe: any trajectory starting in A_1 tends to the left, and enters A_2 after a finite number of iterates. A trajectory starting in A_2, however, must immediately go to either A_3 or A_4. Using the symmetry of the Lorenz equations, we end up with the diagram illustrated in Figure 14.

Of course, we are deliberately ignoring the pre-images of Γ. These points have a finite number of iterates before hitting Γ and vanishing. We have numbered the subrectangles of N as follows: $\cup_{i=1}^{120} N_i^+ \subset A_2$, $\cup_{i=122}^{350} N_i^+ \subset A_1$, and $\Gamma \cap N^+ \subset N_{121}^+$.

Let us now resume our discussion about the contracting regions. There is one such region in each A_i, and they are situated as follows: in A_1, the right-most part ($x_1 > 3.389601$) is contracting. The

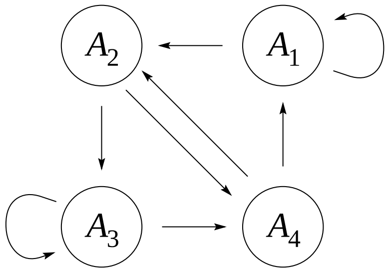

Figure 14: The simplified dynamics of the return map R.

strongest contraction[9] occurs to the far right, and is roughly of a factor 0.8249. All trajectories travel to the left in A_1, and thereby tend to the expanding region of A_1. The contracting region in A_2 is situated to the left ($x_1 < -0.743562$), and the strongest contraction is no less than by a factor 0.2089. This is quite a strong contraction rate, mostly due to the small angle between the attractor and the stable direction, see Figure 15. With our labelling, the expanding region is $\cup_{i=95}^{192} N_i$.

As the geometric model assumes that the attractor is virtually perpendicular to the stable directions, the problem of close tangencies is never discussed in the literature. It is, however, a serious obstruction that must be overcome to prove that there exists an expanding direction. Indeed, by increasing the parameter r slightly, the attractor appears to become tangent to the stable directions, and then there is no hope to save the robust persistence of the attractor, see [6].

We are saved by the fact that orbits entering the out-most parts of A_2 or A_4 have just been very close to the fixed point. Recall that the expansion is very large (unbounded, in fact) in a neighbourhood of the origin. Thus, before we enter a strongly contracting region, we have hopefully already pre-compensated for the coming contraction. We prove that this indeed is the case by using expansion estimates on the pre-images of these regions.

5.3.2 Pre-images and iterates

The computations carried out to prove forward invariance (henceforth called `run#1`) also provide us with a lower bound $\mathcal{E}_i = \mathcal{E}(N_i, \alpha_i^-, \alpha_i^+)$ on the expansion of tangent vectors belonging to the initial cone. Thus any orbit $x_0, x_1, \ldots$, with $x_k \in N_{i_k}$, can be equipped with a corresponding sequence of expansion estimates $\mathcal{E}_{i_0}, \mathcal{E}_{i_1}, \ldots$. We will show that when $\mathcal{E}_{i_k}$ is small, the estimate for the previous iterate $\mathcal{E}_{i_{k-1}}$ is large.

Before doing so, we introduce some notation: let $\langle R\rangle(N_i)$ denote the computed return of N_i, and let $[R](N_i)$ denote the subrectangles of N that intersect $\langle R\rangle(N_i)$, i.e.,

$$\langle R\rangle(N_i) = \bigcup_{j=1}^{k(i)} Q_{i,j} \quad \text{and} \quad [R](N_i) = \{N_k \colon N_k \cap \langle R\rangle(N_i) \neq \emptyset\}.$$

We also define the pre-image of a subrectangle, and its associated expansion estimate:

$$[R^{-1}](N_i) = \{N_k \colon N_i \in [R](N_k)\} \quad \text{and} \quad \mathcal{E}_i^{-1} = \min\{\mathcal{E}_k \colon N_k \in [R^{-1}](N_i)\}.$$

Considering that most subrectangles of N have pre-images in both branches of Λ, and that each such pre-image corresponds to several subrectangles, it is not very practical to iterate backwards more than once using this definition. Also note that, in general, we can not expect to have $[R]([R^{-1}](N_i)) = N_i$.

[9] When we talk about contraction and expansion rates, we always mean the rates restricted to the unstable cone field.

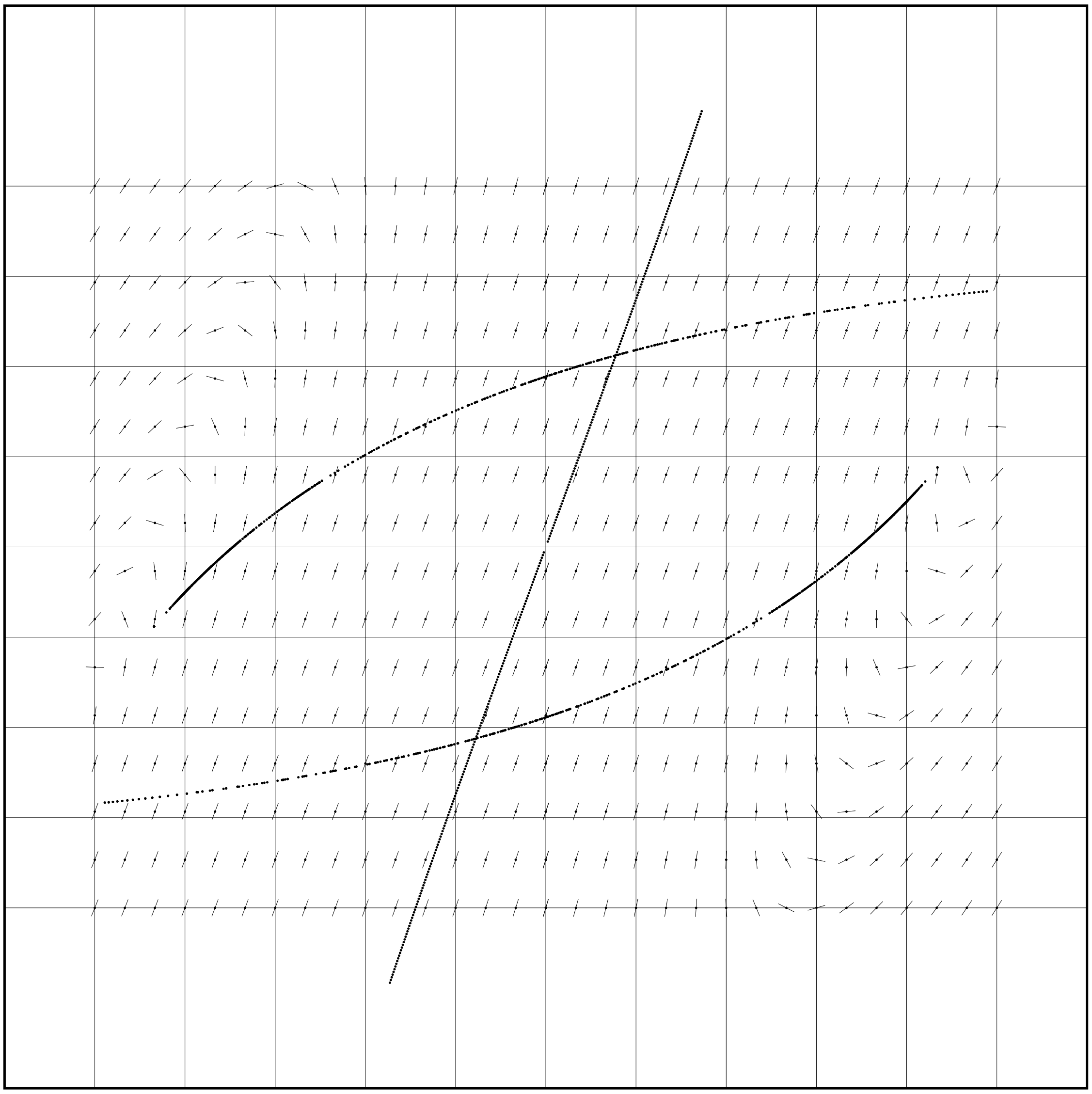

Figure 15: An approximation of Λ with the most contracting directions for one iterate of R. The (almost) straight line cutting across the two branches of Λ is the intersection between the stable manifold of the origin and the return plane. The bounding box is $[-6,6] \times [-6,6] \times \{27\}$.

5.3.3 Pseudo orbits

Set $F^+ = \cup_{i=104}^{129} N_i^+$, $F^- = S(F^+)$, and $F = F^- \cup F^+$. We will consider all orbits starting from the region F, i.e., all orbits with $x_0 \in F$. This selection is based on the fact that all orbits must enter (or originate from) this set sooner or later: an orbit can not cross F without having at least one iterate inside F. Furthermore, F contains the lines $W^s(0) \cap N$, so any line segment being sliced over $W^s(0)$ must also enter F. We will prove the following:

Proposition 5.1 *Consider an orbit $x_0, x_1, \ldots$ with $x_0 \in F$. For every return $x_n \in F$, we have*

$$\min\{|DR^n(x_0)v| \colon v \in \mathfrak{C}\} > 2^n|v|.$$

In particular, this is true for the *first* return to F. Therefore, small line segments must more than double their lengths between two consecutive slicings over $W^s(0)$.

By symmetry, we only have to consider the case $x_0 \in F^+$. The proof can then be split into three different cases:

- $x_1 \in F^+$. This is a technically forced case that can only occur for $x_0 \in N_{128}^+ \cup N_{129}^+$;
- $x_1 \in F^-$. This case deals with all periodic points in F. It can only occur for $x_0 \in \cup_{i=104}^{113} N_i^+$;
- $x_1 \notin F$. This is the most general case, and also requires most work. This situation can occur for all $x_0 \in \cup_{i=112}^{129} N_i^+$.

The case $x_1 \in F^+$:

From `run#1`, it follows that $\mathcal{E}_i > 2$ for $i = 128, 129$. Since all rectangles in F have an associated expansion greater than one, the proposition clearly holds in this case.

The case $x_1 \in F^+$:

We just have to check that $\mathcal{E}_i\mathcal{E}_j > 2$, where $i \in [104, 113]$ and $N_j \in [R](N_i) \cap F$. Since the sequences $\mathcal{E}_{104}, \mathcal{E}_{105}, \ldots, \mathcal{E}_{121}$ and $\mathcal{E}_{129}, \mathcal{E}_{128}, \ldots, \mathcal{E}_{121}$ are increasing, we only have to check the boundary intervals of $[R](N_i) \cap F$. We get the following list of extremal cases:

$$\mathcal{E}_{104}\mathcal{E}_{120} > 10.7 \quad \mathcal{E}_{104}\mathcal{E}_{125} > 3.66 \quad \mathcal{E}_{105}\mathcal{E}_{118} > 5.54 \quad \mathcal{E}_{105}\mathcal{E}_{123} > 5.56$$

$$\mathcal{E}_{106}\mathcal{E}_{116} > 4.20 \quad \mathcal{E}_{107}\mathcal{E}_{114} > 3.38 \quad \mathcal{E}_{108}\mathcal{E}_{112} > 3.00 \quad \mathcal{E}_{109}\mathcal{E}_{110} > 2.77$$

$$\mathcal{E}_{110}\mathcal{E}_{108} > 2.63 \quad \mathcal{E}_{111}\mathcal{E}_{105} > 2.38 \quad \mathcal{E}_{112}\mathcal{E}_{104} > 2.45 \quad \mathcal{E}_{113}\mathcal{E}_{104} > 2.65.$$

As all products are greater than two, the proposition clearly hold in this case too. In particular, this implies that all periodic orbits in F have a Lyapunov exponent greater than $\sqrt{2}$.

The case $x_1 \notin F$:

As already mentioned, this case requires quite a deal of work. This mainly depends on the fact that N_{121} gets sliced over the stable manifold of the origin, and is smeared out over a large region. This results in quite a lot of different cases to consider when we want to follow an orbit. In order to reduce the number of cases, we introduce so called *pseudo orbits* which are guaranteed to produce lower expansion estimates than any true orbit. These orbits are designed to spend as many iterates as possible in the contracting regions, and as few iterates as possible in the expanding regions. By definition, all pseudo orbits terminate in F.

We will construct a main pseudo orbit in A_1. From the output of `run#1`, we see that $[R](N_{350}) = \cup_{i=340}^{344} N_i$. This means that any orbit starting from N_{350} must return within $\cup_{i=340}^{344} N_i$. As this union is contained in the contracting region $N \setminus \cup_{i=95}^{192} N_i$, we assume that N_{344}, the subrectangle situated deepest within the contracting region, contains the return. Continuing one more iterate, we find that $[R](N_{344}) = \cup_{i=334}^{337} N_i$. Again, seeing that the return is contained in the contracting region, we assume that N_{347} contains the return. We continue like this until the orbit enters the expanding region. This gives us the contracting part of our pseudo orbit:

$$N_{350} \to N_{333} \to N_{337} \to N_{330} \to N_{323} \to N_{316} \to N_{309} \to N_{302} \to N_{295} \to N_{288} \to \\ N_{281} \to N_{274} \to N_{267} \to N_{260} \to N_{253} \to N_{246} \to N_{239} \to N_{232} \to N_{225} \to N_{218} \to \\ N_{211} \to N_{198}.$$

As $[R](N_{198}) = \cup_{i=181}^{183} N_i$ is completely contained in the expanding region, we change our algorithm for the next iterates. We now take the iterates as far apart from each other as possible. This will produce an pseudo orbit that leaves the expanding region as soon as possible. We will start from the boundary of F, and work our way backwards: since $[R](N_{142}) = \cup_{i=129}^{132} N_i$, but $[R](N_{143}) = \cup_{i=131}^{133} N_i$, we select N_{142} as the pseudo pre-image of N_{129}. Continuing analogously, we get the expanding part of our pseudo orbit:

$$N_{187} \to N_{169} \to N_{158} \to N_{151} \to N_{142} \to N_{129}.$$

Combining these two parts, we get the full pseudo orbit:

$$N_{350} \to N_{333} \to N_{337} \to N_{330} \to N_{323} \to N_{316} \to N_{309} \to N_{302} \to N_{295} \to N_{288} \to \\ N_{281} \to N_{274} \to N_{267} \to N_{260} \to N_{253} \to N_{246} \to N_{239} \to N_{232} \to N_{225} \to N_{218} \to \\ N_{211} \to N_{198} \to N_{187} \to N_{169} \to N_{158} \to N_{151} \to N_{142} \to N_{129}.$$

Each one of these rectangles has an associated expansion estimate $\mathcal{E}_{i_k}$, and to fill in the blanks we set $\check{\mathcal{E}}_{i_k} = \min\{\mathcal{E}_i \colon i_k \le i < i_{k+1}\}$, and use this estimate for all points in $\cup_{i=i_k}^{i_{k+1}-1} N_i$. It is clear that any true orbit starting from $\cup_{i=i_k}^{i_{k+1}-1} N_i$ must have accumulated an expansion factor of at least $\prod_{j=k}^{27} \check{\mathcal{E}}_{i_j}$ before entering F.

Now that we have this tool in our luggage, let us have a closer look at the remaining case. The rectangles in F^+ whose returns are not completely contained in F are $\cup_{i=112}^{129} N_i^+$. Since we have already covered cases 1 and 2, we only have to consider orbits satisfying $x_1 \in [R](N_i^+) \cap F^c$, $i \in [112, 129]$. For starters, we consider the simple cases (i.e., $i \ne 121$) and stop computing when we enter the expanding region $E = \cup_{i=95}^{192} N_i$. We start with the cases $i \in [112, 120]$:

$$\begin{aligned}
&[R](N_{112}^+) \cap F^c = (\cup_{i=103}^{108} N_i^-) \cap F^c = N_{103}^-; \quad [R](N_{103}^-) \subset F; \quad \mathcal{E}_{112}\mathcal{E}_{103} > 2.35,\\
&[R](N_{113}^+) \cap F^c = (\cup_{i=100}^{106} N_i^-) \cap F^c = \cup_{i=100}^{103} N_i^-; \quad [R](\cup_{i=100}^{103} N_i^-) \subset E; \quad \mathcal{E}_{113}\mathcal{E}_{100} > 2.27,\\
&[R](N_{114}^+) \cap F^c = \cup_{i=97}^{103} N_i^-; \quad [R](\cup_{i=97}^{103} N_i^-) \subset E; \quad \mathcal{E}_{114}\mathcal{E}_{97} > 2.41,\\
&[R](N_{115}^+) \cap F^c = \cup_{i=94}^{100} N_i^-; \quad [R](\cup_{i=94}^{100} N_i^-) \subset E; \quad \mathcal{E}_{115}\mathcal{E}_{94} > 2.52,\\
&[R](N_{116}^+) \cap F^c = \cup_{i=90}^{97} N_i^-; \quad [R](\cup_{i=90}^{97} N_i^-) \subset E; \quad \mathcal{E}_{116}\mathcal{E}_{90} > 2.54,\\
&[R](N_{117}^+) \cap F^c = \cup_{i=85}^{94} N_i^-; \quad [R](\cup_{i=85}^{94} N_i^-) \subset E; \quad \mathcal{E}_{117}\mathcal{E}_{85} > 2.67,\\
&[R](N_{118}^+) \cap F^c = \cup_{i=78}^{90} N_i^-; \quad [R](\cup_{i=78}^{90} N_i^-) \subset E; \quad \mathcal{E}_{118}\mathcal{E}_{78} > 2,90\\
&[R](N_{119}^+) \cap F^c = \cup_{i=60}^{84} N_i^-; \quad [R](\cup_{i=60}^{84} N_i^-) = \cup_{i=150}^{204} N_i^+; \quad \mathcal{E}_{119}\mathcal{E}_{60}\check{\mathcal{E}}_{211}\check{\mathcal{E}}_{198} > 3.36,\\
&[R](N_{120}^+) \cap F^c = \cup_{i=47}^{79} N_i^-; \quad [R](\cup_{i=47}^{79} N_i^-) = \cup_{i=156}^{224} N_i^+; \quad \mathcal{E}_{120}\mathcal{E}_{47}\check{\mathcal{E}}_{225}\check{\mathcal{E}}_{218}\check{\mathcal{E}}_{211}\check{\mathcal{E}}_{198} > 3.90.
\end{aligned}$$

In all computations, we only used three significant digits (truncated) of the expansion estimates. Now we deal with the cases $i \in [122, 129]$:

$$\begin{aligned}
&[R](N_{122}^+) \cap F^c = \cup_{i=55}^{84} N_i^+; \quad [R](\cup_{i=55}^{84} N_i^+) = \cup_{i=150}^{212} N_i^-; \quad \mathcal{E}_{122}\mathcal{E}_{55}\check{\mathcal{E}}_{218}\check{\mathcal{E}}_{211}\check{\mathcal{E}}_{198} > 2.95,\\
&[R](N_{123}^+) \cap F^c = \cup_{i=79}^{89} N_i^+; \quad [R](\cup_{i=79}^{89} N_i^+) \subset E; \quad \mathcal{E}_{123}\mathcal{E}_{79} > 2.86,\\
&[R](N_{124}^+) \cap F^c = \cup_{i=85}^{94} N_i^+; \quad [R](\cup_{i=85}^{94} N_i^+) \subset E; \quad \mathcal{E}_{124}\mathcal{E}_{85} > 2.62,\\
&[R](N_{125}^+) \cap F^c = \cup_{i=90}^{97} N_i^+; \quad [R](\cup_{i=90}^{97} N_i^+) \subset E; \quad \mathcal{E}_{125}\mathcal{E}_{90} > 2.51,\\
&[R](N_{126}^+) \cap F^c = \cup_{i=94}^{100} N_i^+; \quad [R](\cup_{i=94}^{100} N_i^+) \subset E; \quad \mathcal{E}_{126}\mathcal{E}_{94} > 2.60,\\
&[R](N_{127}^+) \cap F^c = \cup_{i=97}^{103} N_i^+ \subset E; \quad [R](\cup_{i=97}^{103} N_i^+) \subset E; \quad \mathcal{E}_{127} > 2.42,\\
&[R](N_{128}^+) \cap F^c = (\cup_{i=100}^{106} N_i^+) \cap F^c = \cup_{i=100}^{103} N_i^+ \subset E; \quad [R](\cup_{i=100}^{103} N_i^+) \subset E; \quad \mathcal{E}_{128} > 2.20,\\
&[R](N_{129}^+) \cap F^c = (\cup_{i=103}^{108} N_i^+) \cap F^c = N_{103}^+ \subset E; \quad [R](N_{103}^+) \subset F; \quad \mathcal{E}_{129} > 2.03.
\end{aligned}$$

Finally, turning to N_{121}^+, we have the expansion estimate $\mathcal{E}_{121} > 9.08$, and the following scheme:

$$[R](N_{121}^+) \cap F^c \subset \cup_{i=1}^{81} N_i; \quad [R](\cup_{i=1}^{81} N_i) = \cup_{i=154}^{350} N_i.$$

By using the same type of estimates as above, we can cope with all rectangles of the return down to N_{24}:

$$[R](N_{24}) = \cup_{i=269}^{279} N_i \sim N_{i_{11}},$$

where $N_{i_{11}}$ is the 11:th element of our main pseudo orbit. This means that we get an expansion factor of at least $\mathcal{E}_{121}\mathcal{E}_{24}\Pi_{j=11}^{27}\check{\mathcal{E}}_{i_j} > 2.02$ before entering F.

For the remaining part of the return $\cup_{i=1}^{23} N_i$, however, this method will not suffice. Instead, we study the outcome of `run#1` more carefully, and conclude that the cone field over $\cup_{i=1}^{23} N_i$ is taken into a cone field with just a 3° opening. Furthermore, this slimmer cone field is preserved by the rectangles $\cup_{i=259}^{350} N_i$. This means that we may use the slimmer cone field on this part of the pseudo orbit, and thereby gain better expansion estimates. By running the program (`run#2`) with the following initial file `INIT`:

```
1               (0 = INPUT_FILE_II [#=96], 1 = INPUT_FILE_I [#=359])
259  350        (InitialN, FinalN)
3.00            (Cone opening in degrees)
0               (0 = sparse output, 1 = more output)
```

we are provided with better expansion estimates $\mathcal{E}_i^{(3)}$ for $i \in [259, 350]$. The improvement is quite good: instead of having $\Pi_{j=1}^{13}\check{\mathcal{E}}_{i_j} > 0.1463$, we may now use $\Pi_{j=1}^{13}\check{\mathcal{E}}_{i_j}^{(3)} > 0.2404$.

The worst situation is of course considering the orbit $N_{121} \to N_1 \to N_{350} \to N_{344} \to \cdots \to N_{147}$, which will have the expansion estimate $\mathcal{E}_{121}\mathcal{E}_1\Pi_{j=1}^{13}\check{\mathcal{E}}_{i_j}^{(3)}\Pi_{j=14}^{27}\check{\mathcal{E}}_{i_j} > 0.443$. This is of course not acceptable for our needs, and we have to do some more work. We will split the initial rectangle N_{121} into 96 subrectangles, and compute (`run#3`) their returns via the following initial file `INIT`:

```
0               (0 = INPUT_FILE_II [#=96], 1 = INPUT_FILE_I [#=359])
1  96           (InitialN, FinalN)
10.0            (Cone opening in degrees)
0               (0 = sparse output, 1 = more output).
```

From the output, we conclude that the pre-expansion of the region $\cup_{i=1}^{23} N_i$ is in fact given by $\mathcal{E}^{-1}(\cup_{i=1}^{23} N_i) > 44.7$. Using this improved estimate, we finally get the expansion estimate

$$\mathcal{E}^{-1}(\cup_{i=1}^{23} N_i)\mathcal{E}_1\Pi_{j=1}^{13}\check{\mathcal{E}}_{i_j}^{(3)}\Pi_{j=14}^{27}\check{\mathcal{E}}_{i_j} > 2.18.$$

This completes the proof of Proposition 5.1, and also provides us with an algorithm for dividing an orbit into non-overlapping pieces $[x_0, \ldots, x_{k_0}], [x_{k_0+1}, \ldots, x_{k_1}], \ldots$ where all but the first piece accumulate an expansion factor greater than 2.

5.4 Topological transitivity and SRB measures

Since the flow of (2) is uniformly volume-contracting and transversal to N, a finite iterate of the return map R is area-contracting on N. This property together with the existence of a forward invariant unstable cone field implies that R admits an invariant stable foliation with $C^{1+\alpha}$ leaves, see §3 in [4]. The singular map f induced by taking quotients along the stable leaves acts on an interval $I = [-a, a]$, and satisfies the following properties:

- The restriction of f to $[-a, 0)$ and $(0, a]$ is of class $C^{1+\alpha}$ with $f'(x) \geq K > 0$ for all $x \neq 0$.
- There exists $C > 0$, $\lambda > 1$ such that $(f^n)'(x) \geq C\lambda^n$, for all $n \geq 0$.
- For any interval $J \subset I$ there exists $n \geq 0$ such that $f^n(J) = I$, i.e., f is *locally eventually onto*.

The last property is not immediate, but can be proved by using Proposition 5.1, see below.

It follows that f admits a unique finite SRB measure μ_f with $\operatorname{supp}(\mu_f) = I$. From this measure it is possible to construct an SRB measure μ_R with $\operatorname{supp}(\mu_R) = \Lambda$ for the return map R, and also an SRB measure μ_X for the flow, see [16].

We will now prove that f is locally eventually onto: The factor 2 that we have been struggling for is motivated by the following simple fact: the set Γ acts as a razor blade, and can cut a passing line segment in half. If neither of the two halves have doubled their lengths before returning to Γ, they could be cut in half again, and thus we could end up with loads of tiny shreds of the initial segment, all of which continue to hit Γ until we are left with nothing but a fine dust.

By Proposition 5.1, however, this can not happen. Any small segment that is cut over Γ will have expanded by more than a factor 2 before returning to Γ. By always selecting the larger half, we get a sequence of longer and longer segments. This will continue until one of them is mapped totally across a *pseudo fundamental domain*, by which we mean set $F_D = \cup_{i=n_0}^{n_1} N_i$ such that no orbit can cross F_D without having at least one iterate in F_D. One such example is the set $F = \cup_{i=104}^{129} N_i$ as used in Proposition 5.1.

When a segment stretches entirely across a pseudo fundamental domain, so will its projection along the stable leaves. This means that the projection covers an interval on the form $[x, f(x)]$. This set is mapped onto the entire interval I within a finite number of iterates. It follows that f is locally eventually onto.

References

[1] J. Guckenheimer, *A strange, strange attractor*, in The Hopf Bifurcation and its Applications by Marsden and McCracken, Springer-Verlag, 1976

[2] J. Guckenheimer, R. F. Williams, *Structural Stability of Lorenz Attractors*, 307-320, Publ. Math. **50**, 1979

[3] S. P. Hasting, W. C. Troy, *A Shooting Approach to the Lorenz Equations*, 298-303, Bull. Amer. Math. Soc. **27**, 1992

[4] M. W. Hirsch, C. C. Pugh, M. Shub, Invariant Manifolds, Lect. Notes in Math. **583**, 1977

[5] E. N. Lorenz, *Deterministic Non-periodic Flow*, 130-141, J. Atmos. Sci. **20**, 1963

[6] S. Luzzatto, M. Viana, *Lorenz attractors without invariant foliations*, in preparation

[7] K. Mischaikow, M. Mrozek, *Chaos in the Lorenz Equations: A Computer-Assisted Proof*, 66-72, Bull. Amer. Math. Soc. **32**, 1995

[8] J. Milnor, *On the Concept of Attractor*, 177-195, Commun. Math. Phys. **99**, 1985

[9] J. Palis, F. Takens, Hyperbolicity & sensitive chaotic dynamics at homoclinic bifurcations, Cambridge University Press, 1993

[10] D. Rand, *The topological classification of Lorenz attractors*, 451-460, Math. Proc. Camb. Phil. Soc. **83**, 1978

[11] C. Robinson, Dynamical Systems, CRC Press, 1995

[12] C. Robinson, *Homoclinic Bifurcation to a Transitive Attractor of Lorenz Type*, 495-518, Nonlinearity **2**, 1989

[13] M. Rychlik, *Lorenz attractors through a Sil'nikov-type bifurcation. Part 1*, 793-821, Ergod. Th. & Dynam. Sys. **10**, 1989

[14] C. L. Siegel, J. K. Moser, Lectures on Celestial Mechanics, Springer-Verlag, 1971

[15] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, Springer-Verlag, 1982

[16] M. Viana, Stochastic Dynamics of Deterministic Systems, Braz. Math. Colloq. **21**, IMPA, 1997

[17] R. F. Williams, *The Structure of Lorenz Attractors*, 321-347, Publ. Math. **50**, 1979