Robust normal forms for analytic vector fields*

Warwick Tucker
Dept. of Mathematics
Uppsala University
Box 480 Uppsala, SWEDEN
warwick@math.uu.se

December 25, 2002

This research was partially supported by NSF Award No. DMS-0107242.

Abstract

The aim of this paper is to provide an alternative to classical linearization techniques, commonly used for describing the qualitative behaviour of trajectories of ordinary differential equations near fixed points. From a quantitative point of view, the classical methods are extremely unstable, and thus lack applicability in real-world situations. The new approach we propose in this paper successfully overcomes the inherent obstructions in the classical methods, and provides a framework which is robust enough to be implemented in computer applications.

1 Introduction

Consider a system of n ordinary differential equations:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ \dot{x}_n &= f_n(x_1, \ldots, x_n), \end{aligned} \tag{1}$$

where the components of the vector field f_i $(i = 1, \ldots, n)$ are analytic in $x_1, \ldots, x_n$. Using vector notation, we express (1) as $\dot{x} = f(x)$.

The particular situation we will be considering is the behaviour of trajectories passing near a saddle fixed point of (1). A *fixed point* of $\dot{x} = f(x)$ is simply a point x^* such that

*2000 Mathematics Subject Classification: Primary 34C20; Secondary 37M99, 65G30.

$f(x^*) = 0$. If we linearize (1) at x^*, we get

$$\begin{aligned} \dot{y}_1 &= \frac{\partial f_1}{\partial x_1}(x_1^*, \ldots, x_n^*)y_1 + \cdots + \frac{\partial f_1}{\partial x_n}(x_1^*, \ldots, x_n^*)y_n \\ &\vdots \\ \dot{y}_n &= \frac{\partial f_n}{\partial x_1}(x_1^*, \ldots, x_n^*)y_1 + \cdots + \frac{\partial f_n}{\partial x_n}(x_1^*, \ldots, x_n^*)y_n \end{aligned} \tag{2}$$

or $\dot{y} = Df(x^*)y$ for short. Here we have translated the fixed point to the origin via $y = x - x^*$, so we are considering small y only. The point is that $Df(x^*)$ is a constant $n \times n$ matrix whose eigenvalues provide us with information regarding the stability of the fixed point. If the matrix $Df(x^*)$ has purely real eigenvalues, not all of the same sign, and all of them non-vanishing, then we say that the fixed point x^* is a *saddle*. If this is the case,

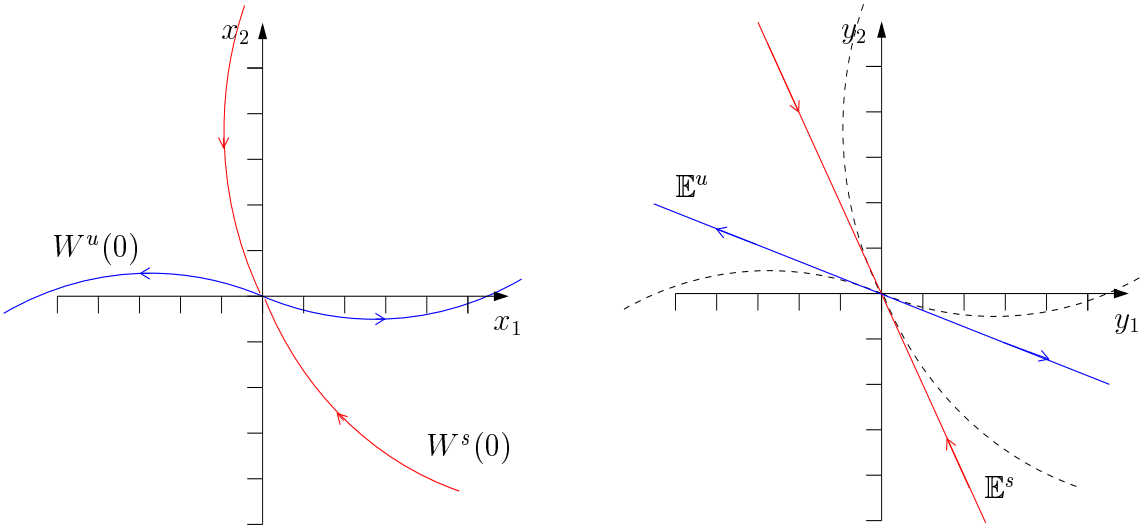

Figure 1: (a) A non-linear planar saddle; (b) its linearization.

elementary linear algebra tells us that there exists a linear change of coordinates $z = Ay$ which brings the system (2) into diagonal form:

$$\begin{aligned} \dot{z}_1 &= \lambda_1 z_1 \\ &\vdots \\ \dot{z}_n &= \lambda_n z_n, \end{aligned} \tag{3}$$

which can be trivially solved. Thus the following two questions can easily be answered for the linear system (3) (and of course (2)):

Q1: How long time does it take a trajectory entering a small neighbourhood of the saddle point to leave the same neighbourhood?

Q2: How much are two nearby trajectories separated when passing through a neighbourhood of the saddle point?

In the limit, question 2 asks for the expansion/contraction rates of a tangent vector passing near the saddle point. For the original (non-linear) system (1), where a linearizing change of coordinates may not exist, the posed questions are not easily answered. This is the topic of the present paper.

2 Linearization procedures

Without any loss of generality, we may assume that the saddle point is located at the origin. This means that the original system (1) can be expressed as

$$\begin{aligned} \dot{x}_1 &= \lambda_1 x_1 + F_1(x_1, \ldots, x_n) \\ &\vdots \\ \dot{x}_n &= \lambda_n x_n + F_n(x_1, \ldots, x_n), \end{aligned} \tag{4}$$

where the functions F_i $(i = 1, \ldots, n)$ contain no linear or constant terms, and are assumed to be analytic in a neighbourhood of the origin. In vector notation, we write $\dot{x} = \Lambda x + F(x)$, with $F(x) = \mathcal{O}(|x|^2)$, and where Λ is the diagonal $n \times n$-matrix defined by

$$\Lambda_{i,j} = \begin{cases} \lambda_i & : \quad i = j, \\ 0 & : \quad i \neq j. \end{cases}$$

Starting from the system (4), we will try to answer the questions posed earlier. Of course, the main difficulty is that we cannot explicitly find a solution to the system. Therefore, the best we can hope for is to obtain bounds on the sought quantities. Nevertheless, we are encouraged by the fact that near a saddle point, the trajectories of (4) should behave roughly like those of its linear counterpart (3). This is motivated by the Hartman-Grobman theorem (see [Ha64]):

Theorem 2.1 (Hartman-Grobman) *Let f be of class C^r, with $r \geq 1$. If x^* is a hyperbolic fixed point of the system $\dot{x} = f(x)$, then there is a homeomorphism h, defined on some neighbourhood of x^*, that takes trajectories of $\dot{x} = f(x)$ to those of the linearized system $\dot{y} = Df(x^*)y$, whilst preserving their orientation w.r.t. time.*

We remind the reader that a fixed point x^* of $\dot{x} = f(x)$ is called *hyperbolic* if $Df(x^*)$ has no purely imaginary eigenvalues. All saddle points are clearly hyperbolic. A *homeomorphism* is a continuous, invertible map with continuous inverse.

As it stands, the theorem is rather weak. A merely continuous map is not likely to preserve much of the structure of the original trajectories, so it is not clear that, from our point of view, the linearized system will accurately describe the non-linear one. Furthermore, the theorem does not easily provide any information regarding the size of the neighbourhood on which h is defined. Remarkably, in dimensions greater than two[1], even if the vector field is assumed to be real analytic, the theorem is *false* if we require h to be of class C^1.

Adding some restraints on the eigenvalues, however, will allow us to give the linearizing change of variables h more regularity.

Definition 2.2 *We say that the eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$ are **resonant** if there exist natural numbers $m_1, \ldots, m_n$ with $|m| = \sum m_i \geq 2$ such that*

$$m\lambda - \lambda_i = \sum_{k=1}^{n} m_k \lambda_k - \lambda_i = 0$$

*for some $i = 1, \ldots, n$. The number $|m|$ is called the **order** of the resonance.*

[1] In the planar case, Hartman showed that if f is of class C^2, then the conjugating function h (and its inverse) can be made to be of class C^1, see [Ha60].

A special case of Sternberg's linearization theorem (see [St57, St58]) tells us that if there are no resonances among the eigenvalues, then the system can be linearized.

Theorem 2.3 (Sternberg) *Let f be of class C^∞. If x^* is a hyperbolic fixed point of the system $\dot{x} = f(x)$, and if the eigenvalues of $Df(x^*)$ are not resonant, then there is a C^∞-diffeomorphism h, defined on some neighbourhood of x^*, that takes trajectories of $\dot{x} = f(x)$ to those of the linearized system $\dot{y} = Df(x^*)y$, whilst preserving their orientation w.r.t. time.*

Here, a *C^r-diffeomorphism* is a C^r invertible map with C^r inverse. For other variants of this theorem, see e.g. [Ne64], [Be78], and [Se85].

Although this result is somewhat better suited for our needs, we must first be able to verify that no resonances occur in our given system. Also, there is still no mention of the size of the neighbourhood on which h is defined. The proof of the above-mentioned theorem is based on the implicit function theorem, which requires a Kantorovich type theorem for explicit bounds on the size of the neighbourhood.

Sternberg's theorem guarantees the existence of a formal power series representation for h. The convergence of this series, however, requires further restraints on the eigenvalues and the vector field.

Definition 2.4 *We say that the eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$ satisfy a **Diophantine** condition of type (κ, τ) if there exists positive κ and τ such that for $i = 1, \ldots, n$ we have $|m\lambda - \lambda_i| \geq \kappa |m|^{-\tau}$ for all $m \in \mathbb{N}^n$ with $|m| \geq 2$.*

The following theorem (see [Si52]) gives sufficient conditions for the convergence of the formal series corresponding to the change of variables :

Theorem 2.5 (Siegel) *Let f be analytic in a neighbourhood of the origin. If $x^* = 0$ is a hyperbolic fixed point of the system $\dot{x} = f(x)$, and if the eigenvalues of $Df(x^*)$ satisfy a Diophantine condition, then there is an analytic change of coordinates h, defined on some neighbourhood of x^*, that takes trajectories of $\dot{x} = f(x)$ to those of the linearized system $\dot{y} = Df(x^*)y$.*

Note that, although the set of eigenvalues satisfying a Diophantine condition with $\tau > 1$ has full Lebesgue measure, the set of resonant eigenvalues is everywhere dense in the set of eigenvalues corresponding to a saddle. This means that not even the *existence* of a *formal* linearizing change of coordinates is guaranteed if we allow for small perturbations of the eigenvalues of $Df(x^*)$. Of course, the Diophantine condition used in Theorem 2.5 is even more fragile. For practical purposes, we wish to allow for small uncertainties in the eigenvalues of $Df(x^*)$. This clearly calls for robust methods that are valid for open sets of eigenvalues.

3 Numerics near fixed points

It may seem strange that one should worry so much about the exact behaviour of trajectories passing near hyperbolic fixed points. After all, as we have just seen, the theory appears to predict that, in most cases, the behaviour is well approximated by the corresponding linear system, restricted to a sufficiently small neighbourhood of the fixed point. As already

pointed out, any such linear system can trivially be solved yielding explicit formulas for the solutions. Thus, from a *mathematical* point of view, the situation is fully understood when the theorems apply.

What could possibly go wrong when *numerically* exploring abovementioned trajectories? First of all, the small neighbourhood on which the linearization holds may very well be much too small with regards to the available machine accuracy. In practice, this would mean that the region where we have a mathematical description of the solutions is too small for the computer to detect. But then, one may ask, is it not irrelevant what actually takes place within this minuscule region? Surely the behaviour in such a small portion of the solution space cannot have any noticeable effect on the larger scales?

Unfortunately, nothing could be further from the truth: it is *exactly* what takes place within these regions that determines the long-time behaviour of the solutions.

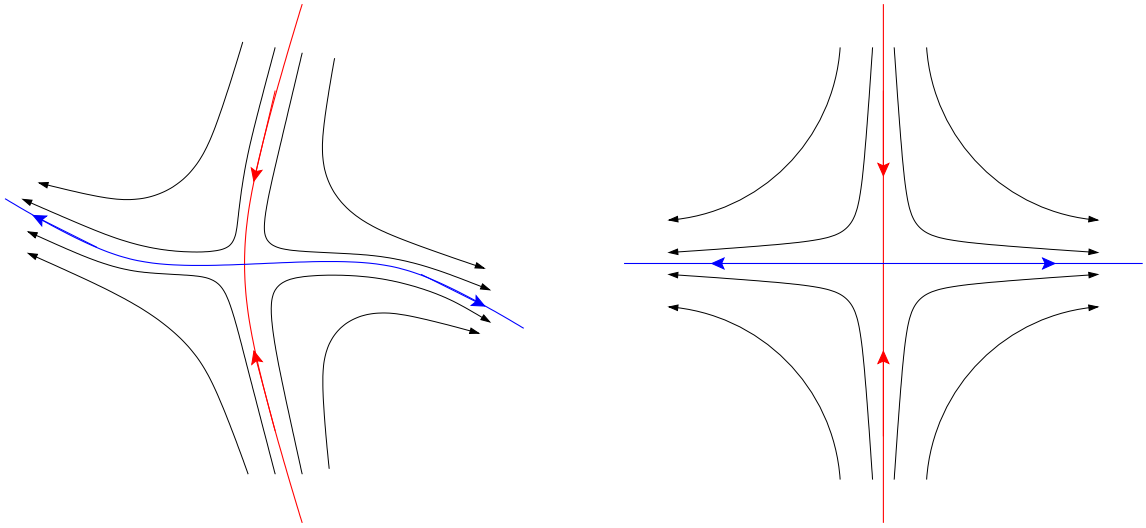

Figure 2: (a) A non-linear diagonal saddle; (b) its linearization.

To illustrate these ideas, consider Figure 2b. It is clear that any trajectory within the ith quadrant, passing near the y_2-axis, will exit the figure along the y_1-axis whilst remaining in the same quadrant. In the nonlinear case (Figure 2a), however, this no longer holds true. The stable and unstable manifolds act as separatrices, and it is the trajectory's relative position to these manifolds that determines its path near the saddle. In the linear case, the invariant manifolds coincide with the coordinate axes, and thus split the (y_1, y_2)-plane into the standard quadrants. In the non-linear case, we do not know the exact locations of the stable and unstable manifolds, and can therefore not even determine whether the trajectory will exit to the left or to the right. Of course, one can obtain estimates on the positions of the invariant manifolds, but this requires some effort.

A somewhat separate concern is that of *flow-times*. A trajectory passing near a fixed point spends a large amount of time in doing so. This is because the vector field is very small near a fixed point, and a trajectory travels with a correspondingly small velocity. When integrating vector fields numerically, it is customary to advance the solutions according to adaptively computed time-steps Δt. The most elementary numerical methods usually take these time-steps to be a pre-defined constant. This is seldom a wise choice, and as most text books on numerical analysis correctly point out, it is wise to decrease the time-

step whenever the modulus of the vector field itself is large, i.e., the product $\Delta t|f(x)|^{1/p}$ should be kept bounded along numerically computed solutions. Here, p denotes the order of the numerical method under consideration. Varying the time-steps in this manner keeps the local discretization errors uniformly bounded, which facilitates estimating the global error. What is usually not mentioned in the literature is that the time-step should also be decreased when the vector field is *small* in modulus. Neglecting to do so is equivalent to completely disregarding the fact that trajectories are *supposed* to spend a long time in the vicinity of a fixed point. If this is not respected, a great deal of information (such as the flow-time spent near the fixed point or the accumulated expansion of tangent vectors) may be lost by prematurely forcing the solutions to leave a neighbourhood of the fixed point.

From a purely numerical point of view, there is no satisfactory way to handle this situation. If the time-steps are decreased as the vector field decreases, the number of steps required to leave a vicinity of the fixed point shoots up and causes a devastating accumulation of local errors, mostly due to rounding errors. On the other hand, if the time-steps are kept bounded from below, information is lost as described above. Therefore it is important to be able to interrupt the numerical integration scheme as soon as a trajectory comes near a fixed point. Once interrupted, analytic methods are used to follow the trajectory until it reaches a safe distance away from the fixed point. At this stage, the computation of the trajectory can once again be accurately handled by the numerics.

4 Normal forms

In what follows, we propose to locally find a close to identity change of coordinates $x = h(y) = y + \phi(y)$ which does *not* bring (4) into a completely linear system, but rather into a system that, in some sense, is close to being linear:

$$\dot{x} = \Lambda x + F(x) \quad \overset{x=y+\phi(y)}{- - - \rightarrow} \quad \dot{y} = \Lambda y + G(y). \tag{5}$$

We call the resulting system $\dot{y} = \Lambda y + G(y)$ a *normal form*. There are of course many choices regarding the structure of G, and we will make a very careful selection. The first property we require from the particular normal form we have in mind is that its unstable and stable manifolds coincide with the appropriate coordinate axes. We then say the the normal form is *rectified*.

In what follows, we will label the eigenvalues of Λ according to

$$\lambda_{s_q} < \cdots < \lambda_{s_1} < 0 < \lambda_{u_1} < \cdots < \lambda_{u_p}.$$

In order for the invariant manifolds to coincide with the coordinate axes, it is necessary that the axes are invariant under the flow. To ensure this, we need a change of variables which, in a fixed neighbourhood of the origin, transforms the original equations $\dot{x} = \Lambda x + F(x)$ into $\dot{y} = \Lambda y + G(y)$, where G satisfies the following conditions:

$$y_{u_1} = \cdots = y_{u_p} = 0 \quad \Rightarrow \quad G_{u_i}(y) = 0 \qquad (i = 1, \ldots, p), \tag{6}$$

and

$$y_{s_1} = \cdots = y_{s_q} = 0 \quad \Rightarrow \quad G_{s_i}(y) = 0 \qquad (i = 1, \ldots, q). \tag{7}$$

In these new coordinates, the unstable manifold coincides with the $(y_{u_1} \ldots y_{u_p})$-plane, and the stable manifold coincides with the $(y_{s_1} \ldots y_{s_q})$-plane, as desired. This will, however, *not*

linearize the flow on the invariant manifolds. As an example, at a point y on the unstable manifold we have $y_{s_1} = \cdots = y_{s_q} = 0$, which brings the normal form into

$$\begin{aligned} \dot{y}_{u_i} &= \lambda_{u_i} y_{u_i} + G_{u_i}(y) \qquad (i = 1, \ldots, p) \\ \dot{y}_{s_i} &= 0 \qquad\qquad\qquad\qquad\quad (i = 1, \ldots, q) \end{aligned}$$

which generally is non-linear in the y_{u_i}-coordinates. An analogous statement can be made for points on the stable manifold. In order to guarantee linear behaviour on the invariant manifolds, we need to impose the additional condition that if a point y is close to the $(y_{u_1} \ldots y_{u_p})$-plane (the unstable manifold) or the $(y_{s_1} \ldots y_{s_q})$-plane (the stable manifold), then the perturbation $G(y)$ is linearly small, i.e.,

$$\min\Big\{ \max_i\{|y_{u_i}|\}, \max_i\{|y_{s_i}|\}\Big\} = \mathcal{O}(\varepsilon) \quad \Rightarrow \quad |G_i(y)| = \mathcal{O}(\varepsilon) \qquad (i = 1, \ldots, n).$$

Depending on the situation at hand, we may want to flatten the normal form even more. Flatness of order ℓ is given by requiring that

$$\min\Big\{ \max_i\{|y_{u_i}|\}, \max_i\{|y_{s_i}|\}\Big\} = \mathcal{O}(\varepsilon) \quad \Rightarrow \quad |G_i(y)| = \mathcal{O}(\varepsilon^\ell) \qquad (i = 1, \ldots, n).$$

In this case, it follows that the components of G can only contain terms of the form $y^m = y_1^{m_1} \ldots y_n^{m_n}$, where the multi-exponent $m \in \mathbb{N}^n$ satisfies both $\sum_{i=1}^p m_{u_i} \geq \ell$ and $\sum_{i=1}^q m_{s_i} \geq \ell$. For future reference, we define the sets

$$\begin{aligned} \mathbb{V}_\ell^n &= \left\{ m \in \mathbb{N}^n : \sum_{i=1}^p m_{u_i} < \ell \quad \vee \quad \sum_{i=1}^q m_{s_i} < \ell \right\}, \\ \mathbb{U}_\ell^n &= \left\{ m \in \mathbb{N}^n : \sum_{i=1}^p m_{u_i} \geq \ell \quad \wedge \quad \sum_{i=1}^q m_{s_i} \geq \ell \right\}. \end{aligned}$$

or, equivalently

$$\begin{aligned} \mathbb{V}_\ell^n &= \left\{ m \in \mathbb{N}^n : \min\Big\{ \sum_{i=1}^p m_{u_i}, \sum_{i=1}^q m_{s_i} \Big\} < \ell \right\}, \\ \mathbb{U}_\ell^n &= \left\{ m \in \mathbb{N}^n : \min\Big\{ \sum_{i=1}^p m_{u_i}, \sum_{i=1}^q m_{s_i} \Big\} \geq \ell \right\}. \end{aligned}$$

In other words, writing G as a formal power series $G(y) = \sum g_m y^m$ (using multi-notation combined with vector notation), we require that

$$g_m \neq 0 \quad \Rightarrow \quad m \in \mathbb{U}_\ell^n .$$

In what follows, we will sometimes omit the superscript n in $\mathbb{U}_\ell^n$ and $\mathbb{V}_\ell^n$.

It is convenient to introduce the notion of *filters* for formal vector-valued power series: given any $f(y) = \sum_{|m|\geq 2} \alpha_m y^m$, we define

$$\langle f(y) \rangle_{\mathbb{U}_\ell} = \sum_{m \in \mathbb{U}_\ell} \alpha_m y^m; \qquad \langle f(y) \rangle_{\mathbb{V}_\ell} = \sum_{m \in \mathbb{V}_\ell} \alpha_m y^m. \tag{8}$$

Note that we always have the decomposition $f(y) = \langle f(y) \rangle_{\mathbb{U}_\ell} + \langle f(y) \rangle_{\mathbb{V}_\ell}$, which splits f into its flat part and its non-flat part. It follows that the non-linear part G of our normal form has flatness of order ℓ if $\langle G(y) \rangle_{\mathbb{U}_\ell} = G(y)$, or equivalently, $\langle G(y) \rangle_{\mathbb{V}_\ell} = 0$.

We stress the fact that flattening a function to order ℓ requires much more effort than simply linearizing it to the same order, i.e., removing all terms $\alpha_m y^m$ with $|m| < \ell$. As an example, in the three-dimensional case with e.g. $\lambda_3 < \lambda_2 < 0 < \lambda_1$, the term $y_1^3 y_2^{992} y_3^5$ is linear to order 1000, but only flat to order 3. In general, flattening a function to order ℓ requires the removal of infinitely many terms, as compared to a finite amount when linearizing to the same order.

5 Main results

Let $\mathcal{S}^n$ denote the space of all real-valued, diagonal $n \times n$-matrices corresponding to the linearization at a saddle (i.e., strictly indefinite matrices), and let $\mathcal{F}_\ell^n$ denote the space of all such matrices whose diagonal elements $\lambda_1, \ldots, \lambda_n$ have no resonances for $m \in \mathbb{V}_\ell^n$:

$$\mathcal{F}_\ell^n = \big\{\Lambda \in \mathcal{S}^n : m \in \mathbb{V}_\ell^n \Rightarrow m\lambda - \lambda_i \neq 0 \quad (i = 1, \ldots, n)\big\}.$$

We will use the following max norms:

$$|y| = \max\{|y_i| : i = 1, \ldots, n\} \qquad \|f\|_r = \max\{|f(y)| : |y| \leq r\}.$$

Theorem 5.1 *Given an integer $\ell \geq 2$ and a system $\dot{x} = \Lambda x + F(x)$, with $\Lambda \in \mathcal{F}_\ell^n$, and where $F(x) = \sum_{|m|\geq 2} a_m x^m$ is analytic, there exist positive constants r_0, r_1, K_0, K_1 and an analytic, close to identity change of variables $x = y + \phi(y)$ with*

$$\|\phi\|_r \leq K_0 r^2 \qquad (r < r_0),$$

such that $\dot{x} = \Lambda x + F(x)$ is transformed into the normal form $\dot{y} = \Lambda y + G(y)$ satisfying $\langle G(y)\rangle_{\mathbb{U}_\ell} = G(y)$ and

$$\|G\|_r \leq K_1 r^{2\ell} \qquad (r < r_1).$$

This theorem tells us that the change of coordinates *and* the resulting normal form exist (as analytic functions) in a fixed neighbourhood of the origin.

Having established the change of coordinates, what can be said about the flow of the resulting normal form? In what follows, we will let $\mathfrak{B}_r$ denote the closed ball (which in our norm looks like a box) centered at the origin, and having radius r. We will refer to the face $\{y \in \mathfrak{B}_r : y_{s_1} = r\}$ as the *lid* of the box $\mathfrak{B}_r$ (recall that λ_{s_1} is the weakest contracting direction of the stable manifold). Within $\mathfrak{B}_r$, we let $\psi(y,t)$ denote the solution to the normal form $\dot{y} = \Lambda y + G(y)$.

We begin with the special case where Λ has only one positive eigenvalue λ_u. In this case the saddle point has a unique unstable direction, and thus any trajectory starting from the lid of $\mathfrak{B}_r$ (except points on the stable manifold of the origin) will exit through an unstable face $\{y \in \mathfrak{B}_r : |y_u| = r\}$. We would like to know how long time a trajectory spends inside the box, and where it exits.

Theorem 5.2 *If Λ has only one positive eigenvalue λ_u, then under the same conditions as in Theorem 5.1, and given any positive constant κ, there is a positive r sufficiently small such that for any trajectory starting from the lid of $\mathfrak{B}_r$, we have the following enclosure of its point of exit:*

$$\psi_u(y, \tau_e(y)) = \operatorname{sign}(y_u) r; \qquad r\big(\tfrac{|y_u|}{r}\big)^{\frac{|\lambda_{s_1}|+\kappa}{\lambda_u-\kappa}} \leq \psi_{s_1}(y, \tau_e(y)) \leq r\big(\tfrac{|y_u|}{r}\big)^{\frac{|\lambda_{s_1}|-\kappa}{\lambda_u+\kappa}},$$

where $\tau_e(y)$ denotes the time spent inside $\mathfrak{B}_r$:

$$\frac{1}{\lambda_u+\kappa}\log\frac{r}{|y_u|} \le \tau_e(y) \le \frac{1}{\lambda_u-\kappa}\log\frac{r}{|y_u|}.$$

If Λ has several negative eigenvalues $\lambda_{s_q} < \cdots < \lambda_{s_1} < 0$, and if we take $\ell > (|\lambda_{s_q}|+1)/(|\lambda_{s_1}|-\kappa)$, then we also have the following enclosures:

$$(y_{s_i}-\kappa r)\big(\tfrac{|y_u|}{r}\big)^{\frac{|\lambda_{s_i}|}{\lambda_u-\sigma_1\kappa}} \le \psi_{s_i}(y,\tau_e(y)) \le (y_{s_i}+\kappa r)\big(\tfrac{|y_u|}{r}\big)^{\frac{|\lambda_{s_i}|}{\lambda_u+\sigma_2\kappa}},$$

where $\sigma_1 = \operatorname{sign}(y_{s_i}-\kappa r)$ and $\sigma_2 = \operatorname{sign}(y_{s_i}+\kappa r)$.

Remark 1 *These additional enclosures can be made somewhat sharper, see Lemma 10.5.*

In the most general setting, we allow for Λ having several positive eigenvalues $0 < \lambda_{u_1} < \cdots < \lambda_{u_p}$. This situation adds the complication of determining through which unstable face of $\mathfrak{B}_r$ a trajectory will exit. It is therefore more appropriate to provide enclosures of the trajectories within the box, and an enclosure of the required exit-time $\tau_e(y)$.

Theorem 5.3 *Under the same conditions as in Theorem 5.1, and given any positive constants κ and r sufficiently small, then for any trajectory starting from the lid of $\mathfrak{B}_r$, we have the following enclosures of the unstable components of its path throughout the box:*

$$|\psi_{u_i}(y,t) - y_{u_i}e^{\lambda_{u_i}t}| \le \frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{u_i}t} \qquad (i=1,\ldots,p),$$

for any α_i satisfying $0 < \lambda_{u_i} < \alpha_i \le \lambda_{u_i} - \ell(\lambda_{s_1}+\kappa)$.

If we take $\ell > |\lambda_{s_q}|/(|\lambda_{s_1}|-\kappa)$, then for any α_i satisfying $0 < \alpha_i \le \lambda_{s_i} - \ell(\lambda_{s_1}+\kappa)$, we also have similar enclosures of the stable components:

$$|\psi_{s_i}(y,t) - y_{s_i}e^{\lambda_{s_i}t}| \le \frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{s_i}t} \qquad (i=1,\ldots,q).$$

As in Theorem 5.2, there exist explicit bounds on the time spent inside $\mathfrak{B}_r$:

$$\tau_e^-(y) \le \tau_e(y) \le \tau_e^+(y),$$

where $\tau_e^{\pm}(y) \nearrow +\infty$ as $\max_{i=1,\ldots,p}\{|y_{u_i}|\} \searrow 0$.

Remark 2 *See Corollary 10.9 for the explicit flow-time bounds $\tau_e^-(y)$ and $\tau_e^+(y)$.*

These theorems have several strengths. First, the constants r_0, r_1, K_0, K_1, α, κ can be explicitly found, and are easy to obtain in terms of ℓ, Λ, and F (naturally κ also depends on r). Second, the change of variables $x = y + \phi(y)$ is analytic for $|y| < r_0$, which means that explicit bounds on its inverse and derivatives can be obtained by Cauchy estimates. The same holds for G when $|y| < r_1$. Furthermore, Theorems 5.2 and 5.3 tell us that solutions to the normal form act very much like those of the completely linearized system. This is not true for a system linearized up a certain high, but finite, order. Finally, the set $\mathcal{F}_\ell^n$, viewed as a subset of $\mathcal{S}^n$, is open and has full Lebesgue measure. We call such a set *robust*: almost all members of $\mathcal{S}^n$ belong to $\mathcal{F}_\ell^n$, and any sufficiently small perturbation of an element in $\mathcal{F}_\ell^n$ remains in $\mathcal{F}_\ell^n$. This allows us to perform the change of coordinates even when we only know the eigenvalues up to some finite degree of accuracy, see e.g. [Tu02]. In contrast to this, we point out that the theorems by Sternberg and Siegel fail on an everywhere dense subset of $\mathcal{S}^n$, and can thus not be used in a noisy environment.

6 The change of variables

Returning to the normal form, we need to know how the vector field $\dot{x} = \Lambda x + F(x)$ is affected by the close to identity change of variables $x = y + \phi(y)$. We have the following identity:

$$\dot{x} = \Lambda(y + \phi(y)) + F(y + \phi(y)) = \Lambda y + \Lambda\phi(y) + F(y + \phi(y)). \tag{9}$$

On the other hand, we also have

$$\begin{aligned} \dot{x} &= \frac{d}{dt}(y + \phi(y)) = (I + D\phi(y))\dot{y} = (I + D\phi(y))(\Lambda y + G(y)) \\ &= \Lambda y + D\phi(y)\Lambda y + G(y) + D\phi(y)G(y). \end{aligned} \tag{10}$$

Comparing the two right-hand sides of (9) and (10) gives

$$D\phi(y)\Lambda y - \Lambda\phi(y) = F(y + \phi(y)) - D\phi(y)G(y) - G(y). \tag{11}$$

For shorthand, we will use the following notation

$$L_\Lambda \phi(y) = D\phi(y)\Lambda y - \Lambda\phi(y).$$

The operator L_Λ 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 (11) on component level, we have

$$L_{\Lambda,i}\phi_i(y) = F_i(y + \phi(y)) - \sum_{j=1}^{n} \frac{\partial \phi_i}{\partial y_j}(y) G_j(y) - G_i(y) \qquad (i = 1, \ldots, n), \tag{12}$$

where

$$L_{\Lambda,i}\phi_i(y) = \sum_{j=1}^{n} \frac{\partial \phi_i}{\partial y_j}(y) \lambda_j y_j - \lambda_i \phi_i(y) \qquad (i = 1, \ldots, n).$$

Note that

$$L_{\Lambda,i}(a_{i,m} y^m) = (m_1\lambda_1 + \cdots + m_n\lambda_n - \lambda_i) a_{i,m_1,\ldots,m_n} y_1^{m_1} \ldots y_n^{m_n} = (m\lambda - \lambda_i) a_{i,m} y^m.$$

The crux is now to choose ϕ so that we produce only flat component functions in the normal form: $\langle G_i(y) \rangle_{\mathbb{U}_\ell} = G_i(y)$. This means that $G_i(y)$ must *not* contain elements on the form $y^m = y_1^{m_1} \ldots y_n^{m_n}$ where the exponent m belongs to $\mathbb{V}_\ell$. By (12), non-flat elements can only come from $F_i(y + \phi(y))$, and any such term can be absorbed by an appropriate choice of ϕ_i provided that the corresponding *divisor* $m\lambda - \lambda_i$ does not vanish. Thus the component functions ϕ_i need only consist of the non-flat terms appearing in the right-hand side of (12), which implies that we should choose ϕ_i such that $\langle \phi_i \rangle_{\mathbb{V}_\ell} = \phi_i$.

By filtering (12), we thus get

$$L_{\Lambda,i}\phi_i(y) = \big\langle F_i(y + \phi(y)) \big\rangle_{\mathbb{V}_\ell} \qquad (i = 1, \ldots, n), \tag{13}$$

and

$$G_i(y) = \big\langle F_i(y + \phi(y)) \big\rangle_{\mathbb{U}_\ell} - \sum_{j=1}^{n} \frac{\partial \phi_i}{\partial y_j}(y) G_j(y) \qquad (i = 1, \ldots, n). \tag{14}$$

We will begin by considering the existence and convergence of ϕ. The recursive scheme (13) can be formally solved by a power series

$$\phi_i(y) = \sum_{|m|=2}^{\infty} a_{i,m} y^m \qquad (i = 1, \ldots, n),$$

where the coefficients are determined by inserting this expression into (13). The existence of a solution ϕ is given by comparing both sides of (13): if $a_{i,m}y^m$ is a term of $\phi_i(y)$ with $|m| = m_1 + \cdots + m_n$, the comparison gives

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

where γ is a polynomial in the coefficients of the terms in ϕ_i $(i = 1, \ldots, n)$ of degree less than $|m|$. Thus the *existence* of ϕ is proved if we show that the divisors $m\lambda - \lambda_i$ do not vanish. As ϕ does not contain constant or linear terms, and since $\langle\phi\rangle_{\mathbb{V}_\ell} = \phi$, the only divisors we need to consider are on the form $m\lambda - \lambda_i$, where $m \in \mathbb{V}_\ell$ and $|m| \geq 2$ (see Figure 3b for a two-dimensional example). In fact, the situation is generally more favourable than this: given an explicit system $\dot{x} = \Lambda x + F(x)$, we only have to consider elements of $\mathbb{V}_\ell$ that actually occur in the absorption process of the change of variables. These depend on the exact form of F, and may be very few compared to the total number of elements of $\mathbb{V}_\ell$.

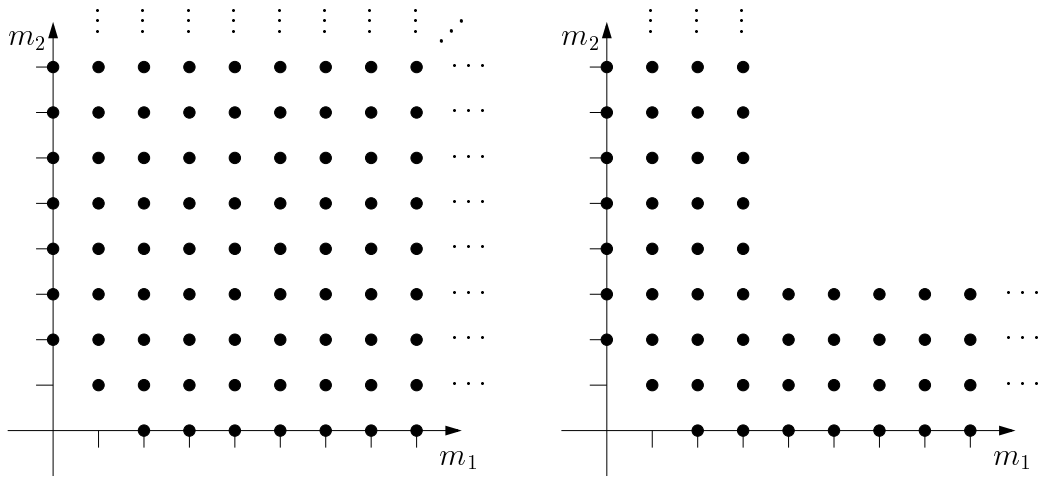

Figure 3: The sets (a) $\{m \in \mathbb{N}^2 : |m| \geq 2\}$; (b) $\{m \in \mathbb{V}_4^2 : |m| \geq 2\}$.

7 Small divisors and existence

In what follows, we let $\lceil x \rceil$ denote the *ceiling* of a real number x, i.e., $\lceil x \rceil = \min\{k \in \mathbb{Z} : x \leq k\}$. We also introduce the numbers $\check{\lambda}$, $\hat{\lambda}$, and $\tilde{\lambda}$ which denote the eigenvalue having the smallest modulus, the eigenvalue having the opposite sign of $\check{\lambda}$ with largest modulus, and the eigenvalue of the same sign as $\check{\lambda}$ with largest modulus, respectively:

$$\check{\lambda} = \begin{cases} \lambda_{s_1} & : \ |\lambda_{s_1}| < |\lambda_{u_1}| \\ \lambda_{u_1} & : \ \text{o.w.} \end{cases} \quad \hat{\lambda} = \begin{cases} \lambda_{u_p} & : \ \check{\lambda} < 0 \\ \lambda_{s_q} & : \ \text{o.w.} \end{cases} \quad \tilde{\lambda} = \begin{cases} \lambda_{s_q} & : \ \check{\lambda} < 0 \\ \lambda_{u_p} & : \ \text{o.w.} \end{cases}$$

Let us begin by stating a lemma that, together with its corollary, proves the existence of a formal series for ϕ for virtually every saddle fixed point.

Lemma 7.1 *If the eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$ are non-resonant for $m \in \mathbb{V}_\ell$, then the divisors $m\lambda - \lambda_i$ are bounded away from zero. Furthermore, for all orders $|m| \geq \Gamma_{\Lambda,\ell} \equiv \ell - 1 + \lceil (\ell-1)|\hat{\lambda}/\check{\lambda}| + \tilde{\lambda}/\check{\lambda} \rceil$, we have the following sharp lower bound:*

$$|m\lambda - \lambda_i| \geq |\big(|m| - (\ell-1)\big)\check{\lambda} + (\ell-1)\hat{\lambda} - \tilde{\lambda}| \qquad (i = 1, \ldots, n).$$

Proof: Take $|m|$ large. Since we are only considering $m \in \mathbb{V}_\ell$, this means that either $\sum_{i=1}^q m_{s_i}$ or $\sum_{i=1}^p m_{u_i}$ is large, but not both (since one of them must be less than ℓ). Although the corresponding eigenvalues have opposite signs, the modulus of the divisor $|m\lambda - \lambda_i|$ must then also be large. There are two cases to consider:

Case 1: $\sum_{i=1}^p m_{u_i} < \ell$. This means that $\sum_{i=1}^q m_{s_i}$ is large, i.e., the divisor $m\lambda - \lambda_i$ is large and negative. We clearly minimize the modulus of the divisor when $m_{u_p} = \ell - 1$, $m_{s_1} = |m| - (\ell - 1)$, and $\lambda_i = \lambda_{s_q}$, which gives

$$|m\lambda - \lambda_i| \geq |(\ell-1)\lambda_{u_p} + \big(|m| - (\ell-1)\big)\lambda_{s_1} - \lambda_{s_q}| \qquad (i = 1, \ldots, n).$$

Case 2: $\sum_{i=1}^q m_{s_i} < \ell$. This means that $\sum_{i=1}^p m_{u_i}$ is large, i.e., the divisor $m\lambda - \lambda_i$ is large and positive. We clearly minimize the modulus of the divisor when $m_{s_q} = \ell - 1$, $m_{u_1} = |m| - (\ell - 1)$, and $\lambda_i = \lambda_{u_p}$, which gives

$$|m\lambda - \lambda_i| \geq |(\ell-1)\lambda_{s_q} + \big(|m| - (\ell-1)\big)\lambda_{u_1} - \lambda_{u_p}| \qquad (i = 1, \ldots, n).$$

Combining both cases, we see that the lowest bound is given by

$$|m\lambda - \lambda_i| \geq |\big(|m| - (\ell-1)\big)\check{\lambda} + (\ell-1)\hat{\lambda} - \tilde{\lambda}| \qquad (i = 1, \ldots, n),$$

which provides the sharp lower bound.

How large must $|m|$ be for this bound to hold? Clearly, the bound is valid starting from the last sign change of $\big(|m| - (\ell-1)\big)\check{\lambda} + (\ell-1)\hat{\lambda} - \tilde{\lambda}$, which happens near the largest (in $|m|$) approximate zero:

$$\big(|m| - (\ell-1)\big)\check{\lambda} + (\ell-1)\hat{\lambda} - \tilde{\lambda} \approx 0.$$

Solving for $|m|$ gives

$$|m| \approx \frac{1}{\check{\lambda}}\big((\ell-1)\check{\lambda} - (\ell-1)\hat{\lambda} + \tilde{\lambda}\big) = \ell - 1 + (\ell-1)|\hat{\lambda}/\check{\lambda}| + \tilde{\lambda}/\check{\lambda}.$$

Rounding up to the nearest integer produces the desired bound:

$$|m| = \ell - 1 + \lceil (\ell-1)|\hat{\lambda}/\check{\lambda}| + \tilde{\lambda}/\check{\lambda} \rceil.$$

Beyond this order, the divisors will increase in modulus with $|m|$, and have the same sign as the eigenvalue of smallest modulus $\check{\lambda}$. □

Remark 3 *In the planar case ($n = 2$), we always have $\check{\lambda} = \tilde{\lambda}$, which gives the bound*

$$|m\lambda - \lambda_i| \geq |(|m| - \ell)\check{\lambda} + (\ell-1)\hat{\lambda}| \qquad (i = 1, 2),$$

which is valid for all $|m| \geq \Gamma_{\Lambda,\ell} \equiv \ell + \lceil (\ell-1)|\hat{\lambda}/\check{\lambda}| \rceil$.

Remark 4 *Note that the asymptotic growth of the divisors is given by*

$$|m\lambda - \lambda_i| \sim |m||\check{\lambda}|.$$

It might appear that requiring the eigenvalues to be non-resonant in $\mathbb{V}_\ell$ is a serious restriction. The following corollary, however, shows that this is in fact almost a completely void demand.

Corollary 7.2 *For any integer $\ell \geq 2$, the set of eigenvalues*

$$(\lambda_{s_1}, \ldots, \lambda_{s_q}, \lambda_{u_1}, \ldots, \lambda_{u_p}) \in \mathbb{R}^q_- \times \mathbb{R}^p_+$$

that are resonant for $m \in \mathbb{V}^n_\ell$ form a closed set of n-dimensional Lebesgue measure zero.

The keyword here is *closed*. This means that the non-resonant eigenvalues form an open set. Furthermore, this set has full measure. Recalling our wish to be allowed some uncertainty in the eigenvalues, this situation is ideal for our needs. The special ordering of the eigenvalues in the statement can be achieved by a simple permutation of the coordinates, and thus causes no loss of generality.

Proof: By Lemma 7.1, there are only a finite number of orders $|m|$ we need to consider. Since each order can give rise to at most a finite number of different resonances, it clearly suffices to show that each such resonance forms a closed set of measure zero in $\mathbb{R}^q_- \times \mathbb{R}^p_+$. But this is obvious: any resonance $m\lambda - \lambda_i = 0$ corresponds to a codimension-1 plane in $\mathbb{R}^{p+q}$ passing through the origin (see Figure 4 for the special case $n = 2$). A finite union of $(n-1)$-dimensional planes certainly forms a closed set of n-dimensional measure zero, as claimed. □

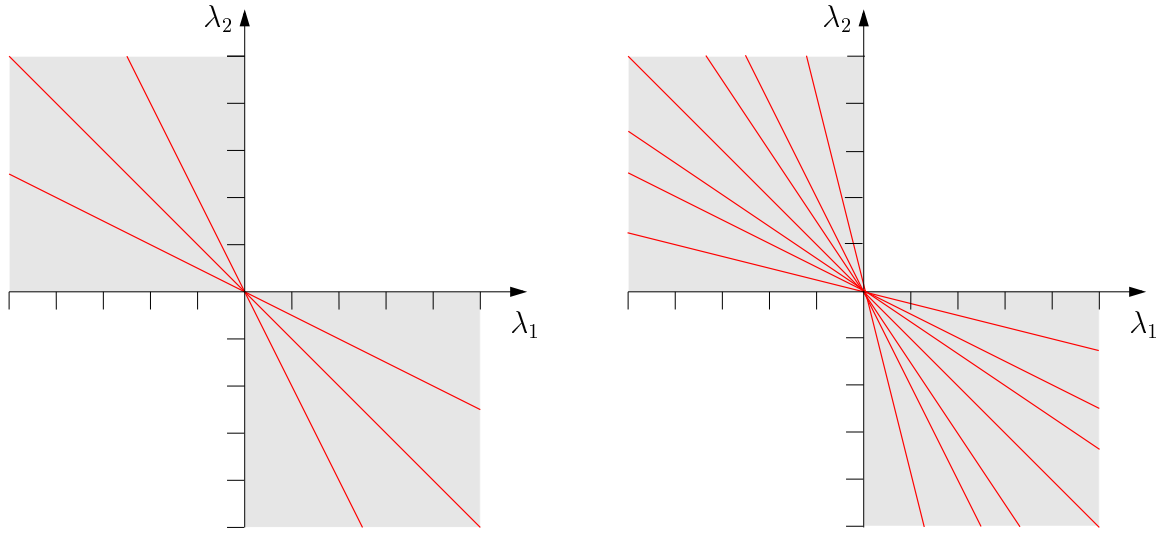

Figure 4: The resonant set in the planar case ($n = 2$) with $\lambda_2 < 0 < \lambda_1$ for (a) $\ell = 5$; (b) $\ell = 6$.

Remark 5 *As the order of flatness ℓ increases, so does the number of resonant planes. In the limit $\ell \to \infty$, the resonant set becomes everywhere dense in $\mathbb{R}^q_- \times \mathbb{R}^p_+$. This limiting case corresponds to completely linearizing the system, i.e., choosing $G \equiv 0$.*

8 Majorants and convergence

Assuming, in what follows, that the formal power series for ϕ defined by (13) exists, we want to show that it also actually converges. To be able to talk about convergence, we need to specify a norm. It is convenient to work in a complex neighbourhood of the origin, and we will use the appropriate max norms:

$$|y| = \max\{|y_i|\colon i = 1, \ldots, n\} \qquad \|f\|_r = \max\{|f(y)|\colon |y| \leq r\}.$$

In order to prove convergence, we follow [SM71] and [Hi76], and use the methods of majorants. If

$$f(\zeta) = \sum_m \alpha_{m_1,\ldots,m_n} \zeta_1^{m_1} \cdots \zeta_n^{m_n} \qquad g(\zeta) = \sum_m \beta_{m_1,\ldots,m_n} \zeta_1^{m_1} \cdots \zeta_n^{m_n}$$

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

$$|\alpha_{m_1,\ldots,m_n}| \leq \beta_{m_1,\ldots,m_n}$$

holds 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.

Suppose that we can find a function $\tilde{F}\colon \mathbb{C}^n \to \mathbb{C}$ such that $F_i \prec \tilde{F}$ $(i = 1, \ldots, n)$ and, together with (13), consider the majorant system

$$\tilde{L}_\Lambda \tilde{\phi}_i(\zeta) = \left\langle \tilde{F}(\zeta + \tilde{\phi}(\zeta)) \right\rangle_{\mathbb{V}_\ell} \qquad (i = 1, \ldots, n), \tag{15}$$

where $\tilde{L}_\Lambda(\zeta^m) = \tilde{\Omega}(m)\zeta^m$ and $\tilde{\Omega}\colon \mathbb{N}^n \to \mathbb{R}$ is defined by

$$\tilde{\Omega}(m) = \min\{|m\lambda - \lambda_i|\colon i = 1, \ldots, n\}.$$

This can be solved formally by a power series

$$\tilde{\phi}_i(\zeta) = \sum_{|m|=2}^{\infty} \tilde{a}_{i,m} \zeta^m \qquad (i = 1, \ldots, n), \tag{16}$$

and it is follows that $\tilde{\phi}_i$ is a majorant of ϕ_i. To see this, compare the two functional equations (13) and (15). In the latter, the divisors appearing on the left-hand side are positive and smaller than or equal to the modulus of those present in (13). Also, the coefficients of $\tilde{F}$, appearing on the right-hand side of (15), are positive and larger than or equal to the modulus of those of F. This implies that the coefficients satisfy $|a_{m_1,\ldots,m_n}| \leq \tilde{a}_{m_1,\ldots,m_n}$ for all m, as claimed.

Furthermore, since both $\tilde{L}_\Lambda$ and the right-hand side of (15) are independent of i, we have $\tilde{\phi}_1 = \cdots = \tilde{\phi}_n$. If we set $\zeta_1 = \cdots = \zeta_n = z$, and find a new function $\check{F}\colon \mathbb{C} \to \mathbb{C}$ such that $\tilde{F}(z, \ldots, z) \prec \check{F}(z)$, we may, together with (15), consider the majorant system

$$\check{L}_\Lambda \check{\phi}(z) = \check{F}(z + \check{\phi}(z)), \tag{17}$$

where $\check{L}_\Lambda(z^k) = \check{\Omega}(k) z^k$ and $\check{\Omega}\colon \mathbb{N} \to \mathbb{R}$ is defined by

$$\check{\Omega}(k) = \min\left\{\tilde{\Omega}(m)\colon |m| = k \quad \wedge \quad m \in \mathbb{V}_\ell\right\}.$$

Again, this can be solved formally by a power series

$$\check{\phi}(z) = \sum_{k=2}^{\infty} \check{a}_k z^k, \tag{18}$$

and, from the same reasoning as above, it is clear that $\tilde{\varphi}(z,\ldots,z) \prec \check{\phi}(z)$. Note that this implies that $\|\phi\|_r \leq \check{\phi}(r)$ in the region of convergence. Thus it suffices to prove the convergence of $\check{\phi}$. We will now present explicit candidates for the abovementioned majorants $\tilde{F}$ and $\check{F}$.

Since we are assuming that F is analytic in a neighbourhood of the origin, we can there identify it with its power series

$$F(\zeta) = \begin{pmatrix} \sum_{|m|=2}^{\infty} c_{1,m}\zeta^m \\ \vdots \\ \sum_{|m|=2}^{\infty} c_{n,m}\zeta^m \end{pmatrix}.$$

Thus if we set

$$\tilde{F}(\zeta) = \sum_{|m|=2}^{\infty} \tilde{c}_m \zeta^m,$$

where $\tilde{c}_m = \max\{|c_{i,m}|\colon i = 1,\ldots,n\}$, we clearly have $F_i \prec \tilde{F}$ $(i = 1,\ldots,n)$, and $\tilde{F}$ has the same radius of convergence as F. Summing the coefficients of all terms having the same degree produces $\check{c}_k = \sum_{|m|=k} \tilde{c}_m$, and if we define

$$\check{F}(z) = \sum_{k=2}^{\infty} \check{c}_k z^k,$$

it follows that $\tilde{F}(z,\ldots,z) \prec \check{F}(z)$. Once again, $\check{F}$ is analytic and has the same radius of convergence as F. Hence, if the solution to (17) converges, then we have $\|\phi\|_r \leq \check{\phi}(r)$ in the region of convergence.

Now, by Lemma 7.1, there exists a positive constant A (depending only on Λ and ℓ) that satisfies $Ak \leq \check{\Omega}(k)$ for all $k = 2, 3, \ldots$. Thus we can replace the operator $\check{L}_\Lambda(z^k) = \check{\Omega}(k)z^k$, by the smaller operator $\hat{L}_\Lambda(z^k) = Akz^k$, which transforms (17) into the following functional equation:

$$\hat{L}_\Lambda \hat{\phi}(z) = \hat{F}(z + \hat{\phi}(z)), \tag{19}$$

where $\hat{L}_\Lambda z^k = Akz^k$, and $\hat{F} = \check{F}$. Substituting $\hat{F}(z) = \sum_{k=2}^{\infty} \hat{c}_k z^k$ and $\hat{\phi}(z) = \sum_{k=2}^{\infty} \hat{a}_k z^k$ gives the formal relation

$$\sum_{k=2}^{\infty} Ak\hat{a}_k z^k = \sum_{i=2}^{\infty} \hat{c}_i \left(z + \sum_{k=2}^{\infty} \hat{a}_k z^k \right)^i. \tag{20}$$

Note that the left-hand side of (20) is simply $Az\hat{\phi}'(z)$ (still only in a formal sense). Also note that all appearing coefficients are non-negative due to the majorization process. Therefore the question regarding the convergence of $\hat{\phi}$ is reduced to that of the convergence of solution to the real ODE

$$\hat{\phi}'(x) = (Ax)^{-1}\hat{F}(x + \hat{\phi}(x)), \qquad \hat{\phi}(0) = 0. \tag{21}$$

Consider the partial sums $\hat{\phi}_d(x) = \sum_{k=2}^{d} \hat{a}_k x^k$. By (21), we have

$$0 \le \hat{\phi}'_{d+1}(x) \le (Ax)^{-1} \hat{F}(x + \hat{\phi}_d(x)) \qquad (0 \le x),$$

which yields the following estimate

$$0 \le \hat{\phi}_{d+1}(x) \le x\hat{\phi}'_{d+1}(x) \le A^{-1} \hat{F}(x + \hat{\phi}_d(x)), \tag{22}$$

Since $\hat{F}(x) = \hat{c}_2 x^2 + \ldots$, there are positive constants r_0 and B_0 such that $0 \le \hat{F}(x) \le B_0 x^2$ for all $0 \le x \le 2r_0$. Also, since $\hat{\phi}_d(x) = \hat{a}_2 x^2 + \cdots + \hat{a}_d x^d$, we can choose r_0 small enough to ensure that $0 \le \hat{\phi}_d(x) \le r_0$ for all $0 \le x \le r_0$. These estimates give

$$0 \le \hat{\phi}_{d+1}(x) \le A^{-1} B_0 (r_0 + r_0)^2 = A^{-1} B_0 (2r_0)^2, \tag{23}$$

for all $0 \le x \le r_0$. By selecting $r_0 \le A/(4B_0)$, we have shown the induction step $\hat{\phi}_d(x) \le r_0 \implies \hat{\phi}_{d+1}(x) \le r_0$ for all $0 \le x \le r_0$. It follows that $\hat{\phi}(x) \le r_0$ for all $0 \le x \le r_0$, which settles the question of convergence of the change of variables $\zeta + \phi(\zeta)$.

9 Convergence of the normal form

All that remains is to prove the convergence of the nonlinear component G of the normal form. Recall that G is recursively defined by

$$G_i(\zeta) = \big\langle F_i(\zeta + \phi(\zeta)) \big\rangle_{\mathbb{U}_\ell} - \sum_{j=1}^{n} \frac{\partial \phi_i}{\partial \zeta_j}(\zeta) G_j(\zeta) \qquad (i = 1, \ldots, n). \tag{24}$$

As there are no small divisors to consider, the existence of a formal solution to (24) is immediate. The question of convergence, however, is complicated by the fact that the recursive formula is made up of two separate contributing terms. Following the spirit of the previous section, we will use majorization techniques to establish the convergence of G.

We begin by reducing the dimension of the range of the problem by considering the majorant system

$$\tilde{G}_i(\zeta) = \big\langle \tilde{F}(\zeta + \tilde{\phi}(\zeta)) \big\rangle_{\mathbb{U}_\ell} + \sum_{j=1}^{n} \frac{\partial \tilde{\phi}_i}{\partial \zeta_j}(\zeta) \tilde{G}_j(\zeta) \qquad (i = 1, \ldots, n), \tag{25}$$

where $\tilde{\phi}$ solves (15), and $F_i \prec \tilde{F}$. This can be solved by a formal power series

$$\tilde{G}_i(\zeta) = \sum_{|m|=2\ell}^{\infty} \tilde{g}_{i,m} \zeta^m, \tag{26}$$

Note that, by our construction of the normal form, we know that the leading coefficients of $\tilde{G}$ with $|m| < 2\ell$ are zero. Furthermore, since $\tilde{\phi}_1 = \cdots = \tilde{\phi}_n$, the right-hand side of (25) is independent of i, and we have $\tilde{G}_1 = \cdots = \tilde{G}_n$. Reducing the dimension of the domain of the problem is achieved by considering the one-dimensional functional equation

$$\hat{G}(z) = \hat{F}(z + \hat{\phi}(z)) + n \cdot \hat{\phi}'(z) \hat{G}(z), \tag{27}$$

where $\hat{\phi}$ solves (19). Again, this can be solved by a formal power series

$$\hat{G}(z) = \sum_{k=2\ell}^{\infty} \hat{g}_k z^k, \tag{28}$$

where the coefficients g_k can be explicitly solved for by rearranging the terms of (27) into

$$\hat{G}(z) = (1 - n\hat{\phi}'(z))^{-1} \hat{F}(z + \hat{\phi}(z)). \tag{29}$$

This expression is valid provided that $1 - n\hat{\phi}'(z)$ is invertible. But, since $\hat{\phi}'(z) = 2\hat{a}_2 z + \ldots$, we can always arrange this by restricting ourselves to sufficiently small z. In other words, the radius of convergence of $\hat{G}(z)$ is at least as large as the smallest radius of convergence of $\hat{F}(z + \hat{\phi}(z))$ and $(1 - n\hat{\phi}'(z))^{-1}$, and is thus positive.

10 The solutions of the normal form

In this section, we will begin by proving a result on the structure of G using information obtained in Section 9. We will use this result to show that the solutions of the normal form act very much like the solutions to the linearized system.

Proposition 10.1 *Under the same conditions as in Theorem 5.1, and given $r_2 < r_1$, there exists a positive K_2 such that, in the open ball $B(0, r_2) = \{y\colon |y| < r_2\}$, we have*

$$|G_i(y)| \le K_2 \max_{i=1,\ldots,p} \{|y_{u_i}|^\ell\} \max_{i=1,\ldots,q} \{|y_{s_i}|^\ell\} \qquad (i = 1, \ldots, n).$$

Proof: Let $G_i(\zeta) = \sum_{m \in \mathbb{U}_\ell} g_{i,m} \zeta^m$, and consider the majorants

$$\tilde{G}(\zeta) = \sum_{m \in \mathbb{U}_\ell} \tilde{g}_m \zeta^m; \qquad \tilde{g}_m = \max_{i=1,\ldots,n} \{|g_{i,m}|\},$$
$$\hat{G}(z) = \sum_{k \ge 2\ell} \hat{g}_k z^k; \qquad \hat{g}_k = \sum_{|m|=k} \tilde{g}_m.$$

We clearly have $G_i \prec \tilde{G} \prec \hat{G}$. Let $\beta(\zeta) = \max_{i=1,\ldots,p}\{|\zeta_{u_i}|\} \max_{i=1,\ldots,q}\{|\zeta_{s_i}|\}$, $\sigma_u(m) = \sum_{i=1,\ldots,p} m_{u_i}$, $\sigma_s(m) = \sum_{i=1,\ldots,q} m_{s_i}$, and suppose that $|\zeta| < r_2 < r_1$. Then we have

$$\begin{aligned}
|G_i(\zeta)| = \left| \sum_{m \in \mathbb{U}_\ell} g_{i,m} \zeta^m \right| &\le \sum_{m \in \mathbb{U}_\ell} |g_{i,m}| |\zeta_1|^{m_1} \ldots |\zeta_n|^{m_n} \le \sum_{m \in \mathbb{U}_\ell} \tilde{g}_m |\zeta_1|^{m_1} \ldots |\zeta_n|^{m_n} \\
&\le \sum_{m \in \mathbb{U}_\ell} \tilde{g}_m \left(\max_{i=1,\ldots,p} \{|\zeta_{u_i}|\} \right)^{\sum_{i=1}^q m_{u_i}} \left(\max_{i=1,\ldots,q} \{|\zeta_{s_i}|\} \right)^{\sum_{i=1}^q m_{s_i}} \\
&= \max_{i=1,\ldots,p} \{|\zeta_{u_i}|^\ell\} \max_{i=1,\ldots,q} \{|\zeta_{s_i}|^\ell\} \sum_{m \in \mathbb{U}_\ell} \tilde{g}_m \max_{i=1,\ldots,p} \{|\zeta_{u_i}|^{\sigma_u(m)-\ell}\} \max_{i=1,\ldots,q} \{|\zeta_{s_i}|^{\sigma_s(m)-\ell}\} \\
&\le \beta(\zeta)^\ell \sum_{m \in \mathbb{U}_\ell} \tilde{g}_m |\zeta|^{|m|-2\ell} = \beta(\zeta)^\ell \sum_{k \ge 2\ell} \hat{g}_k |\zeta|^{k-2\ell} = \beta(\zeta)^\ell |\zeta|^{-2\ell} \sum_{k \ge 2\ell} \hat{g}_k |\zeta|^k
\end{aligned}$$

Now we will use that fact that $\hat{G}$ is analytic. Thus the coefficients $\hat{g}_k$ satisfy $\hat{g}_k \leq DL^k$ for some positive constants D and L. Continuing the estimates, we have

$$\begin{aligned}
|G_i(\zeta)| &\leq \beta(\zeta)^\ell |\zeta|^{-2\ell} \sum_{k\geq 2\ell} \hat{g}_k |\zeta|^k \leq \beta(\zeta)^\ell |\zeta|^{-2\ell} \sum_{k\geq 2\ell} DL^k |\zeta|^k \\
&= \beta(\zeta)^\ell |\zeta|^{-2\ell} D \sum_{k\geq 2\ell} (L|\zeta|)^k = \beta(\zeta)^\ell |\zeta|^{-2\ell} D \frac{(L|\zeta|)^{2\ell}}{1-L|\zeta|} \\
&\leq \beta(\zeta)^\ell \frac{DL^{2\ell}}{1-Lr_2} = K_2 \beta(\zeta)^\ell = K_2 \max_{i=1,\ldots,p} \{|\zeta_{u_i}|^\ell\} \max_{i=1,\ldots,q} \{|\zeta_{s_i}|^\ell\},
\end{aligned}$$

which completes the proof. □

In what follows, we will let $\mathfrak{B}_r$ denote the closed n-box centered at the origin, and having radius r. We will refer to the face $\{\zeta_r \in \mathfrak{B} \colon \zeta_{s_1} = r\}$ as the *lid* of the box $\mathfrak{B}_r$. Recall that s_1 is the index of the negative eigenvalue of the smallest modulus. We will also introduce the constant $\kappa = K_2 r^{2\ell-1}$, which should be thought of as being small compared to the minimal distance between the eigenvalues: $\kappa \ll \min\{||\lambda_i| - |\lambda_j|| \colon i \neq j\}$. We also demand that κ be small compared to the minimal distance between the eigenvalues and the origin: $\kappa \ll \min\{|\lambda_{s_1}|, |\lambda_{u_1}|\}$. This can clearly be arranged by taking r sufficiently small or, if $r < 1$, by taking ℓ large. We begin by stating a lemma which establishes an important dominance property:

Lemma 10.2 *For all trajectories $\psi(\zeta, t)$ of $\dot{y} = \Lambda y + G(y)$ starting from the lid of $\mathfrak{B}_r$, we have*

$$\psi_{s_1}(\zeta, t) \geq |\psi_{s_i}(\zeta, t)| \qquad (i = 2, \ldots, q)$$

throughout the entire box.

Proof: Using Proposition 10.1, the differential equations for $\psi_{s_i}(\zeta, t)$ can be enclosed by the differential inequalities

$$\begin{aligned}
|\dot{\psi}_{s_i}(\zeta, t) - \lambda_{s_i} \psi_{s_i}(\zeta, t)| &= |G_{s_i}(\psi(\zeta, t))| \\
&\leq K_2 \max_{i=1,\ldots,p} \{|\psi_{u_i}(\zeta, t)|^\ell\} \max_{i=1,\ldots,q} \{|\psi_{s_i}(\zeta, t)|^\ell\} \qquad (30) \\
&\leq K_2 r^{2\ell-1} \max_{i=1,\ldots,q} \{|\psi_{s_i}(\zeta, t)|\} = \kappa \max_{i=1,\ldots,q} \{|\psi_{s_i}(\zeta, t)|\}.
\end{aligned}$$

Initially, we have $r = \psi_{s_1}(\zeta, 0) \geq |\psi_{s_i}(\zeta, 0)|$, and by the differential inequalities (30) it follows that, if $\psi_{s_1}(\zeta, 0) = |\psi_{s_i}(\zeta, 0)|$, then $|\psi_{s_i}|$ decreases faster than ψ_{s_1}. Now suppose that after some positive time $t^\star$, we have the first occurrence of the situation $\psi_{s_1}(\zeta, t^\star) = |\psi_{s_i}(\zeta, t^\star)|$ for some $i = 2, \ldots, q$. Then, from (30), we have

$$\begin{aligned}
|\dot{\psi}_{s_1}(\zeta, t^\star) - \lambda_{s_1} \psi_{s_1}(\zeta, t^\star)| &\leq \kappa \psi_{s_1}(\zeta, t^\star), \\
|\dot{\psi}_{s_i}(\zeta, t^\star) - \lambda_{s_i} \psi_{s_i}(\zeta, t^\star)| &\leq \kappa |\psi_{s_i}(\zeta, t^\star)|.
\end{aligned}$$

By the same reasoning as above, $|\psi_{s_i}|$ decreases faster than ψ_{s_1}. Hence $\psi_{s_1}(\zeta, t) \geq |\psi_{s_i}(\zeta, t)|$ for all $i = 2, \ldots, q$ throughout the entire box. □

It now follows that the ζ_{s_1}-component of the flow is monotonically decreasing within $\mathfrak{B}_r$:

Corollary 10.3 *For all trajectories $\psi(\zeta,t)$ of $\dot{y}=\Lambda y+G(y)$ starting from the lid of $\mathfrak{B}_r$, we have*

$$(\lambda_{s_1}-\kappa)\psi_{s_1}(\zeta,t)\leq\dot{\psi}_{s_1}(\zeta,t)\leq(\lambda_{s_1}+\kappa)\psi_{s_1}(\zeta,t)$$

and

$$re^{(\lambda_{s_1}-\kappa)t}\leq\psi_{s_1}(\zeta,t)\leq re^{(\lambda_{s_1}+\kappa)t}$$

throughout the entire box.

Proof: We simply note that $\max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|\}=\psi_{s_1}(\zeta,t)$ in (30). □

In what follows, we will repeatedly utilize the following lemma, which is easily proved by e.g. the method of variation of parameters.

Lemma 10.4 *The linear ODE $\dot{z}=\lambda z+\varepsilon e^{\mu t}$ has the following solution:*

$$z(t)=z(0)e^{\lambda t}+\varepsilon\frac{e^{\lambda t}-e^{\mu t}}{\lambda-\mu}.$$

Regarding the remaining stable components of $\psi(\zeta,t)$, we have the following:

Lemma 10.5 *Given $\ell>|\lambda_{s_q}|/(|\lambda_{s_1}|-\kappa)$, there are positive α_i such that, for all trajectories $\psi(\zeta,t)$ of $\dot{y}=\Lambda y+G(y)$ starting from the lid of $\mathfrak{B}_r$, we have for all $i=2,\ldots,q$*

$$|\psi_{s_i}(\zeta,t)-\zeta_{s_i}e^{\lambda_{s_i}t}|\leq\frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{s_i}t}$$

throughout the entire box.

Proof: Using Lemma 10.2 and Corollary 10.3 together with (30), we can enclose the differential equation for ψ_{s_i} by

$$\begin{aligned}|\dot{\psi}_{s_i}(\zeta,t)-\lambda_{s_i}\psi_{s_i}(\zeta,t)| &= |G_{s_i}(\psi(\zeta,t))|\\ &\leq K_2\max_{i=1,\ldots,p}\{|\psi_{u_i}(\zeta,t)|^\ell\}\max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|^\ell\}\\ &\leq K_2r^\ell\max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|^\ell\}\leq K_2r^\ell|\psi_{s_1}(\zeta,t)|^\ell\\ &\leq K_2r^\ell\Big(re^{(\lambda_{s_1}+\kappa)t}\Big)^\ell=\kappa re^{\ell(\lambda_{s_1}+\kappa)t}.\end{aligned}$$

Using Lemma 10.4, we can explicitly solve for a bound on the perturbation from the linear flow:

$$|\psi_{s_i}(\zeta,t)-\zeta_{s_i}e^{\lambda_{s_i}t}|\leq\kappa r\left|\frac{e^{\lambda_{s_i}t}-e^{\ell(\lambda_{s_1}+\kappa)t}}{\lambda_{s_i}-\ell(\lambda_{s_1}+\kappa)}\right|.$$

By our choice of ℓ, it is clear that there exist positive α_i satisfying $0<\alpha_i\leq\lambda_{s_i}-\ell(\lambda_{s_1}+\kappa)$ $(i=2,\ldots,q)$. This implies that

$$|\psi_{s_i}(\zeta,t)-\zeta_{s_i}e^{\lambda_{s_i}t}|\leq\frac{\kappa r}{\alpha_i}\left|e^{\lambda_{s_i}t}-e^{\ell(\lambda_{s_1}+\kappa)t}\right|\leq\frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{s_i}t},$$

which completes the proof. □

Remark 6 *Naturally, the lemma is also true for $i=1$.*

Remark 7 *If we choose* $\ell > (|\lambda_{s_q}|+1)/(|\lambda_{s_1}|-\kappa)$*, then we can take* $\alpha_i > 1$*, which is used in the estimates of Theorem 5.2.*

Turning to the unstable coordinates, the situation is slightly more delicate. As an example, it is not true that a trajectory will always exit the box through the face corresponding to the strongest expanding coordinate ζ_{u_p}. To illustrate this fact, let us consider the completely linear case $\dot{\zeta} = \Lambda\zeta$ with two unstable directions, ζ_{u_1} and ζ_{u_2}, and assume that a trajectory enters the lid of the box with

$$|\zeta_{u_1}| > r\big(\frac{|\zeta_{u_2}|}{r}\big)^{\lambda_{u_1}/\lambda_{u_2}}. \tag{31}$$

In this situation, even though $\lambda_{u_1} < \lambda_{u_2}$, the trajectory will exit through the face $\{\zeta \in \mathfrak{B}_r : |\zeta_{u_1}| = r\}$. When both quantities of (31) are equal, the trajectory will exit through the intersection of both faces, i.e., through an edge of the box.

Returning to the non-linear situation at hand $\dot{\zeta} = \Lambda\zeta + G(\zeta)$, the dividing lines become inflated as illustrated in Figure 5. Trajectories starting from these *uncertain regions* may exit the box through any one of several faces of $\mathfrak{B}_r$, and with our limited knowledge of G, it is impossible to tell which. Any trajectory starting outside these regions, however, will have a well-defined face of exit.

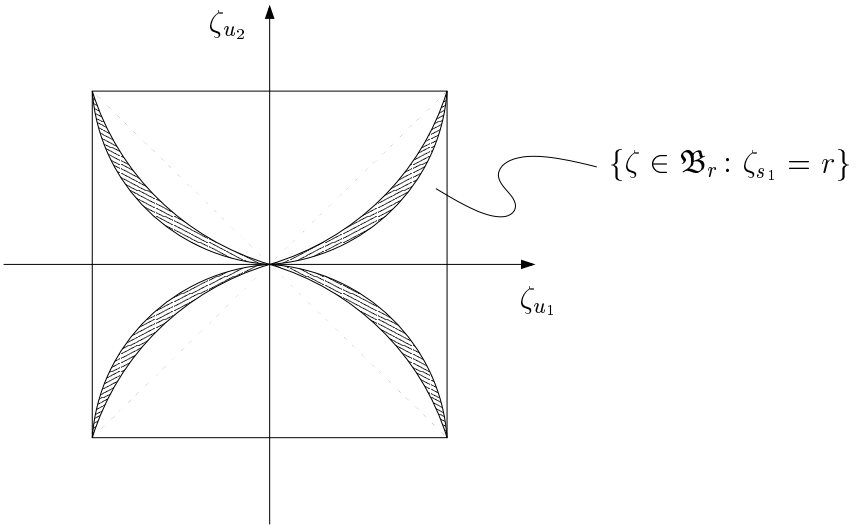

Figure 5: The uncertain regions of the lid of $\mathfrak{B}_r$ with $0 < \lambda_{u_1} < \lambda_{u_2}$.

Another complication is that, when $|\psi_{u_i}(\zeta,t)| \ll |\psi_{u_j}(\zeta,t)|$, we might very well have a situation where G_{u_i} is completely dominated by e.g. a term of the form $a\zeta_{u_j}^k$, where $i \neq j$ and $k \geq \ell$. This means that the following situation could arise:

$$\begin{aligned}\dot{\psi}_{u_i}(\zeta,t) &= \lambda_{u_i}\psi_{u_i}(\zeta,t) + G_{u_i}(\psi(\zeta,t)) \\ &\approx \lambda_{u_i}\psi_{u_i}(\zeta,t) + a\psi_{u_j}(\zeta,t)^k \approx a\psi_{u_j}(\zeta,t)^k \approx G_{u_i}(\psi(\zeta,t)),\end{aligned}$$

which shows that the u_i-coordinate of the normal form has no resemblance to its linear part. This makes a detailed analysis of the corresponding flow somewhat subtle.

A convenient concept in the forthcoming analysis is that of the *dominating unstable component*. This is simply the currently largest unstable component, which we label with the symbol $\hat{\imath}$:

$$|\psi_{u_{\hat{\imath}}}(\zeta,t)| = \max\{|\psi_{u_i}(\zeta,t)|\colon i=1,\ldots,p\}.$$

Note that the dominating unstable component may change along an orbit flowing through the box: $\hat{\imath} = \hat{\imath}(\zeta,t)$.

We begin our treatment of the unstable components by noting that the dominating unstable component acts very much like its linear counterpart.

Lemma 10.6 *While $\psi_{u_{\hat{\imath}}}(\zeta,t) \in \mathfrak{B}_r$ is the dominating unstable component of the trajectory $\psi(\zeta,t)$ of $\dot{y} = \Lambda y + G(y)$, we have*

$$(\lambda_{u_{\hat{\imath}}} - \sigma\kappa)\psi_{u_{\hat{\imath}}}(\zeta,t) \le \dot{\psi}_{u_{\hat{\imath}}}(\zeta,t) \le (\lambda_{u_{\hat{\imath}}} + \sigma\kappa)\psi_{u_{\hat{\imath}}}(\zeta,t)$$

and

$$\psi_{u_{\hat{\imath}}}(\zeta,t_0)e^{(\lambda_{u_{\hat{\imath}}}-\sigma\kappa)(t-t_0)} \le \psi_{u_{\hat{\imath}}}(\zeta,t) \le \psi_{u_{\hat{\imath}}}(\zeta,t_0)e^{(\lambda_{u_{\hat{\imath}}}+\sigma\kappa)(t-t_0)}.$$

Here $\sigma = \operatorname{sign}(\psi_{u_{\hat{\imath}}}(\zeta,t))$, and t_0 is the first time $\psi_{u_{\hat{\imath}}}$ becomes dominating.

Proof: Using Proposition 10.1, the differential equation for $\psi_{u_{\hat{\imath}}}(\zeta,t)$ can be enclosed by the differential inequality

$$\begin{aligned}
|\dot{\psi}_{u_{\hat{\imath}}}(\zeta,t) - \lambda_{u_{\hat{\imath}}}\psi_{u_{\hat{\imath}}}(\zeta,t)| &= |G_{u_{\hat{\imath}}}(\psi(\zeta,t))| \\
&\le K_2 \max_{i=1,\ldots,p}\{|\psi_{u_i}(\zeta,t)|^\ell\} \max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|^\ell\} \\
&\le K_2 r^{2\ell-1} \max_{i=1,\ldots,p}\{|\psi_{u_i}(\zeta,t)|\} = \kappa|\psi_{u_{\hat{\imath}}}(\zeta,t)|,
\end{aligned}$$

which translates into

$$(\lambda_{u_{\hat{\imath}}} - \sigma\kappa)\psi_{u_{\hat{\imath}}}(\zeta,t) \le \dot{\psi}_{u_{\hat{\imath}}}(\zeta,t) \le (\lambda_{u_{\hat{\imath}}} + \sigma\kappa)\psi_{u_{\hat{\imath}}}(\zeta,t),$$

where $\sigma = \operatorname{sign}(\psi_{u_{\hat{\imath}}}(\zeta,t))$. The second statement of the lemma follows by integration. □

By the same reasoning as in the proof of Lemma 10.2, it follows that no weaker unstable component can ever overtake the dominating component:

$$i < \hat{\imath}(\zeta,t_0) \quad \Longrightarrow \quad |\psi_{u_i}(\zeta,t)| < |\psi_{u_{\hat{\imath}(\zeta,t_0)}}(\zeta,t)| \qquad (t_0 \le t).$$

Lemma 10.6 immediately gives a crude upper bound on the time required to exit the box $\mathfrak{B}_r$.

Corollary 10.7 *Let $\hat{\imath}$ be the dominating unstable component at time t_0,i.e., let $|\psi_{u_{\hat{\imath}}}(\zeta,t_0)| = \max\{|\psi_{u_i}(\zeta,t_0)|\colon i=1,\ldots,p\}$. Then the flow-time required to exit the box $\mathfrak{B}_r$ is bounded from above by*

$$\tau_e \le t_0 + \frac{1}{\lambda_{u_{\hat{\imath}}} - \kappa}\log\frac{r}{|\psi_{u_{\hat{\imath}}}(\zeta,t_0)|}.$$

This bound is attained exactly when the dominating component remains dominating throughout the box.

Remark 8 *In the case where we have only one unstable direction, the corollary is valid with $t_0 = 0$. Combined with Lemma 10.6, Corollary 10.3, and Lemma 10.5, we then get the enclosures of Theorem 5.2.*

Returning to the situation where we have several unstable directions, we will make use of Corollary 10.3, which provided an upper bound on the dominating stable component.

Lemma 10.8 *There exist positive α_i such that, for all trajectories $\psi(\zeta,t)$ of $\dot{y}=\Lambda y+G(y)$ starting from the lid of $\mathfrak{B}_r$, we have for all $i=1,\ldots,p$*

$$|\psi_{u_i}(\zeta,t)-\zeta_{u_i}e^{\lambda_{u_i}t}|\leq\frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{u_i}t}$$

throughout the entire box.

Proof: Using Proposition 10.1 and Corollary 10.3, we can enclose the differential equation for ψ_{u_i} by

$$\begin{aligned}|\dot{\psi}_{u_i}(\zeta,t)-\lambda_{u_i}\psi_{u_i}(\zeta,t)| &= |G_{u_i}(\psi(\zeta,t))|\\ &\leq K_2\max_{i=1,\ldots,p}\{|\psi_{u_i}(\zeta,t)|^\ell\}\max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|^\ell\}\\ &\leq K_2 r^\ell\max_{i=1,\ldots,q}\{|\psi_{s_i}(\zeta,t)|^\ell\}\leq K_2 r^\ell|\psi_{s_1}(\zeta,t)|^\ell\\ &\leq K_2 r^{2\ell}e^{\ell(\lambda_{s_1}+\kappa)t}\leq\kappa r e^{\ell(\lambda_{s_1}+\kappa)t}.\end{aligned}$$

Using Lemma 10.4, we can explicitly solve for a bound on the perturbation from the linear flow:

$$|\psi_{u_i}(\zeta,t)-\zeta_{u_i}e^{\lambda_{u_i}t}|\leq\kappa r\left|\frac{e^{\lambda_{u_i}t}-e^{\ell(\lambda_{s_1}+\kappa)t}}{\lambda_{u_i}-\ell(\lambda_{s_1}+\kappa)}\right|.$$

We have already chosen κ small enough to guarantee that $\lambda_{s_1}+\kappa<0$. Therefore, we can find positive α_i such that $0<\lambda_{u_i}<\alpha_i\leq\lambda_{u_i}-\ell(\lambda_{s_1}+\kappa)$ $(i=1,\ldots,p)$. This implies that

$$|\psi_{u_i}(\zeta,t)-\zeta_{u_i}e^{\lambda_{u_i}t}|\leq\frac{\kappa r}{\alpha_i}\left|e^{\lambda_{u_i}t}-e^{\ell(\lambda_{s_1}+\kappa)t}\right|\leq\frac{\kappa r}{\alpha_i}(1-e^{-\alpha_i t})e^{\lambda_{u_i}t},$$

which completes the proof. □

Using these results, we can enclose the time a trajectory starting from the lid of $\mathfrak{B}_r$ spends inside the box.

Corollary 10.9 *For all trajectories $\psi(\zeta,t)$ of $\dot{y}=\Lambda y+G(y)$ starting from the lid of $\mathfrak{B}_r$, the flow-time required to exit the box $\mathfrak{B}_r$ is enclosed by the following inequalities:*

$$\tau_e^-(\zeta)\leq\tau_e(\zeta)\leq\tau_e^+(\zeta),$$

where $\tau_e^-(\zeta)$ and $\tau_e^+(\zeta)$ are defined as follows:

1. *Let $\hat{\imath}=\hat{\imath}(\zeta,0)$ (in a tie, take the largest index), and define*

$$\tau_{\hat{\imath}}^{\pm}(\zeta)=\frac{1}{\lambda_{u_{\hat{\imath}}}\mp\kappa}\log\frac{r}{|\zeta_{u_{\hat{\imath}}}|};$$

2. *For all $i>\hat{\imath}$ compute*

$$\tau_i^-(\zeta)=\frac{1}{\lambda_{u_i}}\log\frac{r}{|\zeta_{u_i}|+\frac{\kappa r}{\alpha_{u_i}}}\quad\text{and}\quad\tau_i^+(\zeta)=\frac{1}{\lambda_{u_i}}\log\frac{r}{\max\{0,|\zeta_{u_i}|-\frac{\kappa r}{\alpha_{u_i}}\}}.$$

Here the constants α_{u_i} are defined by $\alpha_{u_i}=\lambda_{u_i}-\ell(\lambda_{s_1}+\kappa)$;

3. Now define $\tau_e^-(\zeta) = \max\{\tau_j^-(\zeta)\colon j \geq \hat{\imath}\}$ *and* $\tau_e^+(\zeta) = \min\{\tau_j^+(\zeta)\colon j \geq \hat{\imath}\}$.

Note that the flow-time τ_e is infinite exactly when $\zeta_{u_1} = \cdots = \zeta_{u_p} = 0$, i.e., when we enter the box along the stable manifold.

It is now straight-forward to obtain bounds on the trajectory when leaving the box $\mathfrak{B}_r$. Using Corollary 10.9, we simply substitute the bounds on the exit-time $\tau_e(\zeta)$ into the enclosure bounds on the components of the flow. For the stable components, we use Corollary 10.3 and Lemma 10.5. For the unstable components, we use Lemma 10.6 and Lemma 10.8. This results in an interval enclosure I_i for each component, which we can possibly tighten by forming the intersection with the interval $[-r, r]$, i.e., $\psi_i(\zeta, \tau_e(\zeta)) \in I_i \cap [-r, r]$.

References

[Be78] Belitskii, G.R., *Equivalence and Normal Forms of Germs of Smooth Mappings*, Russ. Math. Surv. **33**, 107–177, 1978.

[Ha60] Hartman, P., *On local homeomorphisms of Euclidian spaces*, Bol. Soc. Math. Mexicana **5**, 220–241, 1960.

[Ha64] Hartman, P., Ordinary Differential Equations, John Wiley & Sons, New York, 1964.

[Hi76] Hille, E., Ordinary Differential Equations in the Complex Domain. John Wiley & Sons, New York, 1976.

[Ne64] Nelson, E., Topics in Dynamics I: Flows, Princeton Univerity Press, 1964.

[Po29] Poincaré, H., *Sur les proprietes des fonctions definies par les equations aux differences partielles*, Oeuvres **1**, Gautiers-Villars, Paris, 1929.

[Si52] Siegel, C.L., *Über die analytische Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung*, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl., 21–30, 1952.

[SM71] Siegel, C.L., Moser, J.K., Lectures on Celestial Mechanics, Springer-Verlag, 1971.

[Se85] Sell, G.R., *Smooth Linearization Near a Fixed Point*, Amer. J. Math. **107**, 1035–1091, 1985.

[St57] Sternberg, S., *Local contractions and a theorem of Poincaré*, Amer. J. Math. **79**, 809–824, 1957.

[St58] Sternberg, S., *On the structure of local homeomorphisms of Euclidian n-space – II*, Amer. J. Math. **80**, 623–631, 1958.

[Tu02] Tucker, W., *A Rigorous ODE Solver and Smale's 14th Problem*, Found. of Comp. Math. **2:1**, 53–117, 2002.