# Bifurcations in Discrete-Time Neural Networks – Controlling Complex Network Behaviour with Inputs

Der Technischen Fakultät

der Universität Bielefeld

vorgelegt von

**Robert Haschke**

**zur Erlangung des akademischen Grades**

**Doktor der Naturwissenschaften**

September 2003

# Acknowledgement

# Contents

# Notation

Throughout the thesis the following conventions are adopted:

- Scalars are denoted by simple characters: variables usually by $x, y, z$, parameters by greek letters, e.g. $\alpha, \beta, \gamma$, the time by $t$, and functions by $f(\cdot), \varphi(\cdot, \cdot)$.

- Vectors of scalars or functions are denoted by boldface characters $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{F}(\cdot) = (f_1(\cdot), \ldots, f_n(\cdot))$.

- Applications of usually scalar-valued functions, e.g. $\tanh : \mathbb{R} \to \mathbb{R}$, in their vector-valued form means componentwise application of the corresponding function, i.e. $\boldsymbol{\tanh}(\boldsymbol{x}) = (\tanh(x_1), \ldots, \tanh(x_n))^t$.

- Matrices, which represent linear operators, are denoted by upper Latin characters: $A, B, C$.

## List of symbols

| | |
|---|---|
| $\mathbb{R}$ | the set of real numbers |
| $\mathbb{R}^n$ | the n-dimensional real vector space |
| $\mathbb{C}$ | the set of complex numbers |
| $\mathbb{C}^n$ | the n-dimensional complex vector space |
| | |
| $\varphi, \psi$ | mappings representing the time-one-map of a discrete-time dynamical system |
| | |
| $\mathbb{T}$ | the time space, where $\mathbb{T} \in \{\mathbb{Z}, \mathbb{Z}_+, \mathbb{R}, \mathbb{R}_+\}$ |
| $t$ | the time argument $t \in \mathbb{T}$ |
| | |
| $x_i$ | the i-the component of the vector $\boldsymbol{x}$ |
| i | the imaginary unit |
| | |
| $(\cdot)^t$ | the transpose of a real vector or matrix |
| | or the $t$ fold application of a system's map $\varphi$ or $A$, i.e. $\varphi^t$ and $A^t$ |
| $(\cdot)^*$ | the conjugate transpose of a complex scalar, vector or matrix |
| $\langle \cdot, \cdot \rangle$ | a scalar product |
| $\|\cdot\|$ | a norm |
| $\lambda_i(A)$ | the i-th eigenvalue of $A$ |
| $\det A$ | the determinant of $A$ |
| | |
| $\varphi_x(\bar{x})$ | derivative of $\varphi$ with respect to $x$ evaluated at $x = \bar{x}$, i.e. $\left.\frac{\partial \varphi}{\partial x}\right|_{x=\bar{x}}$ |

# 1 Introduction

Neural network research is a steadily growing field of science attracting researchers from a wide range of disciplines, such as biology, physics, mathematics, computer science, and psychology. Their common goal is to gain insight into the working principles of the brain, which is one of the most fascinating scientific challenges of our time. Nowadays common agreement exists, that the brain is built from an enormous number of neurons which communicate mainly via electrical signals, called action potentials. Although the working principles of single neurons, which can be understood as very simple processing units, and their communication among each other is increasingly well understood, it is not yet clear how these structures result in such complex phenomena like memory, recognition, attention, consciousness, and emotions. Obviously there exists a huge gap between the very limited capabilities of single neurons and the observed complex behaviour of the brain. Presumably this gap cannot be overcome employing the reductionist approach, which is the predominating scientific tenet to understand natural phenomena. It bases on the idea, that each complex system can be reduced to smaller subsystems, whose behaviour can be described by relatively simple fundamental laws. Finally, the original system can be composed from these subsystems again and its behaviour can be completely explained and predicted given the states of its components. The overwhelming success of this approach is reflected by the ubiquitous technical use of the fundamental physical laws, and began with the discovery of the Newtonian laws, which are able to describe the complex motions of the solar system.

A major prerequisite for the application of the reductionist approach is a weak interaction between the components, which constitute a complex system. When Poincaré studied the stability of the solar system in the end of the 19th century, he found that Newton's equations admit very complex and even chaotic solutions if applied to multi-body-problems and thus heavily differ from the smooth elliptic solutions obtained for two-body-problems, where all interactions with other bodies are neglected. Thus, if the prerequisite of weak interaction is violated, the reductionist approach is not applicable anymore. This poses a major problem for the study of such complex networks like the brain, because it seems to be impossible to decompose it into smaller subsystems, which are only weakly interacting. Most of the about $10^{11}$ neurons of the brain are connected with $10^3$ to $10^4$ other neurons, and their dendritic trees cover vast areas of the brain [Kandel et al., 1991]. Recently developed brain imaging techniques strongly indicate, that most mental activities involve simultaneous activation of many brain areas.

The actual challenge is to understand the emergence of complex dynamical behaviour within large networks of strongly interacting, but simple components. Neural networks represent an especially interesting example of such networks – power supply networks, commercial or social networks are other examples. While researchers are mainly interested in prediction and control of the latter systems, the relatively new research field of artificial neural networks aims at imitating high level brain functions employing medium-sized networks of artifical neurons. The goal is on the one hand to identify necessary architectural components to accomplish specific tasks and on the other hand to find general learning rules, which allow a given architecture to adapt or evolve itself in a self-organising manner in order to perform a specific task from a whole class of possible tasks. The search for successful network architectures does not need to incorporate as much neuro-physiological facts as possible, rather it should identify topological structures, which are functionally relevant and construct a model which is as simple as possible and is easily amenable to simulation and mathematical analysis. Nevertheless, the biological example offers a helpful

orientation during this search – after all the brain is a very sophisticated information-processing system, which evolved by nature since millions of years.

During the last years a vast amount of network architectures, neuron types, and learning rules were proposed and it is beyond the scope of this work to provide a taxonomy of the different models. Instead we point out some principal axes along which neural network architectures can be distinguished. For a detailed taxonomy we refer to Kremer [2001].

### Feedforward vs. Recurrent Neural Networks

The most important distinction which has to be made for networks concerns the direction of processing they use. A feedforward network possesses a connection structure, which can be ordered into layers, and it propagates its inputs along this unidirectional path of layers. These networks do not have internal feedback, i.e. loops within the connection structure which allow cyclic propagation of activity. Therefore feedforward networks represent simple nonlinear input-output mappings. On the contrary, recurrent neural networks (RNNs) have feedback and represent dynamical systems, whose dynamics covers all ranges from stationary to oscillatory and even chaotic behaviour. Actually it has been shown, that recurrent neural networks can approximate arbitrarily well any given nonlinear operator, i.e. any mapping of one time-varying sequence (a function of time) to another [Chen and Chen, 1995]. Recently it has been proven, that RNNs can approximate dynamical systems that change continuously or switch between several characteristic behaviours [Back and Chen, 2002]. The ability of RNNs to memorise past input events allows them to draw conclusions from the past input sequence and therefore to react in anticipation of future events, especially they are able to generate non-trivial output sequences autonomously. This capability often has been stressed as generic to use them for identification, modelling, and control of dynamical systems [Narendra and Parthasarathy, 1990; Kolen and Kremer, 2001]. Neural networks have been successfully applied for time series prediction, identification and classification, robot control, object recognition and binding tasks in visual images, associative memories, and nonlinear optimisation. Areas of application range from artificial intelligence to biology, chemistry, physics, bioinformatics, and control. For a coarse overview of recent applications we refer to Hammer and Steil [2002]; Kolen and Kremer [2001].

### Continuous-time vs. Discrete-time Recurrent Neural Networks

Another import distinction is between networks whose evolution is a continuous process and those where the state is updated at discrete time instants only. Clearly continuous-time networks are more appropriate in modelling natural systems, i.e. physical or biological systems, which involve inherently in continuous time. These systems are described by differential equations (ODEs) and there exists a large amount of mathematical tools from differential analysis and dynamical system's theory, which allow an analytical investigation of these systems. On the contrary, discrete-time systems are more easily to simulate on digital computers, because the expensive integration of ODEs can be dropped. This is a major advantage, because simulations of complex dynamical systems become the most important tool to investigate their behaviour if analytical methods fail. Further discrete-time systems allow an investigation of symbolic dynamics, i.e. systems with discrete state space, which emerge naturally with identification tasks of artificial grammars. Many simulations and theoretical results indicate that both types of RNNs are equally powerful with regard to the tasks mentioned in the previous paragraph. Their computational power bases on the *nonlinear* sigmoidal transfer function of single neurons but primarily on the connection structure of the network. Neither the exact modelling of the neurons concerning their biological cellular compartments, ion channels and conductances, nor the type of state evolution are essential

for the emergence of complex behaviour. In the present work we will study discrete-time recurrent neural networks, because they allow fast simulations on today's computers. Most results of this work easily translate to continuous-time RNNs as well; indeed some of our results originate from continuous-time counterparts [Hoppensteadt and Izhikevich, 1997] or were reported for continuous-time networks elsewhere [Beer, 1995].

## Representation of computational results

The third important question is, how the results of a computation should be represented in a neural network model. Hopfield suggested an auto-associative memory model, which was able to memorise several desired patterns as fixed points of the underlying dynamics [Hopfield, 1984; Grossberg, 1988]. When started with an arbitrary initial state, e.g. a noisy input pattern, the convergent dynamics usually approaches one of the attractors, recalling the corresponding pattern. Associative memories are examples of multiple attractor networks, but sometimes a RNN is desirable which possesses a unique global attractor. This type of network is often used in optimisation tasks, where the single attractor corresponds to the single global optimum. Input patterns of these networks alter their dynamics and thus the location of the global fixed point.

Most existing literature considers fixed points attractors as computational objects of RNNs, which allow an easy detection of the termination of a computation. Nevertheless all other types of attractors, e.g. periodic or chaotic attractors, could be used as memorised patterns as well. Clearly in this case, the neuron's activities cannot be used as information sources anymore, but rather more global properties of the attractor have to be used. Due to the dynamical structure of these attractors, it is possible to encode more information per neuron than with fixed point attractors – for example, in case of a periodic attractor we could use average activity (corresponding to neuron activity of a fixed point attractor), frequency and phase of each neuron as information carriers. There have been many attempts to understand the role of non-fixed-point-attractors for information processing in animal and human brains [Skarda and Freeman, 1987; Gray et al., 1989; Friedrich and Laurent, 2001; Kuhn et al., 2003] and for computation [Hoppensteadt and Izhikevich, 2000; Stollenwerk and Pasemann, 1996, for more references see the review by Elbert et al., 1994]. For example, the temporal correlation hypothesis [Milner, 1974; von der Malsburg, 1981] – a popular but controversially discussed approach to feature binding in the brain – postulates a binding between remote neural cell assemblies, if they have synchronised activity. Appropriate models employ networks of coupled oscillators and study their synchronisation and desynchronisation in dependence of the presented input patterns [Campbell and Wang, 1996; Wang and Terman, 1997; Tonnelier et al., 1999; Kazanovich and Borisyuk, 2002].

Sometimes emphasis is placed on the evolving trajectory rather than the final attractor itself. For example, sensorimotor tasks require a complex and continuous trajectory, which can be composed of several shorter trajectories (or subtasks) as well. Such trajectories naturally arise if the attractor structure together with its bassins changes in state space, such that the trajectory is captured by different attractors in series. Rather than assigning representational content to the attractors themselves, the behaviour of an acting individual is then characterised by the structure of the state space, i.e. the bassins of attractions and the possible behavioural trajectories therein. If the individuum interacts with its environment, this space is subject to changes due to varying sensory signals along the trajectory. Therefore recent approaches to cognition and behaviour – additionally to the agent's brain – include its body and environment into the dynamical system under consideration. A discussion of the resulting theory of situated and embodied agents can be found in Clark [1997], while several concrete models of a wide range of cognitive phenomena are reviewed in Port and van Gelder [1995]; Beer [2000]. A detailed discussion of the dynamics of such a situated and embodied system is carried out by Beer [2003], who studies a simple model agent, whose

behaviour and cognitive capabilities arise from nontrivial trajectories in the composed dynamical system. For his agent indeed holds the proverb "The journey is the reward".

### Learning

One of the most appealing properties of neural networks is their learning capability, i.e. the possibility to *learn* a given task from a sufficient amount of training examples. The main body of research concentrates on three different learning concepts: gradient based methods, evolutionary learning and local learning. The first two methods aim at minimising a globally evaluated error functional, which incorporates the desired network behaviour.

Gradient based methods compute the sensitivity of the network output to small changes of the weights and adapt them in dependence of the observed output error. In case of feedforward networks this results in the well known backpropagation algorithm, which can be generalised to RNNs as well [Pearlmutter, 1995]. Recently Atiya and Parlos [2000] summarised many existing learning algorithms for recurrent neural networks within the common framework of constraint optimisation. All gradient based algorithms for recurrent networks suffer from an exponentially fast decreasing error signal, which hampers the learning of long-term dependencies.

On the contrary, evolutionary algorithms search for an optimal solution by trial and error: The performance of a whole population of potential parameter sets is evaluated according to the error functional and the best solutions are altered by means of crossover and random mutation to form a new population [Michaelewicz, 1996]. While this random search cannot get stuck in local minima – like gradient based methods often do – it is not suitable for large problems, because the resulting high-dimensional parameter space cannot be sampled efficiently. Further these algorithms are very sensible to the employed crossover and mutation techniques, the chosen fitness function and other relevant parameters. For a review of employed methods and applications see Husbands et al. [1997a,b]; Meeden and Kumar [1998].

Both, gradient based methods and evolutionary learning are global methods: they evaluate the overall fitness of a given network with respect to a given task. Local learning methods rely on locally available information only, e.g. Hebbian learning uses the pre- and postsynaptic activities to adapt a given weight. The generalised recirculation algorithm introduced by O'Reilly [1996] provides a common framework to understand many local activation-based learning rules as different approximations of the backpropagation algorithm.

## 1.1 Scope and Goals

The goal of this thesis is to contribute to the understanding of the emergence of complex dynamical behaviour in small recurrent neural networks. In various articles the dynamics of RNNs has been studied for different paradigms of artificial networks, including continuous-time and discrete-time networks with rate as well as spike coding. The first conducted studies heavily employed computer simulation techniques to get an overview of the possible dynamical repertoire [Markus and Westervelt, 1989; Renals and Rhower, 1990; Chapeau-Blondeau and Chauvet, 1992]. It was found, that already small networks exhibit complex dynamical behaviour. Wilson and Cowan [1972] proposed a model composed of two neurons to produce oscillatory behaviour and Wang [1991c] proved the existence of chaotic attractors in specific discrete-time networks of two neurons.

Consequently the question arises, how this rich repertoire of dynamical behaviour can be controlled and particularly which parameter sets cause a specific dynamics. The knowledge of these parameter sets, would allow us to directly switch between different dynamical regimes. In order

to tackle this challenging task, recent approaches – besides the purely simulative study – utilise techniques from dynamical systems theory and bifurcation analysis to gain a deeper understanding of the underlying processes [Beer, 1995; Tonnelier et al., 1999; Tiňo et al., 2001].

In this thesis we develop a general approach to analytically compute the bifurcation manifolds of a RNN, which separate different regimes of dynamical behaviour in the parameter space. To this aim we consider discrete-time additive recurrent neural networks, which are a standard model for rate-coded activity dynamics in biologically inspired neural networks. In the discussion of the resulting bifurcation manifolds we concentrate especially on oscillatory dynamics.

Experimental investigations of the neural systems, which underlie the rhythmic activities of biological systems, e.g. swimming and walking, led to the notion of the central pattern generator, CPG [Cohen et al., 1988]. In the past two decades a number of oscillatory neural networks have been developed to model biological CPGs. Examples include the lamprey CPG, which controls swimming motions [Wallén et al., 1992], and the salamander CPG, which is able to switch between a swimming and trotting gait [Ijspeert, 2001]. The characteristic behaviour of these oscillatory networks usually emerges from intrinsic oscillators, which are mutually coupled to evoke phase locking, i.e. all the oscillators eventually oscillate at a single frequency and maintain a constant pattern of relative phases [Hoppensteadt and Izhikevich, 2000]. In contrast, we rely on the fact that diverse dynamics, including oscillatory behaviour, emerges from classical connectionist networks as well, and we attempt to explain their underlying working principles.

Our analysis of two- and three-neuron networks can serve as a starting point to build more complex networks from these basic building blocks. We emphasise that the resulting complex systems will exhibit qualitatively different dynamical behaviour as soon as the interaction between subcomponents becomes too strong. Nevertheless a weak coupling allows to draw some conclusions for the overall network from the known dynamics of its components. For example, Hoppensteadt and Izhikevich [1997] studied several networks of weakly coupled components.

## 1.2 Plan of the manuscript

This introduction exposes complex dynamical systems theory as one of the most challenging fields of research of our time. It motivates the role of artificial neural networks within this context and defines the scope and goals of this thesis.

The next two chapters provide an introduction into dynamical systems theory and bifurcation analysis. While chapter 2 gives an overview of the fundamental terminology and techniques, which are to be used throughout this work, chapter 3 introduces a set of local fixed point bifurcations, which we will investigate in the context of recurrent neural networks. The prerequisites for these bifurcations to occur are presented and their dynamical properties are discussed. A single neuron with self-feedback serves as an example to demonstrate the employed techniques.

In chapter 4 we review some important stability results for time-varying linear systems, which can be employed to prove absolute stability for certain classes of nonlinear dynamical systems including RNNs. In this context, stability of matrices and matrix sets is a major topic. To provide the reader with the necessary terminology, we introduce some frequently used types of matrix stability and review some important theorems in appendix B.

In chapter 5 we identify classes of recurrent networks which have identical dynamical behaviour. We prove, that within a specific class of activation functions, there exists a direct relation between the network parameters – i.e. the connections weights and the inputs to the neurons – and the neuron's parameters, i.e. the actual shape of their activation function. This relationship allows to

study a RNN, employing a single activation function only, and to draw conclusions for any other network which has a different set of activation functions for its neurons. Further we employ the symmetry of common activation functions to infer equivalent dynamical behaviour for a class of weight matrices and corresponding input vectors.

Chapter 6 represents the main contribution of this thesis. We derive analytical expressions for the bifurcation manifolds of fixed point bifurcations in general discrete-time RNNs. To this aim we assume, that the network's weight connection are fixed at arbitrary values, and we consider the external inputs to the neurons as bifurcation parameters. Our expressions allow us to compute bifurcation diagrams in input space fast enough to enable an interactive exploration of these diagrams under variation of the connection weights. This in turn provides the basis to study bifurcation diagrams which include the weights as parameters as well. Our approach is not restricted to a certain class of RNNs – particularly it can be applied to continuous-time networks as well. Furthermore, in principle it allows to compute bifurcation manifolds of large RNNs, although the actual computation becomes very expensive in practice.

In chapter 7 we apply our results in order to compute and discuss the bifurcation diagrams of two-neuron networks. Particularly, we concentrate on weight matrices which are rotation matrices. We present a complete bifurcation diagram in the four-dimensional parameter space including the two external inputs as well as the rotation angle and the scaling factor of the weight matrix. The investigation of the observed bifurcation types reveals two basic mechanisms for the transition from quiescent to oscillatory behaviour. Due to their different properties, both mechanisms can be employed to control either the frequency or the amplitude of the emerging oscillatory behaviour.

In chapter 8 we present methods to study large recurrent networks as well. Although the actual computation of bifurcation manifolds becomes infeasible for networks with more than three neurons, we show that it is sometimes possible to decompose the network into smaller subnets, which can be easily analysed by our approach. Finally, in the spirit of the reductionist approach the dynamical behaviour of the original large network composes from the dynamics of the smaller subnets. As an example we apply the decomposition technique to a three-neuron network, for which we can compute the bifurcation manifolds directly as well. Finally we draw some general conclusions for recurrent neural networks, which result from the presented theory.

The concluding chapter 9 summarises the main results and discusses possible directions of future research.

# 2 Mathematical Preliminaries

This chapter provides fundamental terminology and techniques on dynamical systems theory which are to be used throughout this work. While there exists an exhaustive literature on this topic, we focus on those methods which are essential for this work and are connected to neural networks as dynamical systems. More detailed introductions to this field can be found in Devaney [1989], Reitmann [1996], Stuart and Humphries [1996], Sell and You [2002], and Kuznetsov [1995].

## 2.1 Dynamical Systems

A dynamical system mathematically formalises the concept of a deterministic *process* on the basis of the evolution of some state variables in time following a deterministic *evolution law*[1]. This notion of a dynamical system includes a set $X$ of possible values for these variables – the state space – and a deterministic law which describes their evolution in time $t \in \mathbb{T}$. Depending on whether $\mathbb{T} \subseteq \mathbb{R}$ or $\mathbb{T} \subseteq \mathbb{Z}$, we consider

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(\boldsymbol{x}(t), t) \quad \boldsymbol{x} \in X,\ t \in \mathbb{T} \subseteq \mathbb{R} \tag{2.1}$$

as a *continuous-time dynamical system* and

$$\boldsymbol{x}(t+1) = \boldsymbol{F}(\boldsymbol{x}(t), t) \quad \boldsymbol{x} \in X,\ t \in \mathbb{T} \subseteq \mathbb{Z} \tag{2.2}$$

as a *discrete-time dynamical system*. This rather general definition of a dynamical system has two major drawbacks. First, the existence of solutions for $t \to \infty$ is not guaranteed. Particularly solutions may escape to infinity within finite time. Because dynamical systems theory is focused on the asymptotic behaviour of solutions, existence of solutions $\boldsymbol{x}(t)$ of (2.1) resp. (2.2) for $t \to \infty$ has to be demanded for all initial states $\boldsymbol{x}_0 \in X$ within the state space $X$. Further, the solutions $\boldsymbol{x}(t)$ have to stay in $X$ for all $t \geq 0$.

Second, it is usually demanded that the evolution law $\boldsymbol{F}$ does not depend on time $t$ explicitly, ensuring that the evolution law is constant over time. Otherwise the dynamical system and its properties would change at each moment in time. If $\boldsymbol{F}$ does not depend on time $t$ explicitly, equations (2.1), (2.2) are called *autonomous*. Introducing a new state variable $\tau$, it is always possible to extend (2.1), (2.2) to autonomous equations:

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= \boldsymbol{F}(\boldsymbol{x}(t), \tau(t)) & \boldsymbol{x}(t+1) &= \boldsymbol{F}(\boldsymbol{x}(t), \tau(t)) \\ \dot{\tau}(t) &= 1 & \tau(t+1) &= \tau(t) + 1 \,. \end{aligned} \tag{2.3}$$

Modern literature defines a dynamical system by means of a family of maps $\varphi^t : X \to X$ where $t \in \mathbb{T}$ and $\mathbb{T}$ equals $\mathbb{R}_+$ or $\mathbb{R}$ resp. $\mathbb{Z}_+$ or $\mathbb{Z}$, which obey the following (semi)group equations:

$$\begin{aligned} \varphi^0 &= \mathrm{id} \\ \varphi^{t+s} &= \varphi^t \circ \varphi^s \quad \text{for all } t, s \in \mathbb{T} \,. \end{aligned}$$

---

[1] We consider dynamical systems with state space representation only, but there exist other approaches, i.e. the behavioural approach to dynamical systems [Polderman and Willems, 1998].

The map $\varphi^t$ is assumed to map initial states $\boldsymbol{x}_0 \in X$ into some state $\boldsymbol{x}(t) \in X$ at time $t$ and thus is called *evolution operator*. Note that the reduction of $\mathbb{T}$ to the set $\{\mathbb{R}_+, \mathbb{R}, \mathbb{Z}_+, \mathbb{Z}\}$ ensures existence of solutions for all $\boldsymbol{x}_0 \in X$ and $t \to \pm\infty$. Furthermore the family of evolution operators defines an autonomous system.

If $\mathbb{T}$ equals $\mathbb{R}$ or $\mathbb{Z}$ the dynamical system is called *invertible*, because for each $\varphi^t$ exists $(\varphi^t)^{-1} = \varphi^{-t}$. In these systems an initial state $\boldsymbol{x}_0$ uniquely defines the future states of the system as well as its past behaviour.

Because we concentrate on discrete-time systems, we just note, that an autonomous differential equation (2.1) defines a family of invertible evolution operators $\{\varphi^t\}_{t\in\mathbb{R}}$ and hence an invertible dynamical system if $\boldsymbol{F}$ is locally Lipschitz on some set $M \in \mathbb{R}^n$ and all solutions starting in $M$ are bounded. The latter condition ensures that the solutions of (2.1) are defined for all $t \in (-\infty, +\infty)$. More precisely, the function $t \mapsto \varphi^t(\boldsymbol{x}_0)$ is the solution $\boldsymbol{x}(t) = \varphi^t(\boldsymbol{x}_0)$ to the initial value problem (2.1) with $\boldsymbol{x}(t=0) = \boldsymbol{x}_0$.

For discrete-time systems (2.2) the situation is much simpler – we only have to ensure that the image of $X$ under $\boldsymbol{F}$ is a subset of $X$, i.e. $\boldsymbol{F} : X \to X$. This ensures existence of solutions and it holds $\varphi^t = F^t$ for all $t \in \mathbb{N}$ and $\varphi^0 \equiv \mathrm{id}$. If $\boldsymbol{F}$ is invertible, $\varphi^{-t} = (F^{-1})^t$ is defined as well and we obtain an invertible dynamical system. Hence, a discrete-time dynamical system is completely defined by its time-one-map $\varphi^1 \equiv F$.

We will usually consider $X \subset \mathbb{R}^n$ as state space and assume that the vector-function $\boldsymbol{F} : X \to X$ is sufficiently smooth for our purposes. More precisely, we will use the following additive neural network models

$$\dot{\boldsymbol{x}}(t) \qquad = -\boldsymbol{x}(t) + \boldsymbol{\sigma}\left(W_x \boldsymbol{x}(t) + W_u \boldsymbol{u}(t)\right) \tag{2.4}$$

respective

$$\boldsymbol{x}(t+1) = \qquad \boldsymbol{\sigma}\left(W_x \boldsymbol{x}(t) + W_u \boldsymbol{u}(t)\right), \tag{2.5}$$

where $\boldsymbol{x}(t)$ denotes the *state* or *activity* of the neurons at time $t$ and $\boldsymbol{u}(t)$ the respective input vector. $W_x$ and $W_u$ are weight matrices and $\boldsymbol{\sigma}$ denotes the element-wise application of the node output function, usually a sigmoid, i.e. a function satisfying $\sigma_- < \sigma(x) < \sigma_+$ and $\sigma'(x) > 0$ for some constants $\sigma_-, \sigma_+ \in \mathbb{R}$ and all $x \in \mathbb{R}$. Often an output vector $\boldsymbol{o}(t)$ is computed additionally as a linear combination of the state vector:

$$\boldsymbol{o}(t) = W_o \boldsymbol{x}(t). \tag{2.6}$$

The discrete-time system (2.5) can be regarded as the Euler-integration with time step $\Delta t = 1$ of (2.4) and thus is only a rough approximation of the continuous-time system. It has been shown, that it is not possible to derive any meaningful convergence results of the form "attractors of (2.5) are similar to those of (2.4)", because convergence goes with order $\mathcal{O}(\Delta t^r)$ or worse [Stuart and Humphries, 1996], which ensures similarity of attractors in the limit $\Delta t \to 0$. But because the time step $\Delta t = 1$ is fixed, both systems have different properties and consequently are considered independent of each other in the literature.

Note that the systems (2.4 – 2.5) are non-autonomous in general, because the inputs $\boldsymbol{u}(t)$ depend on $t$ and change the dynamics over time. Because we want to eliminate this additional uncertainty [2] and consider autonomous systems, we will use time-independent inputs $\boldsymbol{u}(t) = \mathrm{const}$ usually.

In order to handle the non-autonomous systems (2.4 – 2.5) as well, it is convenient to think of piecewise constant inputs and *quasistatic* inputs, i.e. inputs changing on a time-scale much slower

---

[2]  According to (2.3) an non-autonomous system possesses the additional dimension $\tau$ in its state space.

than the network dynamics. This allows us to discuss the dynamical behaviour in terms of a sequence of slowly changing autonomous systems. Quasistatic parameter changes guarantee, that the system's state will always converge to an attractor of the autonomous system [Hoppensteadt, 1993]. In the following – unless specified otherwise – we employ constant inputs and consequently obtain autonomous systems.

From the dynamical system's point of view the weight matrix $W_u$ in (2.4, 2.5) can be dropped if we consider $\boldsymbol{u}' = W_u \boldsymbol{u}$ as inputs instead. However, this transformation is relevant for applications having a fixed set of input vectors to allow for a scaling and combination of these fixed input vectors to more useful ones in the context of a particular network dynamics.

Both systems – in their autonomous version – define a dynamical system in terms of the above definition, because (2.4) inherits Lipschitz continuity and boundedness of the sigmoid function $\boldsymbol{\sigma}$ and (2.5) is a mapping from the bounded set $[\sigma_-, \sigma_+]^n$ to itself.

### 2.1.1 Orbits, invariant sets and stability

The evolution of a specific state $\boldsymbol{x}_0$ with time $t$, more precisely the map $t \mapsto \boldsymbol{x}(t; \boldsymbol{x}_0, t_0)$ with $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$ is called *trajectory* of $\boldsymbol{x}_0$ starting at time $t_0$. While trajectories of continuous-time systems are *curves* in the state space $X$ parameterised by the time $t$, trajectories of discrete-time systems are *sequences* in $X$. Neglecting the time order, the set of points described by the trajectory starting in $\boldsymbol{x}_0$ is called *orbit* of $\boldsymbol{x}_0$ and is denoted $\gamma(\boldsymbol{x}_0)$. If the dynamical system is invertible we distinguish between the *positive orbit* $\gamma^+(\boldsymbol{x}_0) = \{\varphi^t(\boldsymbol{x}_0) \,|\, t \in \mathbb{T}, t \geq 0\}$ and the *negative orbit* $\gamma^-(\boldsymbol{x}_0) = \{\varphi^t(\boldsymbol{x}_0) \,|\, t \in \mathbb{T}, t \leq 0\}$ as well.

Any two orbits of an invertible dynamical system either never intersect or are identical. For non-invertible systems it can be only ensured, that trajectories coincide in future, if they meet at some point $\boldsymbol{x}(t)$. Notice, that (2.5) is invertible if and only if both $\boldsymbol{\sigma}$ and $W_x$ are invertible, which will not be assumed in general within this work.

The collection of all possible orbits is called *phase portrait*. Because it is impossible to draw all orbits into a single figure, we will usually select and display several key trajectories to get a schematic view of the phase portrait. Drawing the phase portrait of a discrete-time dynamical system has the additional difficulty that orbits often are a collection of strictly separated points instead of connected curves. Hence it is necessary to visualise the membership of points to specific orbits. The simplest trajectories are fixed points and cycles, which are defined as follows:

**Definition 2.1** Let an autonomous dynamical system be given by $\{\varphi^t\}_{t \in \mathbb{T}}$, then:

(i) $\bar{\boldsymbol{x}} \in X$ is called *equilibrium* or *fixed point*, if $\varphi^t \bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}$ for all $t \in \mathbb{T}$.

(ii) The trajectory of $\boldsymbol{x}_0$ is called *periodic* or *cycle*, if $\varphi^T \boldsymbol{x}_0 = \boldsymbol{x}_0$ for some $T > 0$. The smallest such $T$ is called *period* of the cycle.

(iii) The trajectory of $\boldsymbol{x}(t) = \varphi^t(\boldsymbol{x}_0)$ is called *quasiperiodic*, if it can be written in the form $\boldsymbol{x}(t) = \boldsymbol{G}(t, \ldots, t)$, $t \in \mathbb{T}$ where the mapping $(t_1, \ldots, t_m) \mapsto \boldsymbol{G}(t_1, \ldots, t_m)$ with $\boldsymbol{G} : \mathbb{R}^m \to X$ is continuous and periodic with respect to each argument $t_i$.

Because the definition of a quasiperiodic orbit is rather technical, the difference between quasiperiodic and periodic trajectories is explained below in more detail. Particularly important are isolated periodic or quasiperiodic trajectories, called *limit cycles*. According to the definition of isolated (fixed) points, a quasiperiodic orbit $\gamma(\boldsymbol{x_0})$ is called not isolated, if there exists another quasiperiodic orbit in any $\varepsilon$-neighbourhood of any point $\boldsymbol{x}$ of the orbit $\gamma(\boldsymbol{x_0})$.

As already mentioned, non-invertible discrete-time systems may have different trajectories which meet at some point $T > 0$ in time and coincide for all $t \geq T$. Thus it is interesting to consider the following generalisations. The trajectory of $\boldsymbol{x}_0$ is called *eventually constant*, *eventually periodic* or *eventually quasiperiodic* if there exists some $T > 0$, such that the trajectory of $\varphi^T(\boldsymbol{x}_0)$ is constant, periodic or quasiperiodic respectively. Hence only the asymptotic behaviour is considered and the transient behaviour – described by $\varphi^t(\boldsymbol{x}_0)$ with $t \in [0, T]$ – is neglected.

**Example 2.2** Consider the linear discrete-time system

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}, \quad A = \begin{pmatrix} \cos \omega & -\sin \omega \\ \sin \omega & \cos \omega \end{pmatrix} .$$

Clearly $\bar{\boldsymbol{x}} = 0$ is a fixed point of this system. All other points are periodic if $T = \frac{2\pi}{\omega}$ is rational, i.e. $T = \frac{p}{q}$ for some $p, q \in \mathbb{N}$. Given that $p$ and $q$ are relatively prime, i.e. $\gcd(p, q) = 1$, the period of these trajectories is $T = q$, whereby the trajectory has made $p$ rotations around the origin. If $T$ is irrational the trajectories are quasiperiodic but not periodic, because there exists

$$\boldsymbol{R}(t) = \begin{pmatrix} \cos \omega t & -\sin \omega t \\ \sin \omega t & \cos \omega t \end{pmatrix} \quad \text{such that } \boldsymbol{x}(t) = \boldsymbol{G}(t) := \boldsymbol{R}(t)\boldsymbol{x}_0 .$$

Note that this system has no isolated limit cycles. These require a non-linear system, because each linear system which exhibits a quasiperiodic orbit has infinitely many non-isolated quasiperiodic trajectories.

According to our definition the set of quasiperiodic trajectories comprises the set of periodic trajectories. A quasiperiodic trajectory is periodic if and only if the vector of periods $\boldsymbol{T} = [T_1, \ldots, T_m]^t$ of the mapping $\boldsymbol{G}$ is proportional to a vector of integers $\boldsymbol{n} = \lambda \boldsymbol{T}$, where the factor $\lambda$ has to be rational in case of discrete-time dynamical systems. For a continuous-time system the resulting period is given by $T = \text{lcm}(\boldsymbol{n})/\lambda$, where $\text{lcm}(\boldsymbol{n})$ denotes the least common multiple of the components of the integer vector $\boldsymbol{n}$.

While periodic orbits of continuous-time systems form a closed curve and their period $T$ may be any positive real number, cycles of discrete-time systems form a finite set of $T$ points where $T$ is an integral number. In practice, we commonly observe oscillatory orbits forming a closed curve in discrete-time systems too. They correspond to quasiperiodic orbits, i.e. the existence of a periodic map $\boldsymbol{G} : \mathbb{R} \to X$ with an irrational periodicity. In simulations it is often impossible to distinguish quasiperiodic orbits from periodic orbits with large period $L \gg 100$, because its points become very dense. We reflect these practical considerations by calling both periodic and quasiperiodic trajectories *oscillatory* in the following.

Besides orbits another important class of point sets in state space are *invariant sets*, i.e. sets which do not change under application of the evolution operator. A subset $S \subset X$ is said to be *positive invariant* under the evolution $\varphi^t$ if $\varphi^t S \subseteq S$ for all $t \geq 0$, i.e. positive orbits starting in $S$ reside in $S$. Similarly $S$ is said to be *negative invariant* under $\varphi^t$, if $\varphi^t S \supseteq S$ for all $t \geq 0$. If $\varphi^t$ is invertible this means especially that negative orbits reside in $S$. If $S$ is both positive and negative invariant, then $S$ is said to be *invariant* under $\varphi^t$ and it holds $\varphi^t S \equiv S$.

Note, that an invariant set $S$ consists of orbits and individual orbits form invariant sets. There exist invariant sets of arbitrary complexity and even very simple dynamical systems can have extremely complex invariant sets – for example the horseshoe map possesses an invariant set with fractal boundaries [Hirsch and Smale, 1974].

More regular and analytically tractable invariant sets are $n$-dimensional manifolds. For example a truly quasiperiodic orbit forms an m-dimensional closed manifold if the corresponding map $\boldsymbol{G}$ of
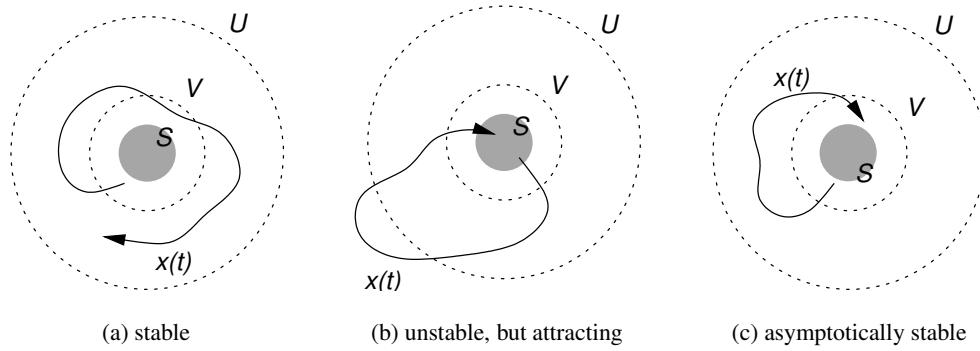
Figure 2.1: Stability of an invariant set $S$, compare with definition 2.3

definition 2.1 requires exactly $m$ arguments. The stability of invariant sets is defined according to Lyapunov:

**Definition 2.3** A positive invariant set $S$ of an autonomous dynamical system, described by $\{\varphi^t\}_{t \in \mathbb{T}}$ is called

(i) *(Lyapunov) stable* , if for any sufficiently small neighbourhood $U \supset S$ there exists a neighbourhood $V \supset S$ such that $\varphi^t \boldsymbol{x} \in U$ for all $\boldsymbol{x} \in V$ and all $t > 0$,

(ii) *attracting*, if there exists a neighbourhood $U \supset S$ such that $\lim_{t \to \infty} \varphi^t \boldsymbol{x} \to S$ for all $\boldsymbol{x} \in U$,

(iii) *globally attracting*, if it attracts all points $\boldsymbol{x}$ within the state space $X$,

(iv) (globally) *asymptotically stable*, if it is stable and (globally) attracting.

While Lyapunov stability ensures that nearby orbits do not leave the neighbourhood of $S$ (fig. 2.1a), the second property ensures that nearby starting orbits converge to $S$ in the asymptotic limit $t \to \infty$ (fig. 2.1b). Both concepts are independent of each other, because there exist invariant sets that are attracting but not stable, since some orbits starting near $S$ eventually approach $S$, but only after an excursion outside any fixed neighbourhood of $S$ [Hahn, 1967].

### 2.1.2 Existence and uniqueness of fixed points

Fixed points are trivial trajectories – they are constant under evolution of the dynamical system. The analytical computation of fixed points according to the equations

$$\boldsymbol{F}(\bar{\boldsymbol{x}}) = 0 \qquad \text{or} \qquad \bar{\boldsymbol{x}} = \boldsymbol{F}(\bar{\boldsymbol{x}}) \tag{2.7}$$

for nonlinear differential equations (2.1) or difference equations (2.2) respectively is in general not possible. Nevertheless there exist fixed point theorems which guarantee the existence and sometimes also the uniqueness of fixed points under certain conditions. In the following we discuss these theorems for discrete-time systems only. Appropriate theorems for continuous-time systems can be found in Stuart and Humphries [1996]. The most important theorem is

**Theorem 2.4 (Brouwer's Fixed Point Theorem)** Suppose $\boldsymbol{F} : B \to \mathbb{R}^n$ is continuous, $B$ is a compact convex subset of $\mathbb{R}^n$, and $\boldsymbol{F}(B) \subseteq B$. Then there exists $\bar{\boldsymbol{x}} \in B$ such that $\boldsymbol{F}(\bar{\boldsymbol{x}}) = \bar{\boldsymbol{x}}$.

A proof of this theorem can be found in Heuser [1994]. Note that the neural network model (2.5) satisfies these conditions and thus possesses at least one fixed point. The uniqueness of a fixed point can be assured by

**Theorem 2.5 (Contraction Mapping Theorem)** Suppose $\boldsymbol{F} : B \to B$ where $B$ is a closed sub-set of a Banach space $X$ and $\boldsymbol{F}$ is a contraction on $B$, i.e. there exists $\mu < 1$ such that

$$\|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\| \leq \mu \|\boldsymbol{x} - \boldsymbol{y}\| \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in B.$$

Then there exists a unique fixed point $\bar{\boldsymbol{x}}$ and each trajectory starting in $B$ converges exponentially fast to it.

### 2.1.3 Limit sets and attractors

The basic objective of dynamical systems theory is to describe the *final* behaviour of some evolving states, i.e. the asymptotic behaviour of the states as $t$ approaches infinity. There might exist attracting fixed points, attracting oscillatory orbits, or even more complex attracting sets. A second important question is, which initial states lead to these final behaviours, i.e. what is the basin of attraction of some attracting set? Obviously all basins of attraction partition the state space into sets with different asymptotic behaviour. If a complete description of all possible attractors and their basins of attraction would be available, the long term behaviour of each initial state could be predicted. If the basins of attraction have smooth boundaries they can be easily drawn into the phase portrait of the dynamical system summarising all asymptotic behaviour. But unfortunately basins of attraction might even have fractal boundaries, which makes it impossible to draw them into a figure. Pasemann [2002] gives an example of a chaotic two-neuron RNN whose attractors have fractal basin boundaries.

A general description of all possible asymptotic states of a dynamical system is captured in the concept of limit sets, which contain limit points of orbits. The $\omega$-limit set $\omega(B)$ of a set $B \subset X$ is defined as

$$\omega(B) := \{\boldsymbol{x} \in X \mid \exists\, t_n \to \infty, \boldsymbol{x}_n \in B : \varphi^{t_n} \boldsymbol{x}_n \to \boldsymbol{x} \text{ for } n \to \infty\}. \tag{2.8}$$

For an invertible dynamical system similarly the $\alpha$-limit set is defined as

$$\alpha(B) := \{\boldsymbol{x} \in X \mid \exists\, t_n \to -\infty, \boldsymbol{x}_n \in B : \varphi^{t_n} \boldsymbol{x}_n \to \boldsymbol{x} \text{ for } n \to \infty\}. \tag{2.9}$$

**Theorem 2.6 (Properties of $\omega$-limit sets)** Let $B \subset X$ be an arbitrary compact set. Then the following holds:

(i) The $\omega$-limit set $\omega(B)$ is given by $\omega(B) = \bigcap_{s \geq 0} \overline{\bigcup_{t \geq s} \varphi^t B}$.

(ii) Since $\omega(B)$ is the intersection of closed sets it follows that $\omega(B)$ is always a closed set in $X$.

If we additionally assume continuity of $\varphi^t$ we have:

(iii) $\omega(B)$ is positive invariant.

(iv) If $\bigcup_{t \geq 0} \varphi^t B$ is compact then $\omega(B)$ is a compact invariant set.

(v) Given a compact positive orbit $\gamma^+(\boldsymbol{x})$ the $\omega$-limit set $\omega(\boldsymbol{x})$ is nonempty. In case of a continuous-time system it is even a connected set. In case of a discrete-time system it is *invariant connected*, i.e. there exist no nonempty, closed invariant and disjoint sets $\omega_1$, $\omega_2 \subset \omega(\boldsymbol{x})$ such that $\omega(\boldsymbol{x}) = \omega_1 \dot\cup \omega_2$.

Because the considered neural network models (2.4) and (2.5) are continuous and bounded, condition (iv) holds and all $\omega$-limit sets of these networks are compact invariant sets. The proofs of the theorem's assertions can be found in Stuart and Humphries [1996] and Sell and You [2002].

Trying to find a precise definition of a basin of attraction in the literature turns out to be difficult, because there exist several perceptions, which properties are important. For our purposes the basin of attraction of $S$ should contain all states $x \in X$, which asymptotically approach $S$, i.e. whose $\omega$-limit set $\omega(x)$ is contained in $S$. In order to yield a reasonable definition, the set $S$ should be invariant, because we are considering some sort of attraction to $S$. Thus trajectories must not leave $S$, although $S$ does not need to be attracting in the sense of definition 2.3, i.e. the basin of attraction of $S$ does not need to contain a neighbourhood of $S$. Hence we define

**Definition 2.7** The *basin of attraction* $\mathcal{B}(S)$ of a compact invariant set $S$ is given by

$$\mathcal{B}(S) := \{ x \in X \,|\, \omega(\{x\}) \subseteq S \} \,.$$

**Definition 2.8** Finally an *attractor* is defined as a compact invariant set $\mathcal{A}$, which attracts a neighbourhood of itself, i.e. there exists a neighbourhood $U$ of $\mathcal{A}$, such that for all $x \in U$ $\mathrm{dist}(\varphi^t(x), \mathcal{A}) \to 0$ for $t \to \infty$ holds. This definition follows Reitmann [1996] and does not demand uniform attraction of $U$ like other authors do [Stuart and Humphries, 1996; Sell and You, 2002][3].

The attractor is said to be *minimal* if it contains a dense positive orbit, i.e. $\overline{\gamma}^+(x) = \mathcal{A}$ for some $x \in \mathcal{A}$, which ensures that the attractor cannot be decomposed into two disjoint attractors. The attractor is called *global* if it attracts every bounded set $B \subset X$.

**Theorem 2.9** Let $\varphi^t : \mathbb{R}^n \to \mathbb{R}^n$ be continuous. Assume that $B \subset X$ is a bounded absorbing set, i.e. $\varphi^t \overline{B} \subset B$ holds for all $t \geq 0$. Then $\omega(B)$ is an attractor which attracts $B$. Furthermore

$$\omega(B) = \bigcap_{t \geq 0} \varphi^t B \,.$$

**Remark:** Again boundedness of the RNNs (2.4) and (2.5) ensures the existence of an attractor. Particularly $\omega(X)$ is an attractor, which is not necessarily minimal and often it can be decomposed into smaller attractors.

## 2.2 Topological Equivalence of Dynamical Systems

Studying dynamical systems, we are not only interested in specific solutions of a specific system, but we want to classify dynamical systems according to their general qualitative behaviour, i.e. the number, position and stability of their invariant sets. This aspect becomes especially important in the context of recurrent neural networks, because we want to know which class of networks is applicable for a specific task, e.g. as an associative memory or central pattern generator. To this end we want to consider two dynamical systems as (locally) equivalent if their (local) phase portraits are similar in a qualitative sense, i.e. if they can be transformed into each other through a continuous transformation:

**Definition 2.10** Two dynamical systems $\varphi^t : X_1 \to X_1$ and $\psi^t : X_2 \to X_2$, $t \in \mathbb{T}$, are called *topologically equivalent* if there exists a homeomorphism $h : X_1 \to X_2$ mapping orbits of the first system onto orbits of the second system, preserving the direction of time. If the parameterisation

---

[3] Uniform attraction guarantees structural stability of the attractor, which we do not need within this work.

of time is preserved by $h$ as well, i.e. the trajectories evolve with the same speed, the systems are called *topologically conjugate*.

They are called *locally topologically equivalent* with respect to open neighbourhoods $U \subset X_1$ and $V \subset X_2$, if a homeomorphism $h$ maps orbits from $U$ onto orbits in $V$, preserving the direction of time.

The systems are called $C^k$-*equivalent* if the homeomorphism $h$ is a $C^k$-diffeomorphism. In this case $h$ is sometimes called *coordinate transformation*.

While this definition holds for both discrete- and continuous-time systems, for discrete-time systems this condition means that the corresponding maps $\varphi^1$ and $\psi^1$ of equivalent systems satisfy the relation

$$\varphi^1 = h^{-1} \circ \psi^1 \circ h \,, \tag{2.10}$$

which preserves time parametrisation automatically, i.e. the two maps are topologically conjugate. Topological equivalence and conjugacy are equivalence relations, i.e. they are reflexive, symmetric and transitive. Equation (2.10) is equivalent to the commutative diagram

$$
\begin{array}{ccc}
X_1 & \xrightarrow{\;\varphi^1\;} & X_1 \\
h \downarrow & & \downarrow h \\
X_2 & \xrightarrow{\;\psi^1\;} & X_2
\end{array}
$$

Given two maps $\varphi$ and $\psi$ which are topologically conjugate with respect to the homeomorphism $h$. Then it immediately follows from the definition, that both systems have the same number of fixed points, periodic and quasiperiodic orbits, which are mapped onto each other by application of the homeomorphism $h$ or its inverse $h^{-1}$.

**Example 2.11** An important example are the two different versions of the neural network model (2.5) considered in literature:

$$\boldsymbol{x}(t+1) = W\boldsymbol{\sigma}(\boldsymbol{x}(t)) + \boldsymbol{u} \tag{2.11}$$

$$\boldsymbol{y}(t+1) = \boldsymbol{\sigma}\left(W\boldsymbol{y}(t) + \boldsymbol{u}\right) \,. \tag{2.12}$$

Clearly we have $\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{\sigma}(W\boldsymbol{x} + \boldsymbol{u})$ and $\boldsymbol{x} = h^{-1}(\boldsymbol{y}) = W^{-1}\left(\boldsymbol{\sigma}^{-1}(\boldsymbol{y}) - \boldsymbol{u}\right)$, where $h$ is an homeomorphism if and only if $W$ is nonsingular and $\boldsymbol{\sigma}$ is an homeomorphism. Thus both systems are not conjugate in general. Nevertheless they have identical dynamical behaviour as was pointed out by Feng and Hadeler [1996], because their trajectories map through the following continuous transformations into each other:

$$
\begin{aligned}
\boldsymbol{x}(t+1) &= W\boldsymbol{y}(t) + \boldsymbol{u} \\
\boldsymbol{y}(t+1) &= \boldsymbol{\sigma}(\boldsymbol{x}(t+1))
\end{aligned}
\tag{2.13}
$$

As discussed in Bhaya et al. [1996] the systems (2.11) and (2.12) can be viewed as two possible realizations of the combined system (2.13) using either $\boldsymbol{x}$ or $\boldsymbol{y}$ as the variable of interest.

### 2.2.1 Local phase portrait at a fixed point

In the following, we classify the local phase portrait of a discrete-time system near a fixed point[4]. For this purpose we drop the time-dependency of the evolution operator $\varphi^t$ and just consider

---

[4] Similar theorems hold for continuous-time systems as well, whereas the Jacobian eigenvalues of continuous-time and discrete-time systems are related by the exponential function, which maps the left negative half plane to the unit circle in the complex plane.

the time-one-map $\varphi \equiv \varphi^1$. The local phase portrait at a fixed point $\bar{\boldsymbol{x}}$ depends crucially on the eigenvalues of the Jacobian matrix evaluated at $\bar{\boldsymbol{x}}$:

$$J(\bar{\boldsymbol{x}}) = D_{\boldsymbol{x}}\varphi(\bar{\boldsymbol{x}}) = \left(\frac{\partial \varphi_i}{\partial x_j}\right)_{i,j}(\bar{\boldsymbol{x}}). \tag{2.14}$$

The possibly complex eigenvalues $\lambda_1, \ldots, \lambda_n$ of $J(\bar{\boldsymbol{x}})$ are sometimes called multipliers of the fixed point $\bar{\boldsymbol{x}}$. Let $n_-$, $n_0$ and $n_+$ denote the numbers of multipliers of $\bar{\boldsymbol{x}}$ lying inside, on, and outside the unit circle $\{\lambda \in \mathbb{C} \,|\, |\lambda| = 1\}$ respectively. If the Jacobian $J(\bar{\boldsymbol{x}})$ has no eigenvalues on the unit circle, i.e. if $n_0 = 0$, $\bar{\boldsymbol{x}}$ is said to be *hyperbolic*. A hyperbolic fixed point is called *sink* or *stable node* if all multipliers are inside the unit circle. It is called *source* or *unstable node*, if all multipliers are outside the unit circle and it is called *saddle* if there exists multipliers both inside and outside the unit circle.

According to the following theorem these definitions are sensible, because the local phase portrait near a hyperbolic fixed point of an arbitrary nonlinear dynamical system is equivalent to the phase portrait of the linear system, given by the Jacobian $J(\bar{\boldsymbol{x}})$. For a proof of this important theorem we refer to Guckenheimer and Holmes [1993].

**Theorem 2.12 (Grobman-Hartman)** Given a discrete-time dynamical system $\boldsymbol{x} \mapsto \varphi(\boldsymbol{x})$ with $\varphi \in C^1$. If $\bar{\boldsymbol{x}}$ is a hyperbolic fixed point of $\varphi$ then the system is – in a neighbourhood of $\bar{\boldsymbol{x}}$ – locally topologically conjugate to its linearisation

$$\boldsymbol{x} \mapsto \bar{\boldsymbol{x}} + J(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}).$$

**Remark:** Since a generic matrix has no eigenvalues on the unit circle, hyperbolicity is a typical property of a generic fixed point of an arbitrary dynamical system.

### 2.2.2 Phase portraits of linear maps

According to the Grobman-Hartman theorem it is important to study different classes of phase portraits of linear maps. Given the linear discrete-time dynamical system

$$\boldsymbol{x} \mapsto A\boldsymbol{x} \quad \boldsymbol{x} \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n} \tag{2.15}$$

the trajectory starting at $\boldsymbol{x}_0$ is given by $\boldsymbol{x}(t) = A^t \boldsymbol{x}_0$. Employing the Jordan normal form[5] $J = P^{-1}AP$ of the matrix $A$, the matrix power can be easily computed as $A^t = PJ^tP^{-1}$ where the power of the Jordan matrix $J$ decomposes into the powers of its Jordan blocks $J_i = \lambda\mathbf{1} + N$ which have the form

$$J_r^t(\lambda) = (\lambda\mathbf{1} + N)^t = \lambda^t\mathbf{1} + \sum_{j=1}^{t}\binom{t}{j}\lambda^{t-j}N^j = \lambda^t\mathbf{1} + \sum_{j=1}^{r-1}\binom{t}{j}\lambda^{t-j}N^j. \tag{2.16}$$

Here the binomial expansion holds because $\lambda\mathbf{1}$ and $N$ commute. Because the $t$-th power of the nilpotent matrix $N$ equals to zero for $t \geq r$ the sum can be truncated for $j \geq r$.

As can be seen from (2.16) all trajectories converge to the fixed point $\bar{\boldsymbol{x}} = 0$, i.e. the origin is globally asymptotically stable, if and only if all eigenvalues $\lambda_i$ have modulus less than one[6] – or equivalently lie strictly inside the unit circle in the complex plane. For this reason a matrix

---

[5] For a short introduction to Jordan normal forms and the employed notation we refer to appendix A.1.
[6] This conclusion holds because $\lambda^{t-j}$ decays faster than $\binom{t}{j} \in \mathcal{O}(t^j)$ grows as $t$ tends to infinity.

$A \in \mathbb{R}^{n \times n}$ is called *Schur stable* if all its eigenvalues are strictly less than unity in absolute value, i.e. if its spectral radius $\rho(A) := \max|\lambda_i|$ is less than one.

The origin is stable but not attracting if and only if all eigenvalues satisfy $|\lambda_i| \leq 1$ and there exist eigenvalues $\lambda_k$ with modulus one, such that all their corresponding Jordan blocks $J_{r_k}(\lambda_k)$ have size $r_k = 1$. In all other cases, i.e. if eigenvalues lie outside the unit circle or if Jordan blocks of size $r_k > 1$ with eigenvalues $|\lambda_k| = 1$ exist, the origin is unstable.

If there exist simple eigenvalues $\lambda_i = 1$ a whole continuum of fixed points exists spanned by the eigenspace corresponding to $\lambda_i$.

In order to study the evolution of a vector $\boldsymbol{x}_0$ further we expand it in terms of the basis of generalised eigenvectors. Let $\boldsymbol{p}_i$ be the $i$-th generalised eigenvector, i.e. the $i$-th column of the matrix $P$. Hence, its $j$-th component will be denoted with $p_{ji}$. Then we can expand $\boldsymbol{x}_0$ as $\sum a_i \boldsymbol{p_i}$ with coefficients $\boldsymbol{a} = P^{-1}\boldsymbol{x}_0$. Because $A$ is a real matrix, its complex eigenvalues $\lambda_i$, their generalised eigenvectors $\boldsymbol{p}_i$ and hence also the corresponding coefficients $a_i$ occur as complex conjugate pairs always. In the following we distinguish between real and complex eigenvalues and we write complex numbers in exponential representation, i.e.

$$\lambda_i = |\lambda_i|e^{\omega_i}$$
$$p_{ji} = |p_{ji}|e^{\theta_{ji}}$$
$$a_i = |a_i|e^{\phi_i}.$$

Using the relation $N\boldsymbol{e}_i = \boldsymbol{e}_{i-1}$ for canonical basis vectors $\boldsymbol{e}_i$ we get now

$$
\begin{aligned}
(A^t \boldsymbol{x}_0)_j = PJ^tP^{-1}P\boldsymbol{a} &= PJ^t\boldsymbol{a} \\
&= \sum_{\mathrm{Im}\,\lambda_i=0} \lambda_i^t a_i p_{ji} + 2\sum_{\mathrm{Im}\,\lambda_i>0} |\lambda_i|^t|a_i||p_{j\,i}|\cos(\omega_i t + \phi_i + \theta_{ji}) \\
&+ \sum_{\mathrm{Im}\,\lambda_i=0}\sum_k^{r_i-1}\binom{t}{k}\lambda_i^{t-k}p_{j\,i-k} + 2\sum_{\mathrm{Im}\,\lambda_i>0}\sum_k^{r_i-1}\binom{t}{k}|\lambda_i|^{t-k}|a_i||p_{j\,i-k}|\cos(\omega_{i-k}t + \phi_i + \theta_{j\,i-k}).
\end{aligned}
$$

$$(2.17)$$

Within this equation we separated on the one hand contributions from Jordan blocks of size $r_i = 1$ (first row) and size $r_i > 1$ (second row) and on the other hand contributions from real and complex eigenvalues. The sums over $k$, which originate from Jordan blocks of size $r_i > 1$, introduce contributions of generalised eigenvectors $\boldsymbol{p}_{i-k}$, even if the corresponding initial components $a_{i-k}$ are zero. Further, the rapidly growing terms $\binom{t}{k}$ cause large amplitude transients of the trajectories before they converge towards the origin due to small eigenvalues (see fig. 2.2).

## Embedding Problem

Given a discrete-time dynamical system $\{\varphi^t\}_{t\in\mathbb{T}_d}$ the question arises whether a continuous-time system $\{\psi^t\}_{t\in\mathbb{T}_c}$ exists such that $\psi^t = \varphi^t$ for all $t \in \mathbb{T}_d$, i.e. the continuous-time system reproduces the discrete one if evaluated at the corresponding discrete times. If this is possible, the discrete-time system is said to be embedded in the continuous-time system. While for a general nonlinear system this embedding problem is difficult to solve (see Utz [1981] for a survey), there exists a necessary and sufficient condition for linear systems [Reitmann, 1996]:

**Theorem 2.13** Let be given a discrete-time system $\{A^t\}_{t\in\mathbb{Z}}$ induced by the linear mapping $A \in \mathbb{R}^{n\times n}$ and a continuous-time system $\{e^{tB}\}_{t\in\mathbb{R}}$ induced by the linear differential equations $\dot{\boldsymbol{x}} = B\boldsymbol{x}$. Then $\{A^t\}$ can be embedded into $\{e^{tB}\}$ if and only if $A^t = e^{tB}$ for all $t \in \mathbb{Z}$ which is

(a) discrete-time system                    (b) embedded continuous-time system
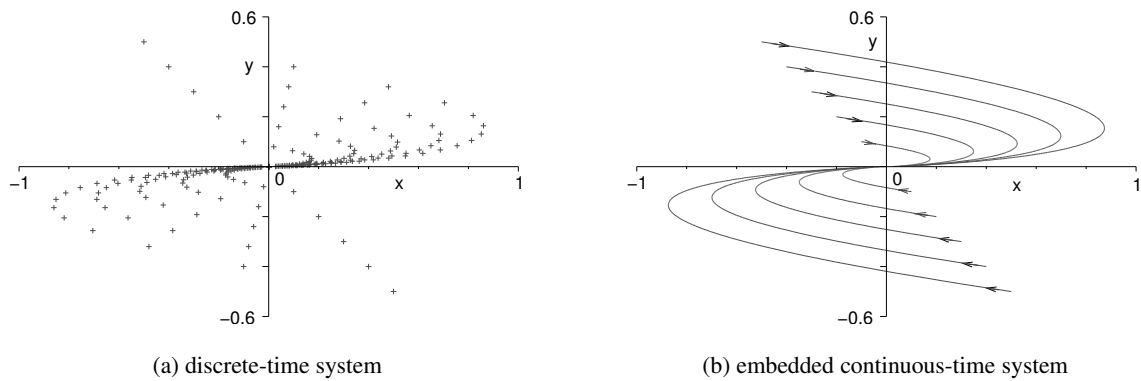
Figure 2.2: Phase Portrait of the linear system $\boldsymbol{x}(t) = J^t \boldsymbol{x}(0)$ with $J = \begin{pmatrix} 0.8 & 1 \\ 0 & 0.8 \end{pmatrix}$.

equivalent to $A = \mathrm{e}^B$. The solution $B$ of this equation for a given matrix $A$, called matrix logarithm, exists and is real if and only if [Horn and Johnson, 1991]

(i) $A$ is nonsingular and

(ii) Jordan blocks corresponding to negative real eigenvalues occur pairwise for every size of these blocks.

Real negative eigenvalues of the matrix $A$ of a discrete-time dynamical system lead to jumps of the trajectory due to sign changes of the corresponding eigenvector components within each time step. A zero eigenvalue sets the corresponding eigenvector component to zero within a single time step. Both types of discontinuous behaviour clearly cannot be observed in continuous-time systems. An exception form negative eigenvalues if their corresponding Jordan blocks occur pairwise. In these cases the even number of corresponding reflections sum up to a rotation about the angle $\pi$ and we can find an embedding continuous-time system, for example:

$$\ln \begin{pmatrix} -\lambda & \\ & -\lambda \end{pmatrix} = \begin{pmatrix} \ln \lambda & \pi \\ -\pi & \ln \lambda \end{pmatrix} \qquad \text{for all } \lambda > 0 \tag{2.18}$$

The Grobman-Hartman theorem allows us to approximate the phase portrait of any nonlinear discrete-time dynamical system near a hyperbolic fixed point by the phase portrait of the linearised system. If this can be embedded into a continuous-time system, we can draw its trajectories as closed curves, which visualises the time evolution more clearly than the single points forming the discrete trajectory (compare figures 2.2a and 2.2b).

Of course, each concrete trajectory of the discrete-time system starting at a point on the orbit of the embedding continuous-time system moves along this orbit in a jumping manner. This orbit is an invariant curve of the continuous-time system trivially – but it is invariant with respect to the discrete-time system as well. Actually it represents a whole bunch of orbits of the discrete-time system. If the Grobman-Hartman theorem is applicable, i.e. in the neighbourhood of a hyperbolic fixed point, such invariant curves exist in nonlinear systems as well.

**Classification of hyperbolic fixed points in the plane**

Using the Grobman-Hartman theorem we now classify the phase portraits of hyperbolic fixed points $\bar{\boldsymbol{x}}$ using the eigenvalues of the Jacobian $J(\bar{\boldsymbol{x}})$. Table 2.1 displays well-known phase portraits
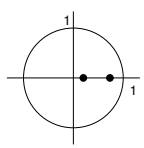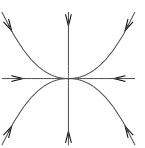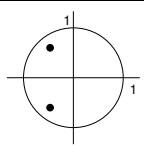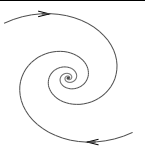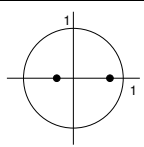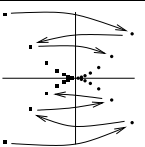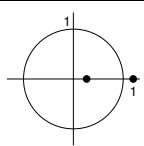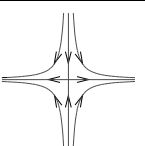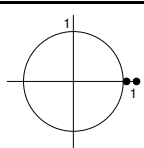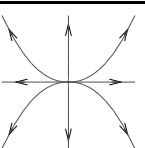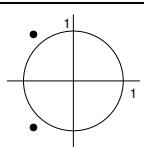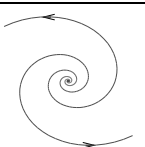
| stability | eigenvalues | phase portrait | $J$ | $\ln J$ |
|---|---|---|---|---|
|  |  | node | $\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$ $0 < \lambda_{1,2} < 1$ | $\begin{pmatrix} \ln \lambda_1 & \\ & \ln \lambda_2 \end{pmatrix}$ |
| stable |  | focus | $\lambda \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}$ $0 < \lambda < 1, \phi \in [0, \pi]$ | $\begin{pmatrix} \ln \lambda & \phi \\ -\phi & \ln \lambda \end{pmatrix}$ |
|  |  | flip node | $\begin{pmatrix} -\lambda_1 & \\ & \lambda_2 \end{pmatrix}$ $0 < \lambda_{1,2} < 1$ |  |
| unstable |  | saddle | $\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$ $0 < \lambda_1 < 1 < \lambda_2$ | $\begin{pmatrix} \ln \lambda_1 & \\ & \ln \lambda_2 \end{pmatrix}$ |
|  |  | node | $\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$ $\lambda_{1,2} > 1$ | $\begin{pmatrix} \ln \lambda_1 & \\ & \ln \lambda_2 \end{pmatrix}$ |
| unstable |  | focus | $\lambda \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix}$ $\lambda > 1, \phi \in [0, \pi]$ | $\begin{pmatrix} \ln \lambda & \phi \\ -\phi & \ln \lambda \end{pmatrix}$ |

Table 2.1:  Topological classification of hyperbolic fixed points in the plane according to the eigenvalues of the Jacobian $J$. The third row gives an example of the orbit, where the number of negative eigenvalues is odd. In all other cases the linear systems are embedabble into continuous-time systems and invariant curves are drawn instead of single orbits. In principle all real eigenvalues $\lambda_{1,2}$ can assume negative values as well, producing the corresponding flipping trajectory, but these cases are omitted here for the sake of clarity.

of linear maps. Because higher dimensional linear maps usually can be split into several lower dimensional sub-maps corresponding to Jordan blocks of the map, we restrict our analysis to the plane here. We can distinguish mainly three types of hyperbolic fixed points: stable nodes/foci, saddles and unstable nodes/foci, characterised by the numbers $n_-$ and $n_+$ of eigenvalues inside and outside the unit circle. Complex eigenvalues corresponding to foci produce spiral orbits. It can be shown that these foci are locally topologically equivalent to non-spiral nodes [Kuznetsov, 1995].

**Example 2.14** Consider the two linear systems $\psi(x) = a \cdot x$ and $\varphi(x) = b \cdot x$. They are topologically equivalent if and only if $a$ and $b$ have the same sign or are zero both. As can be easily verified, a mediating homeomorphism is given by $h(x) = \text{sgn}(x) \cdot |x|^\tau$ with $\tau = \frac{\ln|a|}{\ln|b|}$. The special case $\ln|b| = 0$ can be handled separately.

As we have seen above, flip nodes corresponding to negative real eigenvalues show a jumping trajectory, which is locally topologically equivalent to a spiral trajectory if the eigenvalues occur pairwise. If the number of negative eigenvalues is odd, there remains a reflection and the flip node

is not locally topologically conjugate to a normal node anymore. Using these results we can state the following theorem [Kuznetsov, 1995]:

**Theorem 2.15** The phase portraits of two discrete-time dynamical systems $\{\varphi^t : \mathbb{R}^n \to \mathbb{R}^n\}_{t \in \mathbb{Z}}$ and $\{\psi^t : \mathbb{R}^n \to \mathbb{R}^n\}_{t \in \mathbb{Z}}$ near two hyperbolic fixed points $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{y}}$ are locally topologically conjugate if and only if

  (i) the Jacobian matrices $J(\bar{\boldsymbol{x}})$ and $J(\bar{\boldsymbol{y}})$ have the same number $n_-$ and $n_+$ of eigenvalues with $|\lambda| < 1$ and $|\lambda| > 1$ respectively,

 (ii) they have the same number of zero eigenvalues and

(iii) the signs of the products of real negative eigenvalues inside and outside the unit circle are identical for both fixed points.

The proof is based on the Grobman-Hartman theorem, which ensures that both systems are locally topologically equivalent to their linearisations, which in turn are topologically equivalent due to the conditions on the eigenvalues. Conditions (ii,iii) are specific for discrete-time systems and account for the problem of jumping trajectories which cannot be homeomorph to non-jumping ones if the number of negative real eigenvalues is odd. A similar result holds for continuous-time systems as well, where conditions (ii,iii) can be dropped.

In order to state another important property of hyperbolic fixed points, we introduce the following distance measure on the space $C^k(U)$ of $C^k$ functions $\varphi : U \to U$, where $U$ is an open subset of $\mathbb{R}^n$:

$$d(\varphi, \psi) := \sup_{\boldsymbol{x} \in U} \|\varphi(\boldsymbol{x}) - \psi(\boldsymbol{x})\| + \sup_{\boldsymbol{x} \in U} \|D\varphi(\boldsymbol{x}) - D\psi(\boldsymbol{x})\| . \tag{2.19}$$

Here $\|\cdot\|$ means a vector norm in the first summand and its induced matrix norm in the second summand. It can be shown that the resulting metric space $(C^k(U), d)$ is complete [Reitmann, 1996]. Now we can state the following theorem:

**Theorem 2.16** The set of discrete-time dynamical systems which have a hyperbolic fixed point is open within $(C^k(U), d)$.

**Proof**    Let $\varphi \in C^k(U)$ define a discrete-time dynamical system with a hyperbolic fixed point $\bar{\boldsymbol{x}}$. Hyperbolicity implies that the matrix $\mathbf{1} - J(\bar{\boldsymbol{x}})$ is nonsingular, which guarantees the existence of a fixed point $\bar{\boldsymbol{x}}_\varepsilon$ for any perturbed system $\varphi_\varepsilon$ within some $\varepsilon$-neighbourhood $U_\varepsilon(\varphi)$ of $\varphi$ according to the implicit function theorem. Because this perturbed fixed point depends continuously on $\varphi$, its Jacobian $J(\bar{\boldsymbol{x}}_\varepsilon)$ changes continuously as well.

Using the continuous dependence of eigenvalues on matrix entries [Kato, 1982] we can conclude that the perturbed fixed point is hyperbolic as well, having the same number $n_-$ and $n_+$ of eigenvalues inside resp. outside the unit circle. Thus its stability properties do not change. $\square$

Using theorem 2.15 and continuous dependence of eigenvalues we can further conclude that a locally hyperbolic discrete-time dynamical system is locally *structurally stable*, i.e. any $\varepsilon$-perturbation of $\varphi$ within $U_\varepsilon(\varphi)$ is locally topologically conjugate to $\varphi$. Besides hyperbolicity we have to ensure that Jacobian eigenvalues do not change in a manner that conditions (ii,iii) of theorem 2.15 are violated.

### 2.2.3 Center manifold theory

The dynamics near a nonhyperbolic fixed point is usually very complicated. Nevertheless it is possible to simplify the dynamical system in a neighbourhood of the fixed point reducing its dimension, because the dynamics along hyperbolic directions is almost linear and can be separated. This process is called center manifold reduction.

We consider a discrete-time system, given by a diffeomorphism $\varphi : \mathbb{R}^n \to \mathbb{R}^n$. For an arbitrary fixed point $\bar{x}$ we introduce the stable and unstable invariant sets

$$W^s(\bar{x}) = \{ x \mid \lim_{t \to +\infty} \varphi^t(x) = \bar{x} \}$$
$$W^u(\bar{x}) = \{ x \mid \lim_{t \to -\infty} \varphi^t(x) = \bar{x} \} \,,$$

containing points converging to or diverging from the fixed point $\bar{x}$ under forward iteration of $\varphi$. The following theorem can be traced back to Hadamard [1901] and Perron [1930].

**Theorem 2.17** Let $\varphi$ be a $C^r$-diffeomorphism on $\mathbb{R}^n$ with a hyperbolic fixed point, namely $n_0 = 0, n_- + n_+ = n$.

  (i) The stable and unstable invariant sets $W^s(\bar{x}), W^u(\bar{x})$ are immersed[7] $C^r$-manifolds of dimension $n_-$ and $n_+$ respectively.

  (ii) The intersections of $W^s(\bar{x})$ and $W^u(\bar{x})$ with a sufficiently small neighbourhood of $\bar{x}$ are smooth $C^r$-submanifolds $W^s_{\mathrm{loc}}(\bar{x}), W^u_{\mathrm{loc}}(\bar{x})$, called local stable and unstable manifold at $\bar{x}$. Further at $\bar{x}$ these manifolds are tangent to $E^s$ and $E^u$ respectively, which are the generalised eigenspaces corresponding to eigenvalues of $J(\bar{x})$ inside resp. outside the unit circle.

**Remark:** The global structure of the stable and unstable manifolds can be very complex – only in a small neighbourhood of the fixed point they are smooth submanifolds. As an example consider two hyperbolic fixed points $\bar{x}$ and $\bar{y}$, whose unstable resp. stable invariant manifolds $W^u(\bar{x})$ and $W^s(\bar{y})$ intersect each other at some point $z \notin \{\bar{x}, \bar{y}\}$. Then by definition the whole orbit starting at $z$ belongs to both $W^u(\bar{x})$ and $W^s(\bar{y})$ and is called *homoclinic* if $\bar{x} = \bar{y}$ or *heteroclinic* if $\bar{x} \neq \bar{y}$.

While the positive orbit $\gamma^+(z)$ accumulates at the fixed point $\bar{y}$, the negative orbit $\gamma^-(z)$ accumulates at $\bar{x}$. The infinite number of associated intersections of the stable and unstable manifolds causes them to oscillate in a very complex manner in the neighbourhood of the fixed points $\bar{x}$ resp. $\bar{y}$. The resulting complex structure is sketched in figure 2.3 in case of a homoclinic orbit.

**Theorem 2.18 (Center Manifold Theorem)** If the fixed point $\bar{x}$ is nonhyperbolic, i.e. $n_0 > 0$, there exists a locally defined smooth $n_0$-dimensional invariant manifold $W^c_{\mathrm{loc}}(\bar{x})$, called *center manifold*. Again at $\bar{x}$ it is tangent to the generalised eigenspace $E^c$ corresponding to eigenvalues of $J(\bar{x})$ *on* the unit circle.

The center manifold has the same (finite) smoothness as $\varphi$ in some neighbourhood of $\bar{x}$. Moreover there exists a neighbourhood $U$ of $\bar{x}$, such that if $\varphi^t(x) \in U$ for all $t \in \mathbb{Z}_+$ ($\mathbb{Z}_-$), then $\varphi^t(x) \to W^c_{\mathrm{loc}}(\bar{x})$ for $t \to +\infty$ ($t \to -\infty$).

The last statement especially means, that the center manifold is attracting or repelling if $n_+ = 0$ resp. $n_- = 0$. Finally we rewrite the dynamical system $x \mapsto \varphi(x)$, separating the nonhyperbolic

---

[7]  To be an immersed manifold is a weaker property than to be a smooth submanifold of $\mathbb{R}^n$. Both terms originate from the theory of differentiable manifolds and we refer to Lang [1967] for a detailed introduction to this topic.

$u_1$
$u_2$
$\psi_1$
$\psi_2$

$W^u(\bar{\boldsymbol{x}})$

$\varphi(\boldsymbol{z})$
$\varphi^2(\boldsymbol{z})$
$\varphi^{-1}(\boldsymbol{z})$

$\bar{\boldsymbol{x}}$

$W^s(\bar{\boldsymbol{x}})$

$u_1$
$u_2$
$\psi_1$
$\psi_2$

$W^s(\bar{\boldsymbol{x}})$
$W^u_{\boldsymbol{z}}(\bar{\boldsymbol{x}})$
$W^s$ $\bar{\boldsymbol{x}}$
$W^u$

$\varphi^{-1}(\boldsymbol{z})$
$\boldsymbol{z}$
$\varphi^2(\boldsymbol{z})$
$\varphi(\boldsymbol{z})$

(a) single intersection           (b) infinite intersections
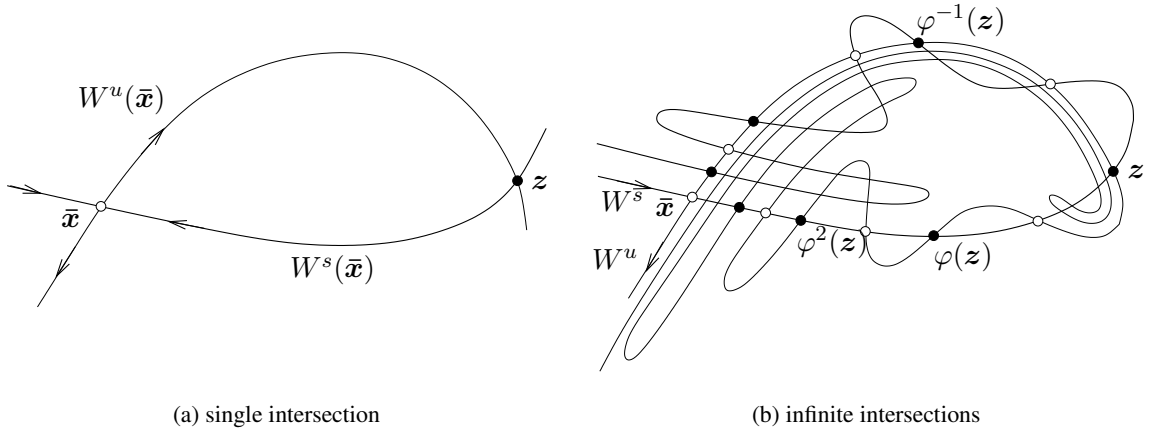
Figure 2.3: Poincaré homoclinic structure resulting from a homoclinic orbit.

($\boldsymbol{u} \in \mathbb{R}^{n_0}$) and hyperbolic (converging ($\boldsymbol{v} \in \mathbb{R}^{n_-}$) and diverging ($\boldsymbol{w} \in \mathbb{R}^{n_+}$)) parts. The center manifold can be locally represented as the graph of a smooth function $H(\boldsymbol{u})$:

$$W^c_{\text{loc}}(\bar{\boldsymbol{x}}) = \bar{\boldsymbol{x}} + \{(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \,|\, (\boldsymbol{v}, \boldsymbol{w}) = H(\boldsymbol{u})\}$$

and we obtain the center manifold reduction:

**Theorem 2.19** The discrete-time dynamical system $\boldsymbol{x} \mapsto \varphi(\boldsymbol{x})$ is locally topologically equivalent to the system

$$\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{w} \end{pmatrix} \mapsto \begin{pmatrix} A^c \boldsymbol{u} \quad + \quad g(\boldsymbol{u}, H(\boldsymbol{u})) \\ A^s \boldsymbol{v} \\ A^u \boldsymbol{w} \end{pmatrix}$$

at the origin. The function $g$ has a Taylor expansion starting with quadratic terms at least. $A^c$, $A^s$ and $A^u$ are matrices in $\mathbb{R}^{n_0 \times n_0}$, $\mathbb{R}^{n_- \times n_-}$, and $\mathbb{R}^{n_+ \times n_+}$ respectively having the eigenvalues of $J_\varphi(\bar{\boldsymbol{x}})$ lying on, inside and outside the unit circle respectively.

From this form, where the hyperbolic $(\boldsymbol{v}, \boldsymbol{w})$-dynamics is separated from the $\boldsymbol{u}$-dynamics on the center manifold, it is easily seen that the interesting nonlinear dynamics happens on the center manifold. But, the center manifold represented in form of the smooth function $H(\boldsymbol{u})$ can be computed approximately only. A recursive procedure allows to compute the coefficients of the Taylor expansion of $H(\boldsymbol{u})$ with increasing order [Kuznetsov, 1995].

### 2.2.4 Hyperbolic periodic orbits

All definitions and theorems concerning hyperbolic fixed points apply equally to periodic orbits, if we keep in mind that all periodic points $\bar{\boldsymbol{x}}_i, i = 1 \ldots T$, of a $T$-periodic orbit are fixed points of the $T$-th iterate $\varphi^T$. Thus we have to determine the eigenvalues of the Jacobians

$$J_{\varphi^T}(\bar{\boldsymbol{x}}_i) = \left(D_{\boldsymbol{x}} \varphi^T\right)(\bar{\boldsymbol{x}}_i) = \prod_{n=1}^{T} \left(D_{\boldsymbol{x}} \varphi^1\right)(\bar{\boldsymbol{x}}_{i-n}) \qquad i = 1 \ldots T \tag{2.20}$$

Because cyclic permutations of the matrices within the product do not change the eigenvalues of the resulting matrix, they do not depend on the selection of the periodic point $\bar{\boldsymbol{x}}_i$.

### 2.2.5 More complex orbits

For even more complex orbits, i.e. quasiperiodic or chaotic orbits, the introduced theory can not be applied. Instead of evaluating a Jacobian with its eigenvalues at a specific point in state space, we have to consider the whole trajectory $\boldsymbol{x}(t) = \varphi^t \boldsymbol{x}_0$ now. For an arbitrary orbit starting at $\boldsymbol{x}_0$ we define the variational equation

$$\boldsymbol{y}(t+1) = \left[ \left( D_{\boldsymbol{x}} \varphi \right) \boldsymbol{x}(t) \right] \boldsymbol{y}(t) \tag{2.21}$$

which is a time-dependent linear dynamical system, which is a linear approximation of the original nonlinear system. Let be given a trajectory $\boldsymbol{y}(t, \boldsymbol{x}_0, \boldsymbol{v})$ of (2.21) starting at $\boldsymbol{v}$. Then the *characteristic exponents* of the semiorbit $\{ \varphi^t \boldsymbol{x}_0 \}_{t \geq 0}$ into *direction $\boldsymbol{v}$* are defined as

$$\chi(\boldsymbol{x}_0, \boldsymbol{v}) = \limsup_{t \to \infty} \frac{1}{t} \ln \| \boldsymbol{y}(t, \boldsymbol{x}_0, \boldsymbol{v}) \| . \tag{2.22}$$

These characteristic exponents describe the evolution of two adjacent trajectories $\varphi^t(\boldsymbol{x}_0)$ and $\varphi^t(\boldsymbol{x}_0 + \varepsilon \boldsymbol{v})$. Because we have the linear approximation

$$\| \varphi^t(\boldsymbol{x}_0 + \varepsilon \boldsymbol{v}) - \varphi^t(\boldsymbol{x}_0) \| \approx \varepsilon \| \left[ \left( D_{\boldsymbol{x}} \varphi \right) \boldsymbol{x}(t) \right] \boldsymbol{v} \| \approx \varepsilon \mathrm{e}^{\chi(\boldsymbol{x}_0, \boldsymbol{v}) t}$$

the orbits approach each other or diverge if $t$ tends to infinity, depending on the sign of the characteristic exponent. Thus the orbit is attracting, if the characteristic exponents are negative for all directions $\boldsymbol{v}$. Having a fixed point or periodic orbit of period $T$, its characteristic exponents equal the logarithms of the eigenvalues of the Jacobian $J_{\varphi^T}(\bar{\boldsymbol{x}})$ in equation 2.20.

Under certain conditions the characteristic exponents are independent of the considered orbit and thus quantify the average rate of convergence or divergence of nearby trajectories in a global sense. In this case they are called *Lyapunov exponents*. Computation of characteristic exponents or Lyapunov exponents for a given dynamical system or a given time series is a field of research of its own and will not be considered here. For more details we refer to Reitmann [1996] and Rangarajan et al. [1998].

# 3 Bifurcations in Dynamical Systems

In the following we consider parameter-dependent dynamical systems

$$\boldsymbol{x} \mapsto \varphi(\boldsymbol{x}, \boldsymbol{\alpha}) \qquad \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\alpha} \in \mathbb{R}^m \tag{3.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ represent state variables and parameters respectively. RNNs are typical examples of parameterised dynamical systems, since the network weights are subject to learning and the network inputs vary in order to reflect environmental changes.

Typically the phase portrait of a system varies as its parameters vary. Again we are especially interested in *qualitative* changes of the phase portrait under parameter variation. If the dynamical behaviour changes dramatically at some parameter value, i.e. a topologically nonequivalent phase portrait appears, we call this a *bifurcation* and the corresponding critical parameter value $\boldsymbol{\alpha}_c$ *bifurcation point*.

The parameter space can be split into regions corresponding to topologically equivalent dynamical systems. If we visualise this partition of the parameter space together with representative phase portraits within each region we obtain a *bifurcation diagram*, which summarises all qualitative properties of the system. Our main goal in later sections 6–8 will be to derive the bifurcation diagrams of low-dimensional RNNs and thus to gain informations about the dynamical repertoire of these networks.

**Example 3.1 (saddle-node bifurcation)** Consider the following one-dimensional RNN

$$x(t+1) = \tanh(wx + u) \qquad x, u, w \in \mathbb{R}, \tag{3.2}$$

where $x$ is the activity and $u$ the external input of the neuron, which is considered as the parameter here. The self-feedback $w$ is assumed to be constant and shall satisfy $w > 1$. Then the system has a single (stable) fixed point for inputs $u < u_-$ and $u > u_+$, while it has three fixed points (two stable and one unstable) if $u \in (u_-, u_+)$. At the boundaries of this interval two new fixed points are created or destroyed through a saddle-node bifurcation. Figure 3.1a shows the corresponding bifurcation diagram in the direct product of the state and parameter spaces. Here stable and unstable fixed points are visualised as solid and dashed lines respectively.

The bifurcation points $u_\mp$ and their associated nonhyperbolic fixed points $\bar{x}_\mp$ can be computed from the fixed point condition and the saddle-node bifurcation condition $\tanh'(w\bar{x} + u)w = (1 - \bar{x}^2)w = 1$, which will be introduced later. Solving these equations we obtain:

$$u_\mp = \operatorname{arctanh}(\bar{x}_\mp) - w\bar{x}_\mp$$

$$\text{with fixed points} \qquad \bar{x}_\mp = \pm\sqrt{\frac{w-1}{w}} \qquad \text{for } w > 1.$$

In order to investigate the stability properties of these nonhyperbolic fixed points it does not suffice to consider the eigenvalue of the corresponding Jacobian, which equals one due to the saddle-node bifurcation condition. Rather we have to consider higher order terms of the Taylor series:

$$\tanh(wx + u_\mp) = \bar{x}_\mp + (x - \bar{x}_\mp) \mp \sqrt{w(w-1)}(x - \bar{x}_\mp)^2 + \mathcal{O}((x - \bar{x}_\mp)^3). \tag{3.3}$$

(a) bifurcation diagram            (b) graph for $u = u_-$            (c) graph for $u = u_+$
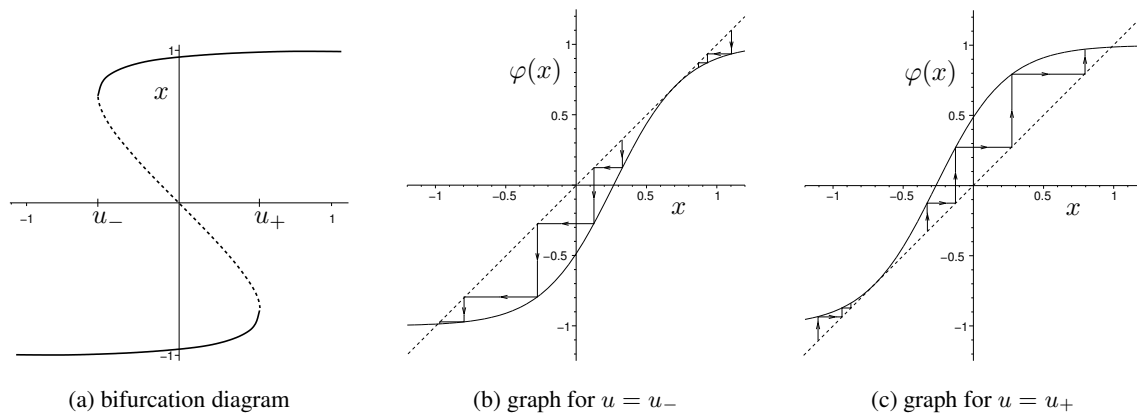
Figure 3.1: saddle-node bifurcations in RNN $x \mapsto \varphi(x) = \tanh(2x + u)$

Considering both the linear and quadratic term we can conclude, that the fixed points $\bar{x}_{\mp}$ are attracting from "outside", i.e. $x > \bar{x}_-$ resp. $x < \bar{x}_+$, and repelling from the other side, which is a typical behaviour at a saddle-node bifurcation. This result can be even better seen from the cobwebs in fig. 3.1b,c. This example shows further, that the stability properties of nonhyperbolic fixed points are complex in general.

In the simplest cases the bifurcation diagram consists of a finite number of regions, which are separated by smooth submanifolds, called *bifurcation boundaries* or *bifurcation manifolds*. In example 3.1 we found two bifurcation points ($u_-$ and $u_+$), separating three dynamical regions. On the bifurcation boundaries themselves we observe a different dynamical regime. When a bifurcation boundary is crossed, the corresponding bifurcation occurs and the dynamical behaviour changes qualitatively – otherwise the dynamical behaviour remains invariant in the sense of topological equivalence.

Bifurcation manifolds are defined by specifying a phase object, i.e. fixed point, periodic orbit, etc., and some bifurcation conditions determining the type of bifurcation. For example the saddle-node bifurcation of a fixed point is characterised by an eigenvalue $\lambda = 1$ of the Jacobian. The number of independent bifurcation conditions is called *codimension* of this bifurcation type and thus only depends on the type of bifurcation. For example, the saddle-node bifurcation has codimension one.

Knowledge of all bifurcation boundaries of neural networks on the one hand allows us to directly choose suitable parameter values to achieve a specific dynamical behaviour. On the other hand this knowledge helps to deepen our understanding of learning algorithms for neural networks. These usually aim at changes of the current dynamical behaviour, and thus might produce bifurcations along the learning path through parameter space. Knowing the bifurcation boundaries in this space we can track the appearing bifurcations and comprehend the actual effects of a learning rule.

In the following sections we summarise existing bifurcation theory relevant for discrete-time neural networks, basically citing results of Kuznetsov [1995] and Hoppensteadt and Izhikevich [1997]. If a bifurcation can be detected in any small neighbourhood of a phase object, i.e. a fixed point $\bar{x}$, the bifurcation is called *local*, otherwise it is called *global*. Although local bifurcations provide only informations on local changes, they can serve as building blocks to construct the global bifurcation diagram of a system. Most theoretical results concentrate on local bifurcations of codimension one or two, but there are also several outstanding results concerning local bifurcations of higher codimension and some global bifurcations.

Higher-codimension bifurcations can be stated in terms of low-codimensional ones in some cases: If the Jacobian matrix decomposes as a direct product of maps exhibiting low-codimension bifurcations each, the high-codimension bifurcation of the original map can be viewed as a direct product of the low-codimension bifurcations.

We can further simplify the analysis of local bifurcations, if we choose the lowest possible state space dimension $n$ and the lowest possible parameter dimension $m$. The needed state space dimension is usually determined by the number of eigenvalues on the unit circle needed for a certain bifurcation. For example a saddle-node bifurcation needs a single eigenvalue on the unit circle, while a Neimark-Sacker bifurcation needs a complex conjugate pair of eigenvalues having modulus one. If higher dimensional dynamical systems are considered, the center manifold reduction can be utilised to investigate the interesting nonhyperbolic dynamics only.

The needed parameter dimension is determined by the codimension of the bifurcation type, because a generic dynamical system needs exactly $m$ parameters to meet $m$ independent bifurcation conditions. These are usually fulfilled at a finite number of bifurcation points only. If we consider systems with more free parameters $m$ than bifurcation conditions $k$, we obtain bifurcation boundaries, which are $(m-k)$-dimensional submanifolds in the parameter space $\mathbb{R}^m$.

## 3.1 Topological equivalence of parameterised dynamical systems

In order to compare bifurcation diagrams of different parameterised dynamical systems we state a definition of topological equivalence of parameterised systems similar to definition 2.10 above.

**Definition 3.2** Two parameterised dynamical systems

$$\varphi_{\boldsymbol{\alpha}}^t : X_1 \to X_1 \quad \boldsymbol{\alpha} \in P_1 \subseteq \mathbb{R}^m, t \in \mathbb{T} \tag{3.4}$$

$$\text{and} \quad \psi_{\boldsymbol{\beta}}^t : X_2 \to X_2 \quad \boldsymbol{\beta} \in P_2 \subseteq \mathbb{R}^m, t \in \mathbb{T} \tag{3.5}$$

are called *topologically equivalent* if

- there exists a homeomorphism $p : P_1 \to P_2$ mapping parameter values $\boldsymbol{\alpha}$ to parameter values $\boldsymbol{\beta} = p(\boldsymbol{\alpha})$ and

- there exists a parameter-dependent homeomorphism $h_{\boldsymbol{\alpha}} : X_1 \to X_2$ mapping orbits of the first system at parameter values $\boldsymbol{\alpha}$ onto orbits of the second system at parameter values $\boldsymbol{\beta} = p(\boldsymbol{\alpha})$, preserving the direction of time.

If the parameterisation of time is preserved by $h_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} \in P_1$ as well, the systems are called *topologically conjugate*.

They are called *locally topologically equivalent* within open neighbourhoods $U \in P_1$, $U_{\boldsymbol{\alpha}} \in X_1$ and $V \in P_2$, $V_{\boldsymbol{\beta}} \in X_2$, if the parameter homeomorphism $p$ maps parameters from $U$ to $V$ and the parameter-dependent state-space homeomorphism $h_{\boldsymbol{\alpha}}$ maps orbits from $U_{\boldsymbol{\alpha}}$ onto orbits in $V_{\boldsymbol{\beta}}$, preserving the direction of time.

**Remark:** For discrete-time dynamical systems parameterisation of time is preserved automatically again, i.e. the systems are topologically conjugate, and we obtain the following commutative

diagram:

$$
\begin{array}{ccc}
P_1 & X_1 \xrightarrow{\;\varphi^1\;} X_1 \\[4pt]
{\scriptstyle p}\downarrow & {\scriptstyle h}\downarrow \qquad\quad \downarrow{\scriptstyle h} \\[4pt]
P_2 & X_2 \xrightarrow[\;\psi^1\;]{} X_2
\end{array}
$$

**Remark:**   The state-space homeomorphism $h_{\boldsymbol\alpha}$ is not required to depend continuously on $\boldsymbol\alpha$, which would imply that the map $(\boldsymbol{x}, \boldsymbol\alpha) \mapsto (h_{\boldsymbol\alpha}(\boldsymbol{x}), p(\boldsymbol\alpha))$ is a homeomorphism from the direct product $X_1 \times P_1$ onto $X_2 \times P_2$. For this reason the above definition is sometimes called *weak* topological equivalence. With respect to this definition, topologically equivalent parameter-dependent systems have topologically equivalent bifurcation diagrams.

## 3.2 Local Bifurcations

To describe local bifurcations of fixed points we usually specify a *topological normal form* for each bifurcation type, i.e. we specify a simple dynamical system $\boldsymbol{x} \mapsto \psi(\boldsymbol{x}, \boldsymbol\alpha)$ such that the system map $\psi$ is *polynomial* in $x_i$ and possesses at $\bar{\boldsymbol\alpha} = 0$ a fixed point $\bar{\boldsymbol{x}} = 0$ meeting $m$ bifurcation conditions determining a codim-$m$ bifurcation of this fixed point. If this system is locally topologically equivalent to each *generic* dynamical system satisfying the same bifurcation conditions at some fixed point $\bar{\boldsymbol{x}}$ and some bifurcation point $\bar{\boldsymbol\alpha}$ it can serve as a normal form for the considered bifurcation type and thus is called *topological normal form*.

In this context a dynamical system $\varphi$ is called generic, if it satisfies a finite number of *genericity* conditions, which have the form of nonequalities $C_i[\varphi] \neq 0$. Each $C_i$ is an algebraic function of certain partial derivatives of $\varphi(\boldsymbol{x}, \boldsymbol\alpha)$ evaluated at the bifurcation point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol\alpha})$. Note, that some bifurcation types do not have a topological normal form, which describes *all* relevant dynamical properties, e.g. the Neimark-Sacker bifurcation. Further topological normal forms are not unique in general. Nevertheless it is always possible to find a simple polynomial system, which exhibits most of the relevant properties and thus helps to improve our understanding of the given bifurcation.

As we can see from theorem 2.15 a necessary condition for a local bifurcation at a fixed point is its non-hyperbolicity. Having a general dynamical system, we distinguish three possibilities, how the hyperbolicity condition can be violated: a simple positive eigenvalue approaches the unit circle at $\lambda = 1$, a simple negative eigenvalue approaches the unit circle at $\lambda = -1$ or a pair of conjugate complex eigenvalues approach the unit circle at $\lambda_{1,2} = \mathrm{e}^{\pm\mathrm{i}\omega}$, $\omega \in (0, \pi)$. These cases are associated to the occurrence of a saddle-node, period-doubling and Neimark-Sacker bifurcation respectively. Obviously all these bifurcation have codimension one, because a single parameter is needed to meet the non-hyperbolicity condition.

In the following sections we list the topological normal forms of these bifurcations and state their exact bifurcation conditions. Further we specify a test function for each bifurcation type, which is easily computable and indicates the possible occurrence of the given bifurcation.

### 3.2.1 Saddle-node bifurcation

In example 3.1 we already considered the saddle-node bifurcation of the one-neuron dynamics. The following theorem describes the general form of a saddle-node bifurcation in a one-dimensional discrete-time dynamical system.

**Theorem 3.3 (Normal form of saddle-node bifurcation)** Suppose the dynamical system

$$x \mapsto \varphi(x, \boldsymbol{\alpha}) \qquad \varphi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$$

with smooth $\varphi$ has a nonhyperbolic fixed point $\bar{x}$ at a bifurcation point $\bar{\boldsymbol{\alpha}}$ with the multiplier $\lambda = \frac{\partial \varphi}{\partial x}(\bar{x}, \bar{\boldsymbol{\alpha}}) = 1$. If further the genericity conditions

(i) $\dfrac{\partial^2 \varphi}{\partial x^2}(\bar{x}, \bar{\boldsymbol{\alpha}}) \neq 0$ (nonzero quadratic term) and

(ii) $\boldsymbol{b} = \dfrac{\partial \varphi}{\partial \alpha}(\bar{x}, \bar{\boldsymbol{\alpha}}) \neq 0$ (transversality condition)

are satisfied, the system is locally topologically conjugate to

$$y \mapsto \beta + y + \sigma y^2 \tag{3.6}$$

with $\sigma = \mathrm{sgn}\left(\frac{\partial^2 \varphi}{\partial x^2}(\bar{x}, \bar{\boldsymbol{\alpha}})\right)$ and $\beta = \langle \boldsymbol{b}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle$.

Assuming $\sigma = +1$, this system exhibits two fixed points if $\beta < 0$, namely $\bar{x}_\pm(\beta) = \pm\sqrt{-\beta}$. The negative fixed point $\bar{x}_-$ is stable, while the other one is unstable. When $\beta$ approaches zero both fixed points coalesce and finally disappear for $\beta > 0$. If $\sigma = -1$ the situation reverses: the two fixed points exist for positive values of $\beta$ and disappear for negative values. In a multidimensional system the stable and unstable fixed points become a node and a saddle respectively, which motivates the name of this bifurcation type.

While the genericity condition (i) ensures that we are not at a Cusp bifurcation point (see next section), the transversality condition (ii) ensures that there is an independent parameter $\beta = \langle \boldsymbol{b}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle$ at all. An appropriate test function for a saddle-node bifurcation of a fixed point is

$$\det\left(D_{\boldsymbol{x}}\varphi(\bar{\boldsymbol{x}}, \boldsymbol{\alpha}) - \mathbf{1}\right) = 0, \tag{3.7}$$

which ensures – together with the fixed point equation $\varphi(\bar{\boldsymbol{x}}, \boldsymbol{\alpha}) = \bar{\boldsymbol{x}}$ – the existence of at least one multiplier $\lambda = 1$ of the Jacobian and thus indicates the possible occurrence of a saddle-node bifurcation. Because the genericity conditions (i) and (ii) are not checked with (3.7) this condition might be satisfied even though there is no saddle-node bifurcation but a bifurcation of higher codimension.

## 3.2.2 Cusp bifurcation

If condition (i) of the saddle-node bifurcation is violated, i.e. the quadratic term is zero, a Cusp bifurcation occurs, which is a codim-2 bifurcation, because two independent equations have to be fulfilled. The following theorem summarises the necessary conditions for a typical Cusp bifurcation.

**Theorem 3.4 (Normal form of Cusp bifurcation)** Suppose the dynamical system

$$x \mapsto \varphi(x, \boldsymbol{\alpha}) \qquad \varphi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}, m \geq 2$$

with smooth $\varphi$ has a nonhyperbolic fixed point $\bar{x}$ at a bifurcation point $\bar{\boldsymbol{\alpha}}$ with the multiplier $\lambda = \frac{\partial \varphi}{\partial x}(\bar{x}, \bar{\boldsymbol{\alpha}}) = 1$ and a zero quadratic term of the Taylor expansion, i.e. $\frac{\partial^2 \varphi}{\partial x^2}(\bar{x}, \bar{\boldsymbol{\alpha}}) = 0$. If further the genericity conditions

(a) bifurcation curve      (b) Cusp surface      (c) Pitchfork bifurcation
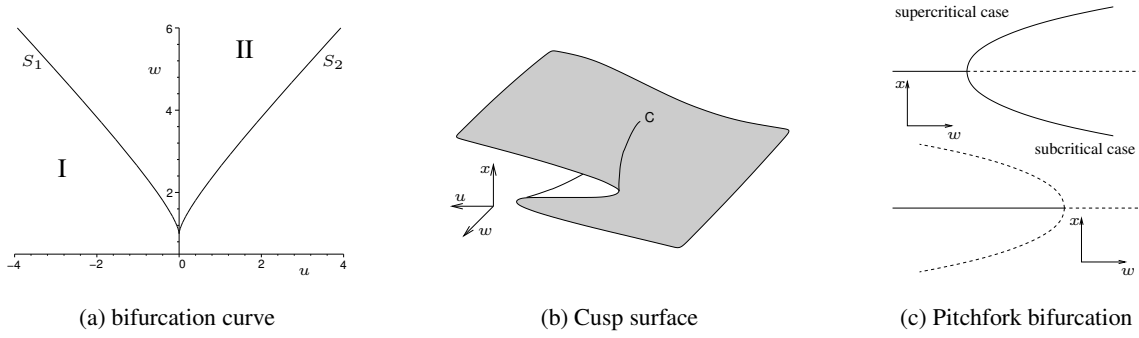
Figure 3.2: Saddle-node bifurcation manifold of RNN (3.2)

(i)  $\dfrac{\partial^3 \varphi}{\partial x^3}(\bar{x}, \bar{\boldsymbol{\alpha}}) \neq 0$ (nonzero cubic term) and

(ii)  $\boldsymbol{b} = \dfrac{\partial \varphi}{\partial \alpha}(\bar{x}, \bar{\boldsymbol{\alpha}})$ and $\boldsymbol{c} = \dfrac{\partial^2 \varphi}{\partial \alpha \partial x}(\bar{x}, \bar{\boldsymbol{\alpha}})$ are linearly independent (transversality condition)

are satisfied, the system is locally topologically conjugate to

$$y \mapsto \beta + (1 + \gamma)\, y + \sigma y^3 \tag{3.8}$$

with $\sigma = \mathrm{sgn}\big(\frac{\partial^3 \varphi}{\partial x^3}(\bar{x}, \bar{\boldsymbol{\alpha}})\big)$, $\beta = \langle \boldsymbol{b}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle$ and $\gamma = \langle \boldsymbol{c}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle$.

Notice, that a Cusp bifurcation always occurs on a saddle-node bifurcation manifold, because in addition to the saddle-node bifurcation condition $\lambda = 1$ the condition on the quadratic term has to be satisfied. As before condition (i) ensures that we are away from a higher codim bifurcation and the transversality condition (ii) ensures that the parameters $\beta$ and $\gamma$ are indeed independent.

**Example 3.5**  To discuss the properties of the Cusp bifurcation we consider once more the one-neuron dynamics (3.2), however with the self-feedback $w$ as a second parameter. The quadratic term of the Taylor series (3.3) vanishes at $w = 1$, such that we observe a Cusp bifurcation at $(\bar{w}, \bar{u}) = (1, 0)$ and $\bar{x} = 0$. The third order term is negative and the parameters $w$ and $u$ are indeed independent of each other, such that the genericity conditions are satisfied as well. Actually we get $\beta = u$, $\gamma = w - \bar{w}$ and $\sigma = -1$ for the topological normal form (3.8). The bifurcation boundary of the saddle-node bifurcation is a one-dimensional manifold in the two-dimensional parameter space with the two branches $S_1$ and $S_2$ which meet tangentially at the Cusp point $(\bar{w}, \bar{u})$ (see fig. 3.2a,b). The resulting wedge separates the parameter plane into two regions. Within region I there is a single stable fixed point, while inside region II coexist three fixed points, two of them stable and one unstable. If the saddle-node bifurcation curve is crossed away from the Cusp point, a nondegenerate saddle-node bifurcation occurs, creating or annihilating two fixed points. At the Cusp point itself all three fixed points coalesce.

If the sign of the cubic term of the topological normal form (3.8) is positive instead, the situation is similar, however we observe a single unstable fixed point within region I and two unstable fixed points together with a stable one within region II. This case is called subcritical Cusp bifurcation in contrast to the former case which is called supercritical. As we have seen in the above example a single neuron cannot exhibit a subcritical Cusp bifurcation, and we will see later that this bifurcation typically cannot be observed in large RNNs as well.

## Hysteresis

Near a Cusp bifurcation point a hysteresis effect occurs. If in fig. 3.2a the parameters are varied along a path starting in the left half of region I two new fixed points are created, if the first bifurcation boundary $S_1$ is crossed. Because initially the "old" fixed point remains stable, we do not observe any changes of the phase portrait in its neighbourhood – the new fixed points are created outside its neighbourhood. If the parameters are varied further into the direction of the secondary bifurcation boundary $S_2$, the newly created unstable fixed point approaches the "old" stable one until they coalesce and annihilate at the boundary. Hence the newly created stable fixed point remains as the unique attractor. This behaviour is illustrated in fig. 3.1a, which shows a cross section of fig. 3.2b at $w = 2$.

## Pitchfork bifurcation

If we cross the bifurcation boundary at the Cusp point tangentially to both branches $S_1$ and $S_2$, we obtain the bifurcation diagram depicted in fig. 3.2c. Because of its form the corresponding bifurcation is called pitchfork bifurcation. Considering the topological normal form (3.8), this corresponds to the additional condition $\boldsymbol{b} = 0$, i.e. $\beta = 0$, resulting in a symmetric system, i.e. $\varphi(y) = -\varphi(-y)$, which has the fixed point $\bar{y} = 0$ always. Because this symmetry is a non-generic property of dynamical systems, the pitchfork bifurcation can be viewed as a *degenerate* Cusp bifurcation. Depending on the sign of the cubic term, we speak of a supercritical ($\sigma < 0$) or subcritical ($\sigma > 0$) pitchfork bifurcation again.

### 3.2.3 Period-doubling bifurcation

If the nonhyperbolic fixed point has a multiplier $\lambda = -1$, a period-2 cycle occurs or disappears. The following theorem defines the corresponding normal form.

**Theorem 3.6 (Normal form of period-doubling bifurcation)** Let the dynamical system

$$x \mapsto \varphi(x, \boldsymbol{\alpha}) \qquad \varphi : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$$

with smooth $\varphi$ have a nonhyperbolic fixed point $\bar{x}$ at a bifurcation point $\bar{\boldsymbol{\alpha}}$ with the multiplier $\lambda = \frac{\partial \varphi}{\partial x}(\bar{x}, \bar{\boldsymbol{\alpha}}) = -1$. If further the genericity conditions

(i) $\dfrac{\partial^3(\varphi \circ \varphi)}{\partial x^3}(\bar{x}, \bar{\boldsymbol{\alpha}}) \neq 0$ (nonzero cubic term of $\varphi^2$) and

(ii) $\boldsymbol{b} = \dfrac{\partial^2 \varphi}{\partial \alpha \partial x}(\bar{x}, \bar{\boldsymbol{\alpha}}) \neq 0$ (transversality condition)

are satisfied, the system is locally topologically conjugate to

$$y \mapsto -(1 + \beta)\, y - \sigma y^3 \tag{3.9}$$

where $\sigma = \mathrm{sgn}\big(\frac{\partial^3 \varphi^2}{\partial x^3}(\bar{x}, \bar{\boldsymbol{\alpha}})\big)$.

The second iterate $\varphi^2$ is locally topologically conjugate to

$$y \mapsto (1 + \beta)\, y + \sigma y^3 \qquad \text{where } \beta = \langle \boldsymbol{b}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle, \tag{3.10}$$

which is the topological normal form of a pitchfork bifurcation. We consider the supercritical and subcritical case separately. If $\sigma < 0$ the permanent fixed point $\bar{y} = 0$ looses its stability if $\beta$ is

varied from negative to positive values. At $\beta = 0$ the nonhyperbolic fixed point $\bar{y} = 0$ is stable too, which is ensured by the positive sign of the cubic term. If $\beta$ becomes positive, the pitchfork bifurcation generates two stable fixed points $\bar{y}_{1,2}$ of the second iterate $\varphi^2$, which correspond to the periodic points of a stable period-2 cycle of $\varphi$. All trajectories, except the one starting at the origin, converge to this limit cycle.

In the subcritical case, if $\sigma > 0$, the situation is as follows: Again the hyperbolic fixed point $\bar{y} = 0$ is stable if $\beta$ is negative, while it is unstable if this parameter is positive. However, at the bifurcation point $\beta = 0$ itself the fixed point is unstable, because the joint effect of the linear and cubic terms amplify small perturbations around the origin. The analysis of the pitchfork bifurcation of the second iterate (3.10) reveals an unstable period-2 cycle for negative parameters which disappears at the bifurcation point.

In both cases the stable fixed point $\bar{y} = 0$ looses its stability if the parameter $\beta$ is varied from negative to positive values. In the first (supercritical) case, there exists an attractor near the origin before and after the bifurcation occurs. Hence, a trajectory starting in a neighbourhood of the origin never leaves far away. For this reason this case is called a *soft* bifurcation. In the latter (subcritical) case for positive parameter values there does not exist an attractor near the origin. This case is called a *hard* bifurcation, because every trajectory starting near the origin will leave this region.

An appropriate test function for the period-doubling bifurcation has to indicate the occurrence of a multiplier $\lambda = -1$ of the Jacobian. Hence we obtain analogously to the saddle-node bifurcation the following test function:

$$\det\left(D_{\boldsymbol{x}}\varphi(\bar{\boldsymbol{x}}, \boldsymbol{\alpha}) + \boldsymbol{1}\right) = 0\,. \tag{3.11}$$

### 3.2.4 Neimark-Sacker bifurcation

In the following we consider the two-dimensional dynamical system

$$\boldsymbol{x} \mapsto A\left[(1 + \beta)\,\boldsymbol{x} + \|\boldsymbol{x}\|^2\,B\,\boldsymbol{x}\right], \tag{3.12}$$

where

$$A = \begin{pmatrix} \cos\theta(\beta) & -\sin\theta(\beta) \\ \sin\theta(\beta) & \cos\theta(\beta) \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} a(\beta) & -b(\beta) \\ b(\beta) & a(\beta) \end{pmatrix}.$$

The functions $\theta(\beta)$, $a(\beta)$ and $b(\beta)$ are assumed to be smooth and shall satisfy $0 < \theta(0) < \pi$ and $\sigma := a(0) \neq 0$. Using polar coordinates we can rewrite this system as

$$\varrho \mapsto (1 + \beta)\,\varrho + a(\beta)\,\varrho^3 + \mathcal{O}(\varrho^5) \qquad \varrho \geq 0$$

$$\phi \mapsto \phi + \theta(\beta) + b(\beta)\,\frac{\varrho^2}{1 + \beta} + \mathcal{O}(\varrho^4). \tag{3.13}$$

which can be analysed easily since the dynamics of $\varrho$ is independent of $\phi$. Actually the $\varrho$-dynamics takes the shape of the normal form of a pitchfork bifurcation and the $\phi$-dynamics describes a rotation about a constant angle, which depends on $\varrho$ and $\beta$. Again we have to consider the supercritical ($\sigma < 0$) and subcritical ($\sigma > 0$) case of the pitchfork bifurcation. For the supercritical case, the stable fixed point

$$\bar{\varrho} = \sqrt{-\frac{\beta}{a(\beta)}} + \mathcal{O}(\beta) \tag{3.14}$$

$\beta < 0$          $\beta > 0$          $\beta < 0$          $\beta > 0$

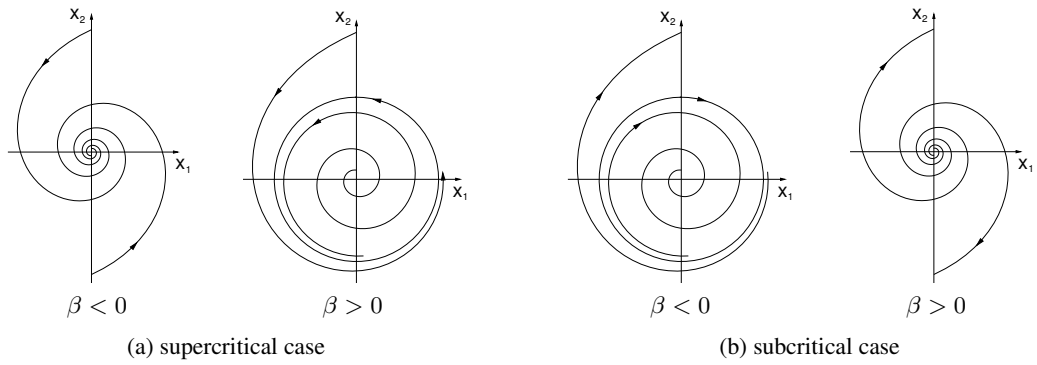(a) supercritical case                              (b) subcritical case

Figure 3.3: Neimark-Sacker bifurcation

appears for positive parameter values $\beta$, which leads in conjunction with the $\phi$-dynamics to the occurrence of an isolated invariant curve, which is unique and asymptotically stable. All trajectories, except the one starting at the origin, approach this curve under forward iteration (fig 3.3a). In the subcritical case there exists an unstable invariant closed curve for $\beta < 0$, which disappears at the bifurcation point $\beta = 0$. The stability properties of the hyperbolic origin are identical in both cases. Thus the subcritical Neimark-Sacker bifurcation leads to a hard loss of stability of the fixed point $\bar{\boldsymbol{x}} = 0$ (fig. 3.3b).

If higher order terms are permitted in (3.12) the $\varrho$-dynamics of the corresponding polar form becomes $\phi$-dependent and the dynamical system is not topologically equivalent to (3.13) anymore. Although higher order terms do not affect the bifurcation of a closed invariant curve, which now differs from a circle, the dynamics on the curve changes heavily. While the latter system, which incorporates higher order terms, generically has only a finite number of periodic orbits, for the simple system (3.12) *each point* on the invariant circle is periodic or quasiperiodic [Kuznetsov, 1995]. For a general dynamical system we have the following theorem:

**Theorem 3.7 ("Normal form" of Neimark-Sacker bifurcation)**
Suppose the two-dimensional dynamical system

$$\boldsymbol{x} \mapsto \varphi(\boldsymbol{x}, \boldsymbol{\alpha}) \qquad \varphi : \mathbb{R}^2 \times \mathbb{R}^m \to \mathbb{R}^2$$

with smooth $\varphi$ has a nonhyperbolic fixed point $\bar{\boldsymbol{x}}$ at a bifurcation point $\bar{\boldsymbol{\alpha}}$ with the pair of complex conjugate multipliers $\lambda_{1,2} = \mathrm{e}^{\pm \mathrm{i}\omega}$. If further the genericity conditions

(i)  $\mathrm{e}^{\mathrm{i}\omega k} \neq 1$ for $k = 1, 2, 3, 4$ (no strong resonances) and

(ii) $D_{\boldsymbol{x}}\varphi(\bar{\boldsymbol{x}}, \boldsymbol{\alpha})$ has complex conjugate eigenvalues $\lambda_{1,2} = r(\boldsymbol{\alpha})e^{\pm \mathrm{i}\omega(\boldsymbol{\alpha})}$ within a small neighbourhood of $\bar{\boldsymbol{\alpha}}$ and it holds $\frac{\partial r(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}(\bar{\boldsymbol{\alpha}}) \neq 0$ (transversality condition)

are satisfied, the system is locally topologically conjugate to

$$\begin{aligned} \varrho &\mapsto (1 + \beta)\,\varrho + \sigma \varrho^3 + \dots \\ \phi &\mapsto \phi + \omega + b(\beta)\varrho^2 + \dots \end{aligned} \tag{3.15}$$

with $\beta = \langle \boldsymbol{b}, \boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}} \rangle$. If further the genericity condition $\sigma \neq 0$ is satisfied, there exists a neighbourhood of $\bar{\boldsymbol{x}}$ in which a unique closed invariant curve bifurcates from $\bar{\boldsymbol{x}}$ as $\boldsymbol{\alpha}$ passes through $\bar{\boldsymbol{\alpha}}$. This curve is asymptotically stable if $\sigma < 0$ (supercritical case) and unstable if $\sigma > 0$ (subcritical case).

The ellipses in equation (3.15) denote higher order terms including both state variables $\varrho$ and $\phi$. They account for a complex dynamics on the invariant curve and cannot be omitted in general. For this reason it is impossible to find a simple topological *normal form* of the Neimark-Sacker bifurcation, which should be locally topologically equivalent to any dynamical system satisfying the Neimark-Sacker bifurcation conditions. The coefficient $\sigma$, whose sign determines the type of bifurcation, can be computed as follows:

$$\sigma = -\operatorname{Re}\left(\frac{(1-2\lambda)\bar{\lambda}^2}{1-\lambda}g_{20}g_{11}\right) + \operatorname{Re}\left(\bar{\lambda}g_{21}\right) - \tfrac{1}{2}|g_{11}|^2 - |g_{02}|^2$$

where

$$\lambda = \mathrm{e}^{\mathrm{i}\omega}$$
$$g_{11} = \tfrac{1}{4}\left(\varphi^1_{xx} + \varphi^1_{yy} + \mathrm{i}(\varphi^2_{xx} + \varphi^2_{yy})\right)$$
$$g_{20} = \tfrac{1}{8}\left(\varphi^1_{xx} - \varphi^1_{yy} + 2\varphi^2_{xy} + \mathrm{i}(\varphi^2_{xx} - \varphi^2_{yy} - 2\varphi^1_{xy})\right)$$
$$g_{02} = \tfrac{1}{8}\left(\varphi^1_{xx} - \varphi^1_{yy} - 2\varphi^2_{xy} + \mathrm{i}(\varphi^2_{xx} - \varphi^2_{yy} + 2\varphi^1_{xy})\right)$$
$$g_{21} = \tfrac{1}{16}\left(\varphi^1_{xxx} + \varphi^1_{xyy} + \varphi^2_{xxy} + \varphi^2_{yyy} + \mathrm{i}(\varphi^2_{xxx} + \varphi^2_{xyy} - \varphi^1_{xxy} - \varphi^1_{yyy})\right)$$

where we used the notation $\varphi(x, y) = (\varphi^1(x, y), \varphi^2(x, y))$ and the subscripts indicate partial differentiation with respect to $x$ or $y$. All derivatives are evaluated at $(\bar{x}, \bar{\alpha})$. If $\sigma$ equals zero we obtain a codim-2 bifurcation, called Chenciner bifurcation. Because we do not encounter this bifurcation type during this work, we skip its detailed discussion and refer to Kuznetsov [1995] instead. The topological normal form (3.15) is often written in complex form $z = x + \mathrm{i}y = \varrho\mathrm{e}^{\mathrm{i}\phi}$:

$$z \mapsto \lambda(\beta)z + \sigma(\beta)|z|^2 z + \mathcal{O}(|z|^4) \tag{3.16}$$

with $\lambda(\beta) = (1 + \beta)\mathrm{e}^{\mathrm{i}\omega(\beta)}$ and $\lambda(0) = e^{\mathrm{i}\omega}$. Here $\sigma(0)$ equals $\sigma$, which determines the type of bifurcation.

Theorem 3.7 guarantees the existence of a closed invariant curve near a Neimark-Sacker bifurcation point. If higher order terms, which are hidden in the ellipses, are involved the dynamics on this curve can be very complicated, including quasiperiodic and periodic orbits of any periodicity. Denote with $\psi : \mathbb{R} \to \mathbb{R}$ the restriction of $\varphi$ to the invariant curve. The key to understand the dynamics on the curve is the *rotation number*, describing the average rotation angle $\langle\Delta\phi\rangle$ on the curve:

$$R = \frac{1}{2\pi}\lim_{k\to\infty}\frac{1}{k}\sum_{t=0}^{k-1}\psi^{t+1}(\phi) - \psi^t(\phi) = \frac{\langle\Delta\phi\rangle}{2\pi}\,.$$

If the rotation number is rational, i.e. $R = \frac{p}{q}$, there exist an even number of $q$-cycles on the curve, where stable and unstable (saddle) periodic points alternate (fig. 3.4). These periodic points correspond to stable and unstable fixed points of $\psi^q$. After $q$ iterations of such a $q$-cycle, the orbit has finished $p$ revolutions on the curve. If the rotation number is irrational and $\psi \in C^2$, a theorem of Denjoy [1932] states, that the dynamics on the curve is topologically equivalent to a pure rotation on the unit circle by the angle $2\pi R$. Thus all orbits are quasiperiodic on the invariant curve. Interestingly the dynamics on the curve is structurally stable if and only if the rotation number is rational and all periodic orbits are hyperbolic. Thus there exist finite parameter windows corresponding to periodic dynamics of constant rotation number. At the borders of these windows the stable and unstable $q$-cycles appear or disappear through saddle-node bifurcations of $\psi^q$. Outside such parameter windows the rotation number is irrational and we observe quasiperiodic orbits only.
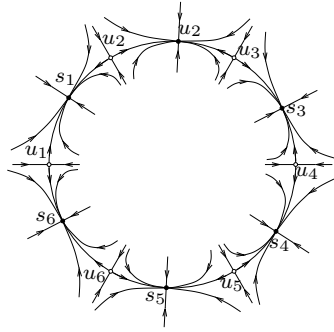
Figure 3.4:   unstable (saddle) and stable period-6  Figure 3.5:   Arnold tongues originating from the
orbits on the invariant curve.                                    unit circle.
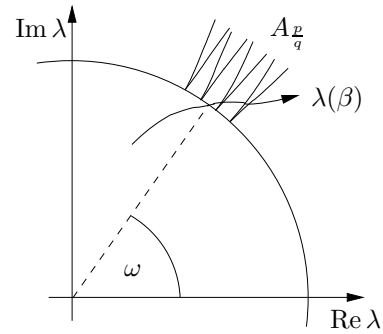
To investigate this phenomenon further, consider eq. (3.16). If the parameter $\beta$ varies, $\lambda(\beta)$ describes a path in the complex plane, crossing the unit circle at $\beta = 0$. Parameter windows corresponding to rational rotation numbers $R = \frac{p}{q}$, define regions $A_{\frac{p}{q}}$ in the complex plane, called Arnold tongues (see fig. 3.5). Because these tongues originate from $q$-th unit roots on the unit circle, a general $\lambda(\beta)$-curve crosses infinitely many tongues, such that an infinite number of high-periodic cycles are born and killed as the parameter $\beta$ varies. Far away from the Neimark-Sacker bifurcation point these tongues can intersect. At such parameter values the invariant curve does not exist anymore [Kuznetsov, 1995].

**Test function**

An appropriate test function for the Neimark-Sacker bifurcation has to indicate the occurrence of complex conjugate eigenvalues on the unit circle. Therefore a possible test function is

$$\prod_{i>j} \lambda_i \lambda_j - 1 = 0 \,. \tag{3.17}$$

To detect a true Neimark-Sacker bifurcation, we have additionally to check that this condition is not fulfilled by two real eigenvalues with unit product ($\lambda_1 \lambda_2 = 1$), indicating a saddle only. The usage of the bialternate product, allows us to express the test function in terms of the Jacobian matrix itself.

**Definition 3.8** The bialternate product[1] of matrices $A$ and $B \in \mathbb{R}^{n \times n}$, denoted by $A \odot B$, is an $m \times m$ matrix with $m = \frac{1}{2}n(n-1)$ defined as

$$(A \odot B)_{(p,q),(r,s)} = \frac{1}{2} \left( \begin{vmatrix} a_{pr} & a_{ps} \\ b_{qr} & b_{qs} \end{vmatrix} + \begin{vmatrix} b_{pr} & b_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} \right) \,, \tag{3.18}$$

where $p, r = 2 \ldots n$ and $q = 1 \ldots p-1, s = 1 \ldots r-1$. The indices are ordered lexicographically.

Due to a theorem of Stéphanos [1900] the eigenvalues of $A \odot A$ equal $\lambda_i \lambda_j$ where $\lambda_i$ denotes the eigenvalues of $A$. Thus we can rewrite the test function (3.17) as

$$\det(A \odot A - \mathbf{1}) = 0 \,. \tag{3.19}$$

This test function together with the fixed point equation $\varphi(\bar{x}, \alpha) = \bar{x}$ defines a $(n-1)$-dimensional manifold in the cross product of the state and parameter space. Subsets on this manifold corresponding to fixed points with complex conjugate eigenvalues on the unit circle indeed define

---

[1]   The bialternate product is called Kronecker product as well.

Neimark-Sacker bifurcation points, whereas points on the manifold corresponding to fixed points with real eigenvalues $\lambda_1 \lambda_2 = 1$ should be ignored. Along the boundaries between these different subsets the corresponding complex or real eigenvalues coalesce at $\lambda_1 = \lambda_2 = \pm 1$. These boundaries define codim-2 bifurcation manifolds, corresponding to 1:1 and 1:2 resonances respectively [Kuznetsov, 1995].

Fixed points, which have real eigenvalues $\lambda_1 \lambda_2 = 1$, are called *neutral saddle* or *neutral flip saddle*, depending on the sign of the eigenvalues. We will discuss the dynamical behaviour in the neighbourhood of such points in the context of oscillatory orbits of RNNs in section 7.3 in more detail.

## 3.3 Soft and hard loss of stability

We have seen, that we have to distinguish between a supercritical and a subcritical type in case of Cusp, period-doubling and Neimark-Sacker bifurcations. In all these cases a pitchfork bifurcation is involved, which originally leads to this distinction. To discuss their difference in more detail, consider a stable fixed point $\bar{x}$ undergoing a bifurcation at the parameter value $\bar{\alpha} = 0$, i.e. a fixed point which looses its stability or disappears at all if the parameter $\alpha$ is varied from negative to positive values. We have to distinguish the following two cases:

- Soft: For every small $\alpha > \bar{\alpha}$ there exists an attractor of small extension, e.g. an oscillatory orbit of small amplitude, which is located in a small neighbourhood of the previously stable fixed point $\bar{x}$.

- Hard: For an arbitrary small $\alpha > \bar{\alpha}$ there exists no attractor in a neighbourhood of $\bar{x}$.

In the first case, every trajectory in the vicinity of the former stable fixed point $\bar{x}$ approaches the new attractor and thus will not leave this neighbourhood. In the second case, there does not exist such an attractor and any trajectory leaves the vicinity of the former fixed point. If the dynamical system describes a physical process having its working point around a given stable fixed point $\bar{x}$, a soft loss of stability is uncritical for this process and may be not noticed at all. But a hard loss of stability may have catastrophic consequences, because the working point is left.

The supercritical cases of the Cusp, pitchfork, period-doubling, and Neimark-Sacker bifurcations correspond to a soft loss of stability, whereas the subcritical cases correspond to a hard loss of stability. This can be easily seen from the discussions in the previous sections. The loss of stability via the saddle-node bifurcation, i.e. disappearance of a stable and an unstable fixed point, is always hard, unless the bifurcation is close to a supercritical Cusp bifurcation, where the existence of a third, stable fixed point is guaranteed. Contrary the creation of a pair of fixed points or periodic orbits through a saddle-node bifurcation of $\varphi^T$ is always soft, because initially existing attractors are not influenced.

# 4 Global Stability

The most fundamental task in the analysis of the dynamical behaviour of the RNN (2.5) is to determine its fixed points and their stability properties. Most neural network applications interpret these stationary or final states as the output of the network or its stored patterns. In these cases it has to be assured that

(i) every trajectory converges to a fixed point and

(ii) fixed points representing stored patterns have a sufficiently large basin of attraction such that only a few trajectories approach other attractors or spurious patterns, which do not correspond to stored patterns at all. The occurrence of such spurious patterns cannot be impeded in general, but there exist learning algorithms which minimise the number of spurious states [Michel et al., 1991; Yune and Michel, 1992; Park et al., 2001].

Many network architectures have a gradient based dynamics which ensures global convergence to stationary states due to the existence of a global Lyapunov function. Among these networks are the classical Hopfield network [Hopfield, 1984; Markus and Westervelt, 1989], the Brain-State-In-A-Box (BSB) [Golden, 1993], and the CLM architecture [Wersing et al., 2001b]. For the construction of an energy or Lyapunov function usually symmetry of the weight matrix $W$ has to be assumed. While the classical Cohen-Grossberg function [Cohen and S. Grossberg, 1983] needs differentiability of the activation function, Feng [1997] constructs a Lyapunov function for discrete-time recurrent neural networks with non-differentiable activation functions.

If neural networks are used to solve optimisation problems we want to ensure global asymptotic stability of a *unique* fixed point which then corresponds to the global optimum. Other fixed points which would correspond to local optima of an energy function should be excluded [Forti, 1996; M. Forti and A. Tesi, 1995]. Deriving sufficient and necessary conditions for global asymptotic stability of a unique fixed point of a specific dynamical system or for a class of systems is a very hard task or leads to conditions which are hard to verify [Molchanov and Liu, 2002].

If strong restrictions on the weight matrices are imposed, conservative stability criteria based on global contraction [Cessac, 1994], symmetry or diagonal dominance [Hirsch, 1989] can be derived, which hold for large classes of dynamical systems. It is worth to mention, that the symmetry of the connection matrix does not ensure robustness of stability with respect to small perturbations [di Marco et al., 2000, 2002]. Hence stability may be lost in the presence of arbitrary small noise on the weights.

The development of stability criteria for sector-bounded nonlinearities within the framework of *absolute stability theory* founded by Lur'e and Postnikov [1944] lead to the powerful *frequency theorem* [Yakubovich, 1973; Tsypkin, 1964], which was used by many authors to prove stability for different classes of dynamical systems [Steil, 1999; Barabanov and Prokhorov, 2002]. This theorem is based on the reduction of a class of nonlinear systems to a class of linear systems for which stability can be proven. Imposing additional assumptions on the nonlinear dynamical systems, stability of these can be concluded as well. Because this theorem usually proves global asymptotic stability of a whole class of dynamical systems, we refer to this type of stability as *absolute stability*. But the theorem can be employed as well to prove local stability of fixed points within a restricted region of admissible weights and inputs [Steil, 2002].

## 4.1 Existence and Uniqueness of Fixed Points

RNNs with sigmoidal, i.e. bounded activation functions, have a compact convex state space, and consequently Brouwer's fixed point theorem guarantees the existence of at least one fixed point. If boundedness of activation functions is not given, other methods have to be used to prove existence of equilibria [Wersing et al., 2001a; Hu and Wang, 2002]. For example, Hu and Wang derive necessary and sufficient conditions for the existence and uniqueness of fixed points for each element in a class of systems defined by a global Lipschitz condition of their activation functions. This class of nonlinear functions includes unbounded and not differentiable functions. The only additional condition needed to prove uniqueness of fixed points is the nonsingularity of all matrices within a matrix polytope, which turns out to be NP-hard to evaluate [Rohn, 1994a; Nemirovskii, 1993]. If in addition the activation function is differentiable, this matrix polytope is identical to the set of Jacobian matrices evaluated at every state vector $x \in X$ and for every nonlinearity within the considered function class. This result is especially interesting because it gives an impression of the class of systems, which have a unique fixed point but do not require differentiability of their activation functions. If additionally continuous differentiability and boundedness of the activation function is given, it is possible to infer *continuous dependency* of the unique fixed point on the input vector $u$ [Bhaya et al., 1996]. Hence, if the RNN (2.5) has a unique fixed point – which can be assured by Hu and Wang's condition – this fixed point continuously varies with $u$.

## 4.2 Global Asymptotic Stability of Fixed Points

In order to prove that all trajectories converge to the unique fixed point, i.e. to prove its global asymptotic stability (gas), modern techniques utilise a global Lyapunov function. Note, that global asymptotic stability excludes the existence of other limit sets, e.g. oscillatory or chaotic attractors. In order to review some important results on absolute stability we begin with systems which have zero input $u = 0$. Later, we will generalise these results to systems with constant non-zero input. All discussed theorems are based on the frequency theorem, i.e. are applicable to slope-bounded nonlinearities and impose conditions on the weight matrix, which ensure simultaneous stability of a whole matrix polytope with respect to *time-varying linear* dynamical systems. A short introduction to the theory of matrix stability – particularly to the terms (diagonal) Schur stability, simultaneous stability and **D**-stability – is given in appendix B. The reader, who is not familiar with these terms, should first consider this review before continuing reading.

Kaszkurewicz and Bhaya [1993] investigated the global asymptotic stability of linear dynamical systems, which are subject to time-varying state-dependent perturbations (4.1) or to time-varying parameter variations (4.2), whose effects can be described by the following nonlinear models:

$$x_i(t + 1) = \sum w_{ij}\phi_j(x_j(t), t) \qquad \text{and} \qquad (4.1)$$

$$x_i(t + 1) = \sum w_{ij}\phi_{ij}(x_j(t), t) \,. \qquad (4.2)$$

The nonlinearities $\phi_j$ and $\phi_{ij}$ satisfy the standard *sector conditions* (see figure 4.1):

$$\mathcal{S} := \{\phi : \mathbb{R} \times \mathbb{T} \to \mathbb{R} \,|\, |\phi(x, t)| \leq |x| \text{ and } \phi(0, t) = 0 \text{ for all } t \in \mathbb{T}\} \,. \qquad (4.3)$$

While the system (4.1) covers all time-varying nonlinear activation functions, the latter system (4.2) additionally allows time-varying weights due to the different perturbation functions $\phi_{ij}$ for each interaction weight $w_{ij}$. Particularly it includes additive perturbations of the form $(w_{ij} + \Delta w_{ij}(t))$ as well as multiplicative perturbations of the form $(w_{ij} \cdot \Delta w_{ij}(t))$. The following sufficient conditions for absolute stability of these systems have been proven:
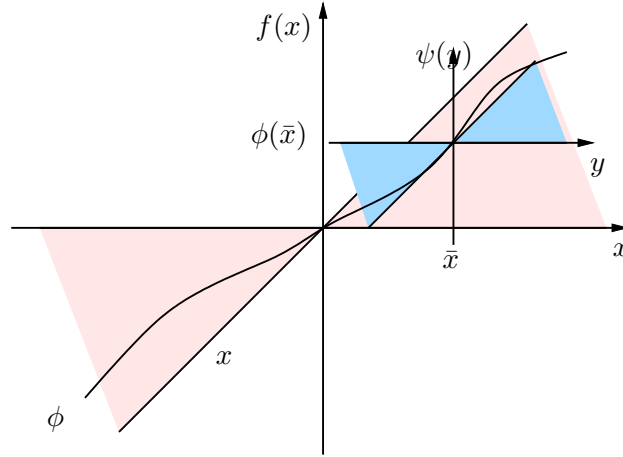
Figure 4.1: The graph of a function $\phi(x)$ which satisfies the standard sector conditions (4.3) is bounded within the sector $f(x) = 0$ and $f(x) = x$. As can be seen from the smaller sector which is obtained by shifting the coordinate frame to $(\bar{x}, \phi(\bar{x}))$, $\psi(y) = \phi(y + \bar{x}) - \phi(\bar{x})$ does not satisfy the sector condition. Hence $\phi(x)$ and $\psi(y)$ do not satisfy the standard incremental sector condition (4.8).

**Theorem 4.1** The fixed point $\boldsymbol{x} = 0$ is globally asymptotically stable

  (i) for all systems (4.1), if $W$ is diagonally Schur stable.

 (ii) for all systems (4.2), if $|W|$ is diagonally Schur stable.

Assuming *identical* nonlinearities $\phi_i = \phi$ in (4.1), Chu and Glover [1999] were able to prove absolute stability with a less restrictive condition on the weight matrix $W$:

**Theorem 4.2** Consider the dynamical system (4.1) with identical nonlinearities $\phi_i = \phi$ from the class of functions:

$$\mathcal{OL} := \{\phi : \mathbb{R} \to \mathbb{R} \,|\, \phi \text{ odd and globally Lipschitz with Lipschitz constant } L = 1\}. \qquad (4.4)$$

If there exists a positive definite and diagonally dominant matrix $P$, such that Stein's inequality $W^t P W - P < 0$ is fulfilled, the fixed point $\boldsymbol{x} = 0$ is globally asymptotically stable.

Clearly $\mathcal{OL}$ is a subset of $\mathcal{S}$. Thus Chu and Glover [1999] considered time-constant, identical nonlinearities $\phi$ satisfying the more restrictive condition $\mathcal{OL}$ in order to obtain a less restrictive condition on $W$. Because one often deals with a time-constant neural network having identical activation functions $\sigma$ for each neuron, this result is quite useful.

The proofs of theorems 4.1(i) and 4.2 rely on the inequality

$$\boldsymbol{\phi}(\boldsymbol{x})^t P \boldsymbol{\phi}(\boldsymbol{x}) \leq \boldsymbol{x}^t P \boldsymbol{x} \quad \text{ for all } \boldsymbol{x} \in \mathbb{R}^n \text{ and all } \boldsymbol{\phi} \in \mathcal{S} \text{ resp. } \mathcal{OL}, \qquad (4.5)$$

where $P$ is the positive definite matrix satisfying Stein's condition for $W$. This inequality guarantees the existence of the quadratic Lyapunov function $V(\boldsymbol{x}) = \boldsymbol{x}^t P \boldsymbol{x}$. Indeed we have

$$\begin{aligned}
\Delta V(\boldsymbol{x}) &= V(\boldsymbol{x}') - V(\boldsymbol{x}) \\
&= \boldsymbol{\phi}(\boldsymbol{x})^t W^t P W \boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{x}^t P \boldsymbol{x} \\
&\leq \boldsymbol{\phi}(\boldsymbol{x})^t W^t P W \boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{x})^t P \boldsymbol{\phi}(\boldsymbol{x}) \\
&= \boldsymbol{\phi}(\boldsymbol{x})^t \left(W^t P W - P\right) \boldsymbol{\phi}(\boldsymbol{x}) \leq 0,
\end{aligned}$$

where the equality holds only if $\phi(\boldsymbol{x}) = \boldsymbol{x} = 0$. These results hold for the related network $\boldsymbol{y}' = \phi(W\boldsymbol{y})$ with the same Lyapunov function as well [Kaszkurewicz and Bhaya, 1993]:

$$\Delta V(\boldsymbol{y}) = \phi(W\boldsymbol{y})^t P\phi(W\boldsymbol{y}) - \boldsymbol{y}^t P\boldsymbol{y}$$
$$\leq \boldsymbol{y}^t \left( W^t PW - P \right) \boldsymbol{y} \leq 0 \,.$$

In the following we want to generalise these results to the following systems with constant, but non-zero input $\boldsymbol{u}$

$$\boldsymbol{x} \mapsto W\phi(\boldsymbol{x}) + \boldsymbol{u} \qquad (4.6)$$
$$\boldsymbol{x} \mapsto \phi(W\boldsymbol{x} + \boldsymbol{u}) \qquad (4.7)$$

While the previously considered system (4.1) possesses the origin as its unique fixed point due to the property $\phi(0) = 0$, we now have to deal with a different fixed point $\bar{\boldsymbol{x}} \neq 0$. Using the coordinate transformation $\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{x} - \bar{\boldsymbol{x}}$ we can transform these systems into

$$\boldsymbol{y} \mapsto W \left( \phi(\boldsymbol{y} + \bar{\boldsymbol{x}}) - \phi(\bar{\boldsymbol{x}}) \right) =: W\boldsymbol{\psi}(\boldsymbol{y}) \qquad (4.6')$$
$$\boldsymbol{y} \mapsto \phi(W\boldsymbol{y} + W\bar{\boldsymbol{x}} + \boldsymbol{u}) - \phi(W\bar{\boldsymbol{x}} + \boldsymbol{u}) =: \boldsymbol{\psi}(W\boldsymbol{y}) \,, \qquad (4.7')$$

which obviously have the origin has a fixed point again (see fig 4.1). The main issue to proof gas of the transformed system – with activation functions $\boldsymbol{\psi}(\boldsymbol{y})$ – is under which conditions $\psi_i$ inherits the properties of $\phi_i$ with respect to the membership in the function classes $\mathcal{OL}$ and $\mathcal{S}$. Clearly oddness is lost in almost all cases[1], so we have to rely on the more general theorem 4.1. The sector condition (4.3) for $\psi$ reads

$$0 \leq \frac{|\psi(y)|}{|y|} = \frac{|\phi(y + \bar{x}) - \phi(\bar{x})|}{|y|} \leq 1 \quad \text{for all } y \in \mathbb{R}.$$

To hold for an arbitrary input $\boldsymbol{u}$ and thus arbitrary $\bar{\boldsymbol{x}} \in X$ this is equivalent to the global Lipschitz condition

$$0 \leq \frac{|\phi(x) - \phi(y)|}{|x - y|} \leq 1 \quad \text{for all } x \neq y \in R. \qquad (4.8)$$

Thus we obtain the following theorem [Hu and Wang, 2002]:

**Theorem 4.3** The dynamical systems (4.6) and (4.7) are globally asymptotically stable for all $\phi_i(x)$ satisfying condition (4.8) and for all inputs $\boldsymbol{u} \in \mathbb{R}^n$, if $W$ is diagonally Schur stable.

**Remark:** Condition (4.8) is also known as *incremental sector condition*. If $\phi$ is differentiable for all $x$ this is equivalent to the condition $0 \leq |\phi'(x)| \leq 1$ for all $x \in R$.

Theorem 4.3 generalises the result of Bhaya et al. [1996] to non-differentiable and unbounded nonlinearities. Hu and Wang present even more general results for recurrent neural networks of the form

$$\boldsymbol{x}' = A\boldsymbol{x} + B\boldsymbol{\sigma}(W\boldsymbol{x} + u)$$

where $A$ and $B$ are diagonal matrices, $W$ is an arbitrary matrix and $\sigma_i$ are monotone increasing activation functions satisfying (4.8). Absolute stability of this system is equivalent to absolute stability of the linear time-varying system

$$\boldsymbol{y}(t + 1) = [A + BD(t)W] \boldsymbol{y}(t) =: M(t)\boldsymbol{y}(t) \quad \text{with } D(t) \in [-\boldsymbol{1}, \boldsymbol{1}], \qquad (4.9)$$

---

[1] two exceptions are for example $\phi(x) = x$ and $\phi(x) = \sin(x)$

which can be proven, if the matrix polytope $[A - BW, A + BW]$ is *simultaneously* Schur stable, which is easily shown for diagonally stable $A + BW$ (see appendix B for details). Finding a non-diagonal Lyapunov function for this problem is much more difficult and was considered in many previous works [Bauer et al., 1993; Molchanov and Liu, 2002, and references therein]. There exist necessary and sufficient conditions for (4.9) to be globally asymptotically stable, which rely on matrix products of arbitrary length of vertex matrices of the matrix polytope [Molchanov and Liu, 2002]. These conditions imply NP-hard problems in general and hence are difficult to verify [Rohn, 1994a; Nemirovskii, 1993].

**Remark:** As was pointed out by Barabanov and Prokhorov [2002] all modern methods to prove absolute stability, do not utilise properties of a specific nonlinearity $\sigma$, but rather infer stability for a large class of nonlinearities bounded by a linear time-varying system (4.9). They argue, that employing properties of the nonlinearities should improve inferred stability bounds – much like nonlinearities make nonlinear systems functionally superior to linear systems.

But actually, a necessary condition for absolute stability of an arbitrary nonlinear system is local stability of all possible fixed points. As we have seen, if a fixed point is hyperbolic, the linearisation at this fixed point determines its local stability. Only, if it is nonhyperbolic, higher order terms have to be taken into account. For RNNs of type (4.6) or (4.7) it is always possible to find an input vector $u$, such that an arbitrary point in state space becomes a fixed point of the network – simply use $u = \bar{x} - W\phi(\bar{x})$ or $u = \phi^{-1}(\bar{x}) - W\bar{x}$ respectively. Thus, if we regard absolute stability, we have to consider all possible linearisations of the nonlinear system, which leads naturally to sector-based methods.

Nevertheless certain properties of the nonlinear activation functions, influence the stability of a specific network and might improve stability results for these specific systems. Hu and Lin [2001] give an example for a BSB model having a Schur stable connectivity matrix, which is not globally asymptotically stable. But, if the piecewise linear activation function is replaced by a smooth sigmoid function, for example the hyperbolic tangent, it becomes globally asymptotically stable, even though both activations functions have the same maximal gain.

# 5 Equivalence Classes of RNNs

If we are interested in qualitative properties of RNNs it is important to identify classes of networks, which have similar phase portraits, i.e. which are (globally) topologically equivalent. This allows us to study the behaviour of a single, e.g. an especially simple member of this class in detail and to draw conclusions for all other members subsequently. In the following we will study two different types of these equivalence classes: First we consider classes of equivalent activation functions and second we consider classes of equivalent parameter sets, i.e. weights and inputs. In both cases the resulting phase portraits are affine transformations of the original phase portraits.

## 5.1 Classes of activation functions

Tiňo et al. [2001] introduced the following class $\mathcal{S}_0$ of sigmoid activation functions:

$$\sigma_{\alpha,\beta,\mu}(x) = \frac{\alpha}{1 + \mathrm{e}^{-\mu x}} + \beta \quad \alpha, \mu \in \mathbb{R}_+, \beta \in \mathbb{R} \tag{5.1}$$

mapping $\mathbb{R}$ into the interval $[\beta, \alpha + \beta]$. This function class includes two commonly used transfer functions: the Fermi function ($\sigma_{1,0,1}$) and the hyperbolic tangent ($\sigma_{2,-1,2}$). Every function within this class can be represented by the hyperbolic tangent using the affine transformation

$$\sigma_{\alpha,\beta,\mu}(x) = \tfrac{1}{2}\alpha \tanh(\tfrac{1}{2}\mu x) + \tfrac{1}{2}\alpha + \beta \,. \tag{5.2}$$

In the following theorem we show, that this equation results in an interrelation of transfer function parameters $[\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}]$ and network parameters $[W, \boldsymbol{u}]$.

**Theorem 5.1** Consider the RNNs (2.11) and (2.12), which might have different activation functions $\sigma_i = \sigma_{\alpha_i,\beta_i,\mu_i}$ for each neuron $i$. We summarise their parameters within diagonal matrices, which are denoted by $\boldsymbol{\mu} = \mathrm{diag}(\mu_i)$, $\boldsymbol{\alpha} = \mathrm{diag}(\alpha_i)$, and $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_n]^t$. Further we define the vector $\boldsymbol{c} = [\tfrac{1}{2}\alpha_1 + \beta_1, \ldots, \tfrac{1}{2}\alpha_n + \beta_n]^t$. Each RNN of type (2.11) or (2.12) is globally topologically conjugate to

$$\boldsymbol{x}'(t+1) = W' \tanh(\boldsymbol{x}'(t)) + \boldsymbol{u}' \qquad \text{and} \qquad \boldsymbol{y}'(t+1) = \tanh\left(W'\,\boldsymbol{y}'(t) + \boldsymbol{u}'\right)$$

respectively, where the parameter homeomorphisms and the parameter-dependent state homeomorphisms are defined by:

$$W' = \tfrac{1}{4}\boldsymbol{\mu} W \boldsymbol{\alpha} \tag{5.3}$$

$$\boldsymbol{u}' = \tfrac{1}{2}\boldsymbol{\mu}\left(W\boldsymbol{c} + \boldsymbol{u}\right) \tag{5.4}$$

$$\boldsymbol{x}'(t) = \tfrac{1}{2}\boldsymbol{\mu}\boldsymbol{x}(t) \qquad \text{and} \qquad \boldsymbol{y}'(t) = 2\boldsymbol{\alpha}^{-1}\left(\boldsymbol{y}(t) - \boldsymbol{c}\right) \,. \tag{5.5}$$

**Proof**  If we consider an RNN of type (2.12) and employ equation (5.2) we obtain:

$$\boldsymbol{y}(t+1) = \boldsymbol{\sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}\left(W\boldsymbol{y}(t) + \boldsymbol{u}\right)$$
$$= \tfrac{1}{2}\boldsymbol{\alpha} \tanh\left(\tfrac{1}{2}\boldsymbol{\mu}\,W\boldsymbol{y}(t) + \tfrac{1}{2}\boldsymbol{\mu}\boldsymbol{u}\right) + \boldsymbol{c}\,.$$

Applying the coordinate transformation $\boldsymbol{y}(t) = \frac{1}{2}\boldsymbol{\alpha}\boldsymbol{y}'(t) + \boldsymbol{c}$ finally yields:

$$\boldsymbol{y}'(t+1) = \tanh\left(W'\,\boldsymbol{y}'(t) + \boldsymbol{u}'\right)\;.$$

The other network type (2.11) can be handled analogously and results in the state homeomorphism $\boldsymbol{x}'(t) = \frac{1}{2}\boldsymbol{\mu}\boldsymbol{x}(t)$. Notice, that we obtain the same parameter transformations in both cases, but different state transformations. Actually, the parameters $W$ and $\boldsymbol{u}$ define the properties of the RNN and have to be identical for both network types, which have identical dynamics (compare example 2.11). $\square$

According to this theorem, each RNN which has transfer functions $\sigma_{\alpha_i,\beta_i,\mu_i}$, weight matrix $W$ and input vector $\boldsymbol{u}$ can be translated into an equivalent RNN, which uses the hyperbolic tangent as the transfer function for all neurons and which has $W'$ resp. $\boldsymbol{u}'$ as transformed parameter sets. Conversely for each RNN with the hyperbolic tangent as transfer function and parameter sets $W'$, $\boldsymbol{u}'$ we can find a parameter set $[W, \boldsymbol{u}] \in \mathcal{Q} = \mathbb{R}^{n \times n} \times \mathbb{R}^n$ such that a corresponding RNN with arbitrary other activation functions within $\mathcal{S}_0$ has the same dynamical behaviour. Actually their trajectories differ by an affine transformation only.

Because this correspondence between transfer function parameters $[\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}]$ and network parameters $[W, \boldsymbol{u}]$ exists, it suffices to study RNNs with the hyperbolic tangent as transfer function for all neurons. This choice is particularly attractive due to the high symmetry of the hyperbolic tangent. If the correct parameter set $[W, \boldsymbol{u}]$ to achieve a specific dynamical behaviour has been found, the system can be easily transformed into another system, if the activation functions have to be changed, for example to achieve a scaling of the neuron activities according to (5.5). Furthermore we can conclude that all relevant parameters are included already in the weight matrix $W$ and the input vector $\boldsymbol{u}$. The transfer function does not need to be parametrisable too. Consequently we will drop the parameter subscripts and write $\sigma$ instead of $\sigma_{\alpha,\beta,\mu}$ for an arbitrary function within $\mathcal{S}_0$.

We can rewrite the transformations (5.3) – (5.5) in the framework of topologically conjugate parameter-dependent dynamical systems (definition 3.2), which results in the following commutative diagrams:

$$
\begin{array}{ccccc}
\mathcal{Q} & & X_1 & \xrightarrow{\;\varphi = W\,\boldsymbol{\sigma}_{\alpha,\beta,\mu}(\boldsymbol{x}) + \boldsymbol{u}\;} & X_1 \\
p \downarrow & & h_{\boldsymbol{\mu}} \downarrow & & \downarrow h_{\boldsymbol{\mu}} \\
\mathcal{Q} & & X_2 & \xrightarrow[\;\psi = W'\,\tanh(\boldsymbol{x}') + \boldsymbol{u}'\;]{} & X_2
\end{array}
\qquad
\begin{array}{ccc}
X_1 & \xrightarrow{\;\varphi = \boldsymbol{\sigma}_{\alpha,\beta,\mu}(W\,\boldsymbol{y} + \boldsymbol{u})\;} & X_1 \\
h_{\alpha,\beta} \downarrow & & \downarrow h_{\alpha,\beta} \\
X_2 & \xrightarrow[\;\psi = \tanh(W'\,\boldsymbol{y}' + \boldsymbol{u}')\;]{} & X_2
\end{array}
$$

with homeomorphisms

$$p(W, \boldsymbol{u}) = (W', \boldsymbol{u}') = \left(\tfrac{1}{4}\,\boldsymbol{\mu}\,W\,\boldsymbol{\alpha}, \tfrac{1}{2}\,\boldsymbol{\mu}(W\boldsymbol{c} + \boldsymbol{u})\right)$$

$$h_{\boldsymbol{\mu}}(\boldsymbol{x}) = \tfrac{1}{2}\,\boldsymbol{\mu}\,\boldsymbol{x} \qquad \text{and} \qquad h_{\alpha,\beta}(\boldsymbol{y}) = 2\boldsymbol{\alpha}^{-1}(\boldsymbol{y} - \boldsymbol{c})\;.$$

## Properties of functions in $\mathcal{S}_0$

The meaning of the parameters $\alpha$, $\beta$ and $\mu$ of $\sigma_{\alpha,\beta,\mu}$ can be interpreted as follows: While $\alpha$ defines the length of the image interval $[\beta, \alpha + \beta]$, $\beta$ controls its offset. The maximal steepness or gain of the function is achieved at $x = 0$ and equals to $\frac{1}{4}\mu\alpha$. All functions within $\mathcal{S}_0$ have a sigmoidal shape, i.e. they are strict monotonically increasing, saturate to $\beta$ resp. $\alpha + \beta$ as $x$ approaches $\mp\infty$, and their derivatives have a single isolated maximum at $x = 0$. As can be seen from fig. 5.1a the

$u_1$
$u_2$
$\psi_1$
$\psi_2$

$\frac{1}{2}\alpha + \beta$

$0$

$\frac{1}{4}\mu\alpha$

$\psi(x)$

$\alpha + \beta$

$C$

$\beta$

$u_1$
$u_2$
$\psi_1$
$\psi_2$

$\frac{1}{4}\mu\alpha$

$\psi(x)$

$0$

$C\beta$ $\qquad \frac{1}{2}\alpha + \beta$ $\qquad \alpha + \beta$

(a) graph of $\sigma_{\alpha,\beta,\mu}$ $\qquad$ (b) graph of $\psi_{\alpha,\beta,\mu}$
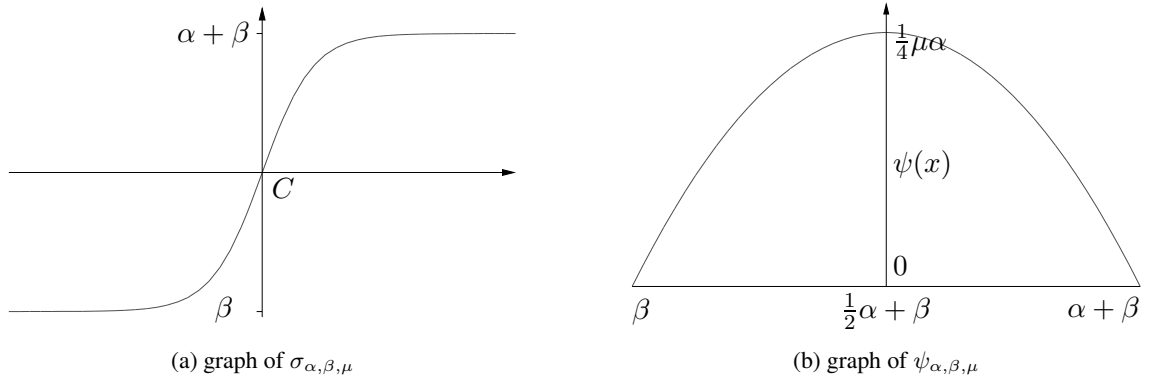
Figure 5.1: Graph of hyperbolic tangent $\in \mathcal{S}_0$ together with its corresponding $\psi$ function.

graph of $\sigma_{\alpha,\beta,\mu}$ has a special point at $C = [0, \frac{1}{2}\alpha + \beta]$, where the function reaches its maximal gradient $\frac{1}{4}\mu\alpha$. Furthermore the graph is point-symmetric with respect to $C$, i.e.

$$\sigma_{\alpha,\beta,\mu}(x) - c = -\sigma_{\alpha,\beta,\mu}(-x) + c \qquad \text{where} \quad c = \tfrac{1}{2}\alpha + \beta\,. \tag{5.6}$$

All functions within $\mathcal{S}_0$ are diffeomorph with the derivative

$$\sigma'_{\alpha,\beta,\mu} = \psi_{\alpha,\beta,\mu} \circ \sigma_{\alpha,\beta,\mu} \tag{5.7}$$

$$\text{where} \quad \psi_{\alpha,\beta,\mu}(x) = \frac{\mu}{\alpha}(x - \beta)(\alpha + \beta - x)\,. \tag{5.8}$$

Thus, the derivative of $\sigma$ can be expressed as a quadratic function of $\sigma$ itself. In section 6 we will employ this property to simplify our analytical expressions for bifurcation manifolds of RNNs. More concretely, we have to evaluate the derivative of the iterative map of a RNN (2.12) at its fixed point $\bar{x}$, which comprises the derivatives $\boldsymbol{\sigma'}$ for which we obtain:

$$\bar{\boldsymbol{x}} = \boldsymbol{\sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}(W\bar{\boldsymbol{x}} + \boldsymbol{u}) \tag{5.9}$$

$$\boldsymbol{\sigma'}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}(W\bar{\boldsymbol{x}} + \boldsymbol{u}) \overset{(5.7)}{=} \boldsymbol{\psi}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}\big(\boldsymbol{\sigma}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}(W\bar{\boldsymbol{x}} + \boldsymbol{u})\big) \overset{(5.9)}{=} \boldsymbol{\psi}_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\mu}}(\bar{\boldsymbol{x}})\,. \tag{5.10}$$

Obviously, the derivatives $\boldsymbol{\sigma'}$ can be computed component-wise as a simple nonlinear (quadratic) function of the *fixed point* only – neither the weight matrix nor the input is involved directly. Consequently, for each value $\psi$ exist two preimages $\bar{x}_\pm$ under the derivative $\sigma'_{\alpha,\beta,\mu}$ (compare fig. 5.1b):

$$x_\pm = \psi_{\alpha,\beta,\mu}^{-1}(\psi) = c \pm \sqrt{\frac{\alpha^2}{4} - \frac{\alpha}{\mu}\psi} \qquad \text{for all } \psi \in [0, \tfrac{1}{4}\mu\alpha]\,. \tag{5.11}$$

An exception constitutes the maximal value of the gain, where only a single preimage exists.

## 5.2 Classes of parameters

The symmetry (5.6) of functions in $\mathcal{S}_0$ is the source for another equivalence class, which is introduced by appropriate reflections of the neuron activities $\boldsymbol{x}$.

**Theorem 5.2** Let $S = \text{diag}(s_i)$ be a diagonal matrix with elements $s_i = \pm 1$, which defines a multidimensional reflection about the origin. The corresponding reflection about the symmetry point $\boldsymbol{c}$ in state space is defined by:

$$T(\boldsymbol{y}) = S(\boldsymbol{y} - \boldsymbol{c}) + \boldsymbol{c} \tag{5.12}$$

According to its definition in theorem 5.1 the symmetry vector $c$ depends on the choice of activation functions $\sigma_i$. Each RNN of type (2.11) or (2.12) is globally topologically conjugate to

$$\boldsymbol{x}'(t+1) = W' \tanh(\boldsymbol{x}'(t)) + \boldsymbol{u}' \qquad \text{and} \qquad \boldsymbol{y}'(t+1) = \tanh\left(W' \boldsymbol{y}'(t) + \boldsymbol{u}'\right)$$

respectively, where the parameter and state homeomorphisms are defined by:

$$W' = S W S \tag{5.13}$$

$$\boldsymbol{u}' = S(\boldsymbol{u} + W\boldsymbol{c}) - S W S \boldsymbol{c} \tag{5.14}$$

$$\boldsymbol{x}'(t) = S\boldsymbol{x}(t) \qquad \text{and} \qquad \boldsymbol{y}'(t) = T\boldsymbol{y}(t) \,. \tag{5.15}$$

The corresponding commutative diagrams have the form:

$$
\begin{array}{ccc}
\mathcal{Q} & X_1 \xrightarrow{\varphi = W\,\boldsymbol{\sigma}(\boldsymbol{x}) + \boldsymbol{u}} X_1 & X_1 \xrightarrow{\varphi = \boldsymbol{\sigma}(W\,\boldsymbol{y} + \boldsymbol{u})} X_1 \\
p \downarrow \quad & S \downarrow \qquad\qquad \downarrow S & T \downarrow \qquad\qquad \downarrow T \\
\mathcal{Q} & X_2 \xrightarrow[\psi = W'\,\boldsymbol{\sigma}(\boldsymbol{x}') + \boldsymbol{u}']{} X_2 & X_2 \xrightarrow[\psi = \boldsymbol{\sigma}(W'\,\boldsymbol{y}' + \boldsymbol{u}')]{} X_2
\end{array}
$$

**Proof**    The proof is based on the multidimensional version of equation (5.6), which yields

$$\boldsymbol{\sigma} \circ S = T \circ \boldsymbol{\sigma} \,.$$

Employing the identities $S = S^{-1}$ and $T = T^{-1}$, which are typical for reflection operators, we obtain for a RNN of type (2.12):

$$\boldsymbol{\sigma}(W'\boldsymbol{y}' + \boldsymbol{u}') = T\left(\boldsymbol{\sigma}(W[T^{-1}\boldsymbol{y}'] + \boldsymbol{u})\right) \tag{5.16}$$

$$= \boldsymbol{\sigma}\left(S W[S(\boldsymbol{y}' - \boldsymbol{c}) + \boldsymbol{c}] + S\boldsymbol{u}\right) \tag{5.17}$$

$$= \boldsymbol{\sigma}\left(S W S\boldsymbol{y}' + S(\boldsymbol{u} + W\boldsymbol{c}) - S W S\boldsymbol{c}\right) \,. \tag{5.18}$$

By comparison of the parameters on both sides we obtain the weight matrix $W'$ and the input vector $\boldsymbol{u}'$. The proof for the network type (2.11) can be done analogously. $\square$

A similar[1] equivalence class of network parameters was already considered by Pasemann [2002] for networks having the Fermi function $\sigma_{101}$ as their activation function. The theorem presented here generalises Pasemann's result to a broad class of activation functions. As can be seen from (5.15) the symmetry point in state space is given by zero or $c$, depending on whether type (2.11) or type (2.12) of the network dynamics is used. The transformation of the input vector (5.14) cannot be written in the form (5.12) in general. Consequently there exists no symmetry point in input space for these reflections.[2]

The theorem states the equivalence of $2^n$ parameter sets $[W, \boldsymbol{u}]$ for networks with *fixed* activation functions. The effect of applying the weight transform $W' = S W S$ is swapping the signs of all columns and rows for which $s_i = -1$ is satisfied, i.e. $w'_{ij} = s_i s_j w_{ij}$. Together with an appropriate input vector, whose transformed form becomes most simple if $c = 0$, the resulting network exhibits equivalent dynamical behaviour with respect to the original RNN.

---

[1]  Pasemann studied sign changes of single state components, i.e. within our notation a single $s_i = -1$. Of course the homeomorphism $S$ or $T$ considered here can be composed of multiple applications of Pasemann's operator.

[2]  An exception, which holds for arbitrary weight matrices, is $S = -1$. In this case the symmetry point in input space is given by $-W\boldsymbol{c}$. If the vector $\boldsymbol{c}$ equals zero, e.g. for the hyperbolic tangent, the symmetry point in input space is the origin – also independently of the chosen reflection operator $S$.

# 6 Bifurcation Manifolds in Input Space

In the present chapter we employ results from bifurcation theory to derive analytical expressions for bifurcation manifolds of local fixed point bifurcations in fully connected RNNs, which have arbitrary activation functions $\sigma$. This chapter constitutes the main contribution of this work. Although many authors studied bifurcation manifolds of discrete- and continuous-time neural networks before, they employed numerical methods only and often restricted their analysis to simplified connection matrices. For example Borisyuk and Kirillov [1992] have done a detailed bifurcation analysis of the Wilson-Cowan oscillator using continuation techniques[1] to compute bifurcation curves. Beer [1995] studied selected bifurcation curves of continuous-time two-neuron networks and Pasemann determined bifurcation curves of various discrete-time RNNs by simulation of the underlying dynamics [Pasemann, 1993b,a, 1995, 1997, 1999, 2002]. The first attempt to compute bifurcation manifolds of RNNs analytically was made by Hoppensteadt and Izhikevich [1997], who employed specific properties of the Fermi function to derive bifurcation curves of continuous-time networks.

In contrary, our approach is applicable to networks with arbitrary activation functions, though the derived expressions become especially simple for functions within the class $\mathcal{S}_0$. In addition our approach allows to compute bifurcation manifolds of arbitrary dimension, while continuation techniques are restricted to bifurcation curves, i.e. one-dimensional manifolds. This enables us to compute codim-1 bifurcations, i.e. saddle-node, period-doubling, and Neimark-Sacker bifurcations, within a parameter space of more than two dimensions. Particularly we will study bifurcation manifolds of three-neuron networks in chapter 8.

## 6.1 The general approach

We consider discrete-time recurrent neural networks of the type (2.12), i.e.

$$\boldsymbol{x} \mapsto \boldsymbol{\sigma}(W\boldsymbol{x} + \boldsymbol{u}) \qquad \boldsymbol{x} \in \mathbb{R}^n \ . \tag{6.1}$$

In order to study their bifurcation manifolds in input space, we assume, that the network's weights are constant or at least varying on a slow time scale with respect to the time scale of the network inputs. The inputs themselves are treated as bifurcation parameters in the $n$-dimensional parameter space. Thus for fixed connection weights, the input space is decomposed into regions of different, i.e. topologically non-equivalent, dynamical behaviour. Knowledge of the corresponding bifurcation boundaries allows to switch between different dynamical regimes, for example steady states and oscillations, by choosing inputs within the corresponding parameter region.

The bifurcation boundaries corresponding to codim-1 bifurcations of a fixed point are defined by the fixed point condition

$$\bar{\boldsymbol{x}} = \boldsymbol{\sigma}\Big(W\bar{\boldsymbol{x}} + \boldsymbol{u}\Big) \tag{6.2}$$

---

[1] Continuation methods numerically compute a curve, which is implicitly defined by a set of equations. An introduction to continuation techniques in bifurcation theory is given by Kuznetsov [1995], who is the author of the continuation package CONTENT as well [Kuznetsov and Levitin].

| bifurcation type | eigenvalue condition | necessary test condition |
|---|---|---|
| saddle node | $\lambda = +1$ | $\det(J(\bar{\boldsymbol{x}}) - \mathbf{1}) = 0$ |
| period doubling | $\lambda = -1$ | $\det(J(\bar{\boldsymbol{x}}) + \mathbf{1}) = 0$ |
| Neimark-Sacker | $\lambda_{1,2} = \mathrm{e}^{\pm\mathrm{i}\omega}$ $\mathrm{e}^{\pm\mathrm{i}\omega\cdot k} \neq 1$ for $k = 1, 2, 3, 4$ | $\det(J(\bar{\boldsymbol{x}}) \odot J(\bar{\boldsymbol{x}}) - \mathbf{1}) = 0$ |

Table 6.1: Bifurcation conditions for codim-1 bifurcation in discrete-time dynamical systems

– which actually represents $n$ independent nonlinear equations – and a further condition, which differs with respect to different bifurcation types. The resulting system of $n+1$ nonlinear equations with $2n$ variables ($\bar{\boldsymbol{x}}$ and $\boldsymbol{u}$) defines a $n-1$-dimensional manifold in the direct product of the state and the input space. Its projection to the input space defines potential bifurcation boundaries – potential boundaries, because the employed test functions represent necessary conditions for the occurrence of the corresponding bifurcations only. They do not take into account their associated genericity conditions. Hence, it has to be verified subsequently, that the obtained bifurcation manifolds indeed represent bifurcations of the given type at all points on the manifold.

All test functions, which we derived in section 3.2 for all considered bifurcation types, are summarised for convenience in table 6.1. They have the form of determinant equations involving the Jacobian matrix $J(\bar{\boldsymbol{x}})$ at an unknown fixed point $\bar{\boldsymbol{x}}$. This Jacobian has the following simple form:

$$J(\bar{\boldsymbol{x}}) = D(\bar{\boldsymbol{x}})\,W$$
$$\text{where} \quad D(\bar{\boldsymbol{x}}) = \mathrm{diag}(\boldsymbol{\sigma}'(W\bar{\boldsymbol{x}} + \boldsymbol{u}))\,.$$

The key idea to find analytical expressions for the bifurcation manifolds of networks with arbitrary activation functions $\boldsymbol{\sigma}$ is to regard the derivatives

$$\boldsymbol{\psi} := \boldsymbol{\sigma}'(W\bar{\boldsymbol{x}} + \boldsymbol{u})$$

as free parameters initially, instead of the input coordinates $\boldsymbol{u}$ themselves. Then the bifurcation conditions (table 6.1) are expressions in terms of the weight matrix $W$ and the variables $\boldsymbol{\psi}$ only. These equations are algebraically solvable often and their solutions define $n-1$-dimensional manifolds in $\boldsymbol{\psi}$-space. It remains to find a corresponding fixed point in state space and a bifurcation point $\boldsymbol{u}$ in input space for each point $\boldsymbol{\psi}$ on this solution manifold. Because the derivatives have a limited range, say $\sigma_i' \in [\alpha_i, \beta_i]$, only solutions within the hypercube

$$\mathcal{H} = \prod_{i=1}^{n} [\alpha_i, \beta_i] \subseteq \mathbb{R}^n \tag{6.3}$$

are practical solutions, for which fixed points $\bar{\boldsymbol{x}}$ and associated bifurcation points $\boldsymbol{u}$ can exist. In fact, this limited range of feasible derivatives reflects an incremental sector condition (compare chapter 4), which is met by the chosen set of activation functions $\boldsymbol{\sigma}$. In order to find the pair $[\bar{\boldsymbol{x}}, \boldsymbol{u}]$ for a given bifurcation point in $\boldsymbol{\psi}$-space, we have to solve the equations

$$\boldsymbol{\sigma}'(W\bar{\boldsymbol{x}} + \boldsymbol{u}) = \boldsymbol{\psi} \tag{6.4}$$
$$\text{and} \quad \boldsymbol{\sigma}(W\bar{\boldsymbol{x}} + \boldsymbol{u}) = \bar{\boldsymbol{x}} \tag{6.5}$$

which formally have the solution

$$\bar{\boldsymbol{x}} = \boldsymbol{\sigma}(\boldsymbol{\xi}) \quad \text{for all} \quad \boldsymbol{\xi} \in Z(\boldsymbol{\psi}) \tag{6.6}$$
$$\boldsymbol{u} = \boldsymbol{\xi} - W\bar{\boldsymbol{x}} = \boldsymbol{\xi} - W\boldsymbol{\sigma}(\boldsymbol{\xi}) \tag{6.7}$$

$$\text{where} \quad Z(\boldsymbol{\psi}) = \boldsymbol{\sigma}'^{-1}(\boldsymbol{\psi}) = \{\boldsymbol{\xi} \,|\, \boldsymbol{\sigma}'(\boldsymbol{\xi}) = \boldsymbol{\psi}\} \tag{6.8}$$
$$\text{such that} \quad \boldsymbol{\xi} \equiv W\bar{\boldsymbol{x}} + \boldsymbol{u}\,. \tag{6.9}$$

Here we denote with $Z(\boldsymbol{\psi})$ the set of preimages of $\boldsymbol{\sigma}'$ for a given $\boldsymbol{\psi}$. Similar expressions can be obtained for RNNs of type (2.11) as well. We summarise the results for both network types in the following theorem:

**Theorem 6.1** Let be given a recurrent neural network of type (2.11) or (2.12) with activation functions $\sigma_i$, which satisfy an incremental sector condition each. The corresponding hypercube of feasible derivatives is denoted by $\mathcal{H}$. Then for each $\boldsymbol{\psi} \in \mathcal{H} \subset \mathbb{R}^n$ exists at least one input vector $\boldsymbol{u}$, such that the network exhibits a fixed point $\bar{\boldsymbol{x}}$, for which holds

$$\boldsymbol{\psi} = \boldsymbol{\sigma}'(\bar{\boldsymbol{x}}) \qquad \text{for RNNs of type (2.11)} \qquad (6.10)$$

$$\text{or} \quad \boldsymbol{\psi} = \boldsymbol{\sigma}'(W\bar{\boldsymbol{x}} + \boldsymbol{u}) \qquad \text{for RNNs of type (2.12).} \qquad (6.11)$$

The input vector(s) and the corresponding fixed point(s) are given by:

$$\boldsymbol{u} = \boldsymbol{\xi} - W\boldsymbol{\sigma}(\boldsymbol{\xi}), \qquad \boldsymbol{\xi} \in \boldsymbol{\sigma}'^{-1}(\boldsymbol{\psi}) \qquad (6.12)$$

and $\bar{\boldsymbol{x}} = \boldsymbol{\xi}$ or $\bar{\boldsymbol{x}} = \boldsymbol{\sigma}(\boldsymbol{\xi})$ in dependence of the network type.

Typical sigmoid activations functions have two preimages for each admissible $\psi_i$ (compare figure 5.1b), but in principle even more preimages are possible. The number of possible preimages determines the number of different branches of the bifurcation manifold in input space corresponding to a single bifurcation point in $\boldsymbol{\psi}$-space. Thus, for the typical activation functions we obtain up to $2^n$ different branches of a bifurcation manifold in input space. As we will see in the detailed analysis below, the solution manifolds in $\boldsymbol{\psi}$-space split into several branches as well, and their number increases exponentially with $n$. Thus the numerical computation of these solution manifolds becomes very complicated for high-dimensional RNNs. Furthermore a visualisation of the bifurcation manifolds in the $n$-dimensional embedding space requires projections to a three-dimensional space if $n > 3$. For this reason we first concentrate on simple RNNs with two or three neurons.

Notice, that the bifurcation manifolds in $\boldsymbol{\psi}$-space are *independent* of the activation functions $\boldsymbol{\sigma}$. Only the range of their derivatives restricts the region in $\boldsymbol{\psi}$-space corresponding to feasible solutions. Apart from that all activation functions possess the same bifurcation manifolds in $\boldsymbol{\psi}$-space, which only depend on the weight matrix. Thus this space is particularly qualified to study bifurcations of fixed points. Actually Tiňo et al. [2001] investigated the stability of fixed points of RNNs of type (2.12) with activation functions within $\mathcal{S}_0$ by a partition of this $\boldsymbol{\psi}$-space.

Especially interesting is the fact that there is a direct relation between fixed points $\bar{\boldsymbol{x}}$ and their corresponding derivatives $\boldsymbol{\psi}$, which even holds component-wise. Actually we have

$$\bar{\boldsymbol{x}} = \boldsymbol{\sigma}'^{-1}(\boldsymbol{\psi}) \qquad \text{or} \qquad \bar{\boldsymbol{x}} = \boldsymbol{\sigma}(\boldsymbol{\sigma}'^{-1}(\boldsymbol{\psi})) \qquad (6.13)$$

for networks of type (2.11) and (2.12) respectively. Of course this relation is not a one-to-one mapping, because both, $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ might have multiple preimages – but their number is finite in general. Thus for each given $\bar{\boldsymbol{x}}$ we can find a finite number of corresponding $\boldsymbol{\psi}$-values and vice versa. Notice that, the fixed point, which satisfies (6.10) or (6.11) for a given $\boldsymbol{\psi}$, is not unique. With the associated input vector (6.12) the RNN can exhibit rather several other stable and unstable fixed points or even more complex attractors.

Being able to compute bifurcation manifolds in input space from those in $\boldsymbol{\psi}$-space, it remains to derive analytical expressions for the latter. The next sections are concerned with this task. We consider saddle-node, period-doubling and Neimark-Sacker bifurcations of fixed points in the following.

## 6.2 Saddle-node bifurcation

The bifurcation condition of a saddle-node bifurcation ($\det(J - \mathbf{1}) = 0$) leads to a polynomial which is linear with respect to every $\psi_i$. Especially we have for two- and three-neuron networks:

$$n = 2 : \qquad\qquad 1 - \psi_1 w_{11} - \psi_2 w_{22} + \psi_1 \psi_2 \det W = 0 \qquad\qquad (6.14)$$

$$
n = 3 : \qquad
\begin{aligned}
& -1 + \psi_1 w_{11} + \psi_2 w_{22} + \psi_3 w_{33} \\
& \quad - [\psi_1 \psi_2 \det W_{(3)} + \psi_2 \psi_3 \det W_{(1)} + \psi_1 \psi_3 \det W_{(2)}] \\
& \quad + \psi_1 \psi_2 \psi_3 \det W = 0
\end{aligned}
\qquad (6.15)
$$

where $W_{(i)}$ denote the submatrices of $W$ obtained by deletion of $i$-th row and column. These equations have a solution if at least one coefficient of the occurring monoms is nonzero. Particularly $W$ has to be truly recurrent – upper or lower triangular matrices, corresponding to feed-forward networks obviously cannot produce any bifurcations. Without loss of generality we assume that we can solve for $\psi_1$ and obtain for:

$$n = 2 : \qquad\qquad \psi_1 = \frac{-1 + \psi_2 w_{22}}{-w_{11} + \det W \psi_2} \qquad\qquad (6.16)$$

$$n = 3 : \quad \psi_1 = \frac{+1 - \psi_2 w_{22} - \psi_3 w_{33} + \psi_2 \psi_3 (w_{22} w_{33} - w_{23} w_{32})}{w_{11} - \psi_2 (w_{11} w_{22} - w_{21} w_{12}) - \psi_3 (w_{11} w_{33} - w_{13} w_{31}) + \psi_2 \psi_3 \det W} \quad (6.17)$$

Whenever the denominator of these rational functions becomes zero, we observe a discontinuity of the solution manifold in $\psi$-space. Solving the denominator for its zeros leads to a second rational function, which defines a $n - 2$-dimensional discontinuity manifold in $\psi$-space separating two discontinuous branches of the bifurcation manifold. This second rational function again has singularities introduced by zeros of its denominator which can be solved for, leading to a third rational function. This procedure can be iterated until the denominator becomes constant. Altogether we obtain up to $2^{n-1}$ disjoint branches of the bifurcation manifold in $\psi$-space (see fig. 6.1a).

The handling of this exponentially growing number of branches is the main difficulty encountered if we try to compute bifurcation manifolds for networks of more than $n = 3$ neurons – formally more complex RNNs can be studied analogously. In fig. 6.1 the bifurcation manifolds in $\psi$-space and in input space are shown for a relatively simple weight matrix of a three neuron network with the hyperbolic tangent as activation function. For this reason we can restrict the analysis to the hypercube $\mathcal{H} = (0, 1]^3$ in $\psi$-space. Because the number of branches of the bifurcation manifold in input space is too large to visualise concisely, only the eight branches corresponding to the green lower left branch in $\psi$-space are drawn, each with a different colour. Already this small selection introduces a complex partition of the input space into different regions.

## 6.3 Cusp bifurcation

In order to determine the location of Cusp points on the saddle-node bifurcation manifold in input space we have to perform the center manifold reduction for each nonhyperbolic fixed point $\bar{x}$ corresponding to a bifurcation point on this manifold. We have a Cusp point at hand, if the second derivative of the system's mapping reduced to the center manifold equals zero. Due to the use of this higher order derivative, it is not possible to locate Cusp points on the bifurcation manifold in $\psi$-space directly, which was obtained from a condition on the first derivative only. Contrary, in input space Cusp points are characterised by sharp, i.e. nondifferentiable, points on the bifurcation
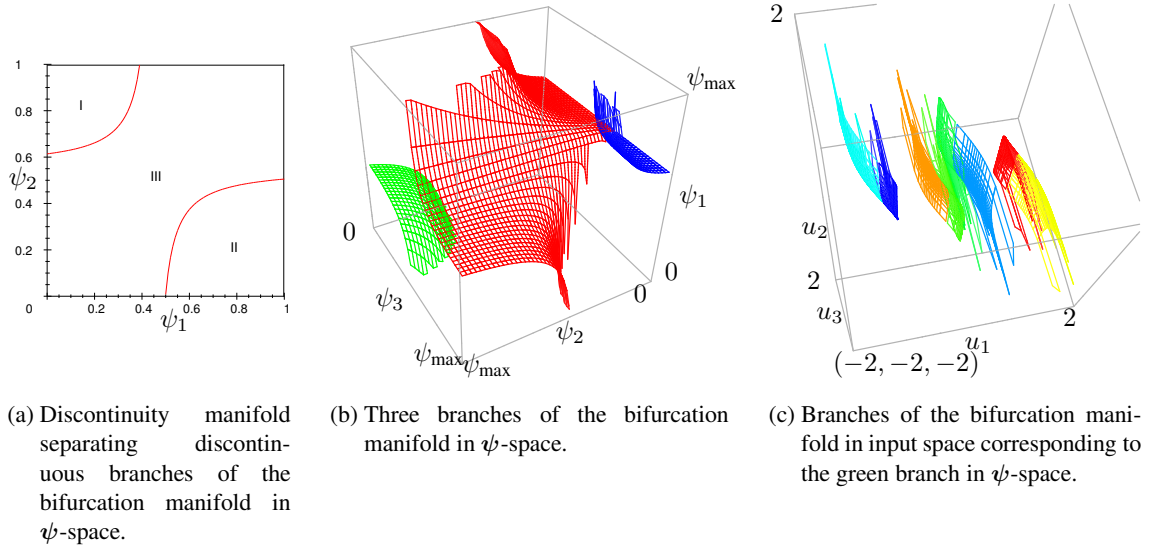
(a) Discontinuity manifold separating discontinuous branches of the bifurcation manifold in $\psi$-space.

(b) Three branches of the bifurcation manifold in $\psi$-space.

(c) Branches of the bifurcation manifold in input space corresponding to the green branch in $\psi$-space.

Figure 6.1: Saddle-node bifurcation manifolds for a RNN with $W = \left( \begin{smallmatrix} 1.61 & 0.78 & 0 \\ -0.78 & 1.63 & 0.59 \\ 0 & -0.59 & 1.63 \end{smallmatrix} \right)$ and $\sigma = \tanh$.

manifold.

In the following we want to derive an expression for the Cusp condition in RNNs of type (6.1). To this aim, we have to find the quadratic term of the Taylor series of the reduced system map. We use the projection method to perform the center manifold reduction, which is explained in a general form in Kuznetsov [1995]. Given a nonhyperbolic fixed point $\bar{x}$ with a single multiplier $\lambda = 1$ the system map in the vicinity of $\bar{x}$ can be written as

$$\bar{x} + \Delta x \equiv x \mapsto \sigma(Wx + u) = \bar{x} + D(\bar{x})W(x - \bar{x}) + \phi(x - \bar{x}) \tag{6.18}$$

where $\phi$ is a nonlinear smooth function with the Taylor expansion

$$\phi(\Delta x) = \tfrac{1}{2} B(\Delta x, \Delta x) + \mathcal{O}(\|\Delta x\|^3)$$
$$\text{where} \quad B_k(y, z) = \sigma_k''(w_k \bar{x} + u_k)\, y^t w_k^t w_k z \quad k = 1, \dots, n\,.$$

Here $w_k$ denotes the $k$-th row vector of the weight matrix $W$. The center manifold reduction is performed by decomposing $\Delta x$ into a part within the center eigenspace and an orthogonal part. To this aim we denote an eigenvector to the eigenvalue $\lambda = 1$ by $c$, such that it holds $J(\bar{x})\, c = c$. The projection to the center eigenspace, which is defined by $\{\alpha c \,|\, \alpha \in \mathbb{R}\}$, is performed by the adjoint eigenvector $p$, which satisfies $J^t(\bar{x})\, p = p$. For convenience we normalise $p$ with respect to $c$, such that $\langle p, c \rangle = 1$. With these definitions we obtain the decomposition

$$\Delta x = uc + v \qquad \text{with } u = \langle p, \Delta x \rangle \text{ and } v = \Delta x - \langle p, \Delta x \rangle c\,.$$

Rewriting (6.18) in terms of $[u, v]$ leads to the decomposed system

$$u \mapsto u + \langle p, \phi(uc + v) \rangle \tag{6.19}$$
$$v \mapsto D(\bar{x})Wv + \phi(uc + v) - \langle p, \phi(uc + v) \rangle c\,. \tag{6.20}$$

Developing the first of these equations into a Taylor series with respect to $u$, the system map, which is reduced with respect to the center manifold, yields:

$$u \mapsto u + \tfrac{1}{2} q u^2 + \mathcal{O}(u^3) \qquad \text{with } q = \langle p, B(c, c) \rangle\,. \tag{6.21}$$

The Cusp bifurcation condition now reads $q = \langle p, B(c, c) \rangle = 0$, which is a complicated expression in terms of $\psi$ and $W$.

**Example 6.2** In the following we derive this expression for the two-neuron network, and consequently obtain an implicit condition for a Cusp bifurcation on the saddle-node bifurcation manifold in $\boldsymbol{\psi}$-space, which can be solved subsequently with numerical methods. To this aim we have to perform the center manifold reduction and to evaluate the expression $q = \langle \boldsymbol{p}, \boldsymbol{B}(\boldsymbol{c}, \boldsymbol{c}) \rangle = 0$ in terms of $\boldsymbol{\psi}$ and $W$. First of all, we have to find an eigenvector and its adjoint to an eigenvalue $\lambda = 1$ of the Jacobian $J(\boldsymbol{\psi})$. The condition to have an eigenvalue $\lambda = 1$ is given by equation (6.14). Without loss of generality we choose the following projection vectors:

$$\boldsymbol{c} = \begin{pmatrix} \psi_2 w_{22} - 1 \\ -\psi_2 w_{21} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{p} = \begin{pmatrix} \psi_2 w_{22} - 1 \\ -\psi_1 w_{12} \end{pmatrix} . \tag{6.22}$$

Employing (6.14) it can be easily verified that these vectors are eigenvectors of $J(\boldsymbol{\psi})$ and $J^t(\boldsymbol{\psi})$ respectively. They do not satisfy the normalisation condition $\langle \boldsymbol{p}, \boldsymbol{c} \rangle = 1$, because the resulting constant factor can be dropped if the equality $q = 0$ is considered.

The second derivatives $\sigma''(\boldsymbol{w}_k \bar{\boldsymbol{x}} + u_k)$, which we denote for abbreviation by $\eta_k$, have to be expressed in terms of $\psi_k$ as well. To this aim we consider activation functions from $\mathcal{S}_0$, such that we obtain the following expressions:

$$\begin{aligned} \eta_k := \sigma_k''(\boldsymbol{w}_k \bar{\boldsymbol{x}} + u_k) = \sigma_k''(\xi_k) = \sigma_k''\big(\sigma_k'^{-1}(\psi_k)\big) \qquad &\text{(in general)} \\ = \psi'\big(\sigma_k(\xi_k)\big) \cdot \sigma_k'(\xi_k) = \psi'(x_k)\,\psi_k = \psi'(\psi^{-1}(\psi_k))\,\psi_k \qquad &\text{(for } \sigma_k \in \mathcal{S}_0\text{)} \\ = \pm\, 2\,\psi_k \sqrt{\frac{\mu^2}{4} - \frac{\mu}{\alpha}\psi_k} \; , \qquad & \end{aligned} \tag{6.23}$$

where we employed equations (5.10) and (5.11) from section 5.1 to simplify the expression $\psi'\big(\sigma_k(\xi_k)\big)$. Notice, that $\psi_k = \sigma_k'(\xi_k)$ denotes a variable, while $\psi$, $\psi'$, and $\psi^{-1}$ denote the quadratic function (5.8), its derivative, and its inverse (5.11) respectively. If we choose the hyperbolic tangent as activation function, expression (6.23) simplifies further:

$$\eta_k = \pm\, 2\,\psi_k\,\sqrt{1 - \psi_k}\,.$$

It remains to evaluate the bifurcation condition $q = 0$, for which we obtain:

$$\begin{aligned} q = \langle \boldsymbol{p}, \boldsymbol{B}(\boldsymbol{c}, \boldsymbol{c}) \rangle = \sum_{k=1}^{2} \eta_k p_k (\boldsymbol{w}_k \boldsymbol{c})^2 = 0 \\ = \eta_1\,(\psi_2 w_{22} - 1)(\psi_2 \det W - w_{11})^2 - w_{12} w_{21}^2 \eta_2 \psi_1 \, . \end{aligned}$$

Together with (6.23) and (6.14) this yields a system of equations, which involves several rational functions in $\psi_1$ and $\psi_2$. This system can be solved numerically, to obtain Cusp points on the bifurcation manifold in $\boldsymbol{\psi}$-space, which subsequently can be transformed back to the input space.

## 6.4 Period-doubling bifurcation

The period-doubling bifurcation can be handled analogously to the saddle-node bifurcation due to the similarity of their bifurcation conditions. For two- and three-neuron networks we get:

$$n = 2: \qquad\qquad 1 + \psi_1 w_{11} + \psi_2 w_{22} + \psi_1 \psi_2 \det W = 0 \tag{6.24}$$

$$n = 3: \qquad \begin{aligned} +1 + \psi_1 w_{11} + \psi_2 w_{22} + \psi_3 w_{33} \\ + [\psi_1\psi_2 \det W_{(3)} + \psi_2\psi_3 \det W_{(1)} + \psi_1\psi_3 \det W_{(2)}] \\ + \psi_1\psi_2\psi_3 \det W = 0 \end{aligned} \tag{6.25}$$

where $W_{(i)}$ denote the submatrices of $W$ obtained by deletion of $i$-th row and column again. Solving for $\psi_1$ we obtain:

$$n = 2: \qquad \psi_1 = \frac{-1 - \psi_2 w_{22}}{+ w_{11} + \det W \psi_2} \qquad (6.26)$$

$$n = 3: \qquad \psi_1 = \frac{-1 - \psi_2 w_{22} - \psi_3 w_{33} - \psi_2 \psi_3 (w_{22} w_{33} - w_{23} w_{32})}{w_{11} + \psi_2 (w_{11} w_{22} - w_{21} w_{12}) + \psi_3 (w_{11} w_{33} - w_{13} w_{31}) + \psi_2 \psi_3 \det W} \qquad (6.27)$$

## 6.5 Neimark-Sacker bifurcation

The Neimark-Sacker bifurcation condition involves the bialternate product of the Jacobian matrices $J(\bar{x})$, which is a matrix of size $\frac{1}{2} n(n-1)$. We can decompose this product as follows:

$$J \odot J = (DW) \odot (DW) = (D \odot D)(W \odot W),$$

where $D \odot D$ is a diagonal matrix with all possible product pairs $\psi_i \psi_j$, $i < j$ on its diagonal. For example, in $n = 3$ dimensions we obtain:

$$D \odot D = \begin{pmatrix} \psi_1 \psi_2 & & \\ & \psi_1 \psi_3 & \\ & & \psi_2 \psi_3 \end{pmatrix}.$$

Each $\psi_i$ appears $n - 1$ times in this matrix. Thus the bifurcation condition, which is a determinant expression, becomes a polynomial where each $\psi_i$ occurs with degree $n - 1$. Thus, there exists no analytical solution to this problem if $n$ becomes larger than five. But already in three dimensions we have to deal with numerous branches of the solution manifold, making the analysis complex enough. We study the two-neuron and three-neuron cases separately in the following.

### Two neurons.

For a two-neuron network the bialternate product becomes a simple real number, i.e. $J \odot J = \psi_1 \psi_2 \det W$, and the bifurcation condition has the simple form:

$$\det(J \odot J - \mathbf{1}) = \psi_1 \psi_2 \det W - 1 = 0, \qquad (6.28)$$

which can be easily solved for $\psi_1$. The discontinuity at $\psi_2 = 0$ of the solution manifold may lead to separate branches of the bifurcation manifold if the incremental sector condition allows derivatives smaller than zero. But in most cases derivatives have to be positive or at least non-negative, such that only a single branch of the bifurcation manifold can be observed. If $\det W \leq 0$ we cannot find a positive solution at all, thus a Neimark-Sacker bifurcation cannot occur and the corresponding oscillation cannot be observed.

### Three neurons.

For a three-neuron network the bialternate product becomes a $3 \times 3$-matrix and we denote the matrix $W \odot W$ by

$$M := W \odot W = \begin{pmatrix} w_{22} w_{11} - w_{21} w_{12} & w_{23} w_{11} - w_{21} w_{13} & w_{23} w_{12} - w_{22} w_{13} \\ w_{32} w_{11} - w_{31} w_{12} & w_{33} w_{11} - w_{31} w_{13} & w_{33} w_{12} - w_{32} w_{13} \\ w_{32} w_{21} - w_{31} w_{22} & w_{33} w_{21} - w_{31} w_{23} & w_{33} w_{22} - w_{32} w_{23} \end{pmatrix}.$$

With this abbreviation the Neimark-Sacker bifurcation condition of a three-neuron network becomes:

$$
\det(J \odot J - \mathbf{1}) = \begin{vmatrix} \psi_1\psi_2 M_{11}{-}1 & \psi_1\psi_2 M_{12} & \psi_1\psi_2 M_{13} \\ \psi_1\psi_3 M_{21} & \psi_1\psi_3 M_{22}{-}1 & \psi_1\psi_3 M_{23} \\ \psi_2\psi_3 M_{31} & \psi_2\psi_3 M_{32} & \psi_2\psi_3 M_{33}{-}1 \end{vmatrix}
$$

$$
= -1 + (\psi_1\psi_2 M_{11} + \psi_1\psi_3 M_{22} + \psi_2\psi_3 M_{33})
$$
$$
- [\psi_1^2\psi_2\psi_3 \det M_{(3)} + \psi_1\psi_2\psi_3^2 \det M_{(1)} + \psi_1\psi_2^2\psi_3 \det M_{(2)}]
$$
$$
+ \psi_1^2\psi_2^2\psi_3^2 \det(M) = 0 ,
$$

(6.29)

where again $M_{(i)}$ denotes the matrix obtained from $M$ by deletion of the $i$-th row and column. Equation (6.29) is a quadratic equation of the form $a\psi_1^2 + b\psi_1 + c = 0$ where

$$
a = \psi_2^2\psi_3^2 \det(M) - \psi_2\psi_3(M_{11}M_{22} - M_{21}M_{12})
$$
$$
b = \psi_2 M_{11} + \psi_3 M_{22} - \psi_2\psi_3^2(M_{22}M_{33} - M_{23}M_{32}) - \psi_2^2\psi_3(M_{11}M_{33} - M_{13}M_{31})
$$
$$
c = \psi_2\psi_3 M_{33} - 1
$$

with solutions (compare equation (A.9) in appendix)

$$
\psi_1^+ = \frac{q}{a} \qquad \psi_1^- = \frac{c}{q} \qquad \text{where } q = -\frac{1}{2}\left(b + \operatorname{sgn}(b)\sqrt{b^2 - 4ac}\right) .
$$

As shown in appendix A.4 we have two different discontinuity manifolds separating continuous branches of the roots: $a = 0$ and $b^2 = 4ac$. These discontinuities in $a$-$b$-$c$-space implicitly define discontinuity manifolds in $\psi_2$-$\psi_3$-space as well. While the first equation ($a = 0$) leads to a quadratic equation in $\psi_2$ and $\psi_3$, which can be easily solved, the second equation ($b^2 = 4ac$) leads to a 4-th order polynomial in $\psi_2$, $\psi_3$, which cannot be solved algebraically without enormous effort. This latter manifold separates the real solution space (corresponding to admissible bifurcation points) from the complex solution space (corresponding to invalid bifurcation points). We skip the explicit and costly computation of their separating manifold, but rely on the type of solutions (real or complex) to distinguish between both regions. The discontinuity curves introduced through $a = 0$ are defined by

$$
\psi_2 = 0 \qquad \psi_3 = 0 \qquad \text{and} \qquad \psi_3(\psi_2) = \frac{M_{11}M_{22} - M_{21}M_{12}}{\psi_2 \det M} .
$$

These curves theoretically separate the $\psi_2$-$\psi_3$-plane into six different regions (see fig. 6.2a). However, because the derivatives and consequently $\boldsymbol{\psi}$ are usually positive only the two regions I and V remain, which are separated by the positive branch of $\psi_3(\psi_2)$.

## 6.6 Adaptive Step Size

Numerical computation of the bifurcation manifolds for purposes of visualisation relies on sampling data points on a regular grid in the $\psi_2$-$\psi_3$-plane and subsequently computing $\psi_1(\psi_2, \psi_3)$ according to the derived formulas. In order to achieve a uniform sampling of the data points in either $\boldsymbol{\psi}$-space or input space, the step sizes $\Delta\psi_k, k = 2\ldots n$ should be chosen according to the curvature of the bifurcation manifold at each sampling point $\boldsymbol{\psi}$ resp. $\boldsymbol{u}(\boldsymbol{\psi})$ on the $n-1$-dimensional bifurcation manifold. Hence, if we want to achieve a maximal grid spacing of size $h$ into each direction $\psi_i$ respective $u_i$ we have to choose:

$$
\Delta\psi_k = h\left(\max\left|\frac{\partial\psi_1}{\partial\psi_k}\right|, 1\right)^{-1} \qquad \text{for uniform } \boldsymbol{\psi}\text{-spacing,}
$$

$$
\Delta\psi_k = h\left(\max_{i=1\ldots n}\left|\frac{\partial u_i}{\partial\psi_k}\right|\right)^{-1} \qquad \text{for uniform } \boldsymbol{u}\text{-spacing.}
$$
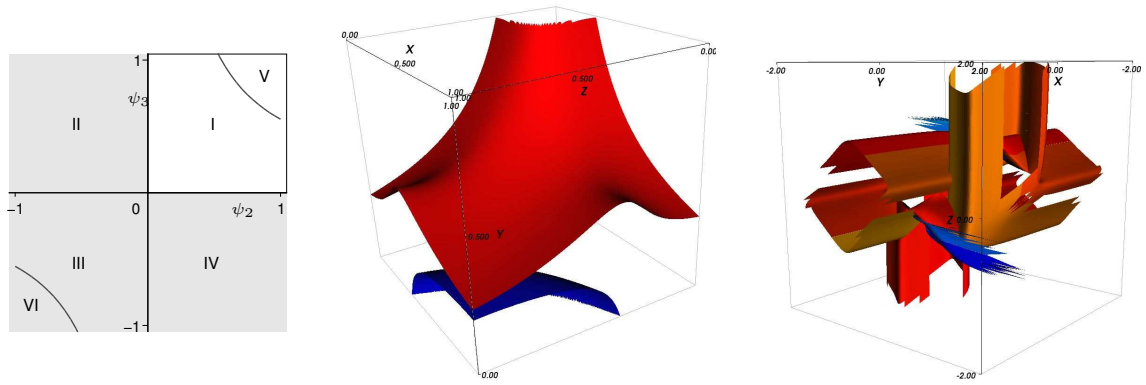
(a) Discontinuity   manifolds in $\psi_2$-$\psi_3$-space.

(b) Bifurcation manifold in $\psi$-space.

(c) Bifurcation manifold in input space.

Figure 6.2:   Neimark-Sacker bifurcation manifolds for a RNN with $W = \begin{pmatrix} 1.61 & 0.78 & 0 \\ -0.78 & 1.63 & 0.59 \\ 0 & -0.59 & 1.63 \end{pmatrix}$ and $\sigma = \tanh$. Different branches have different colours.

The necessary partial derivatives can be easily computed using the chain rule

$$\frac{\partial u_i}{\partial \psi_k} = \sum_{j=1}^{n} \frac{\partial u_i}{\partial x_j} \frac{\partial x_j}{\partial \psi_k} \, . \tag{6.30}$$

Starting from the fixed point equation (6.2) we obtain for the first partial derivative:

$$\frac{\partial u_i}{\partial x_j}(\bar{x}_j) = \delta_{ij} \frac{\partial \sigma_i^{-1}}{\partial x_j} = \frac{\delta_{ij}}{\psi_j} - w_{ij} \, ,$$

where we used the identity

$$\frac{\partial \sigma_i^{-1}(x)}{\partial x}(\bar{x}_i) = \frac{1}{\sigma_i'(\sigma_i^{-1}(\bar{x}_i))} = \frac{1}{\sigma_i'(\xi_i)} = \frac{1}{\psi_i}$$

To derive the second partial derivative involved in (6.30), we resume to activation functions within class $\mathcal{S}_0$ again. According to (5.10) we have $\bar{x}_j = \psi^{-1}(\psi_j)$ which finally leads to:

$$\frac{\partial x_j}{\partial \psi_k} = \frac{\partial \psi^{-1}}{\partial \psi_j} \frac{\partial \psi_j}{\partial \psi_k} = \frac{1}{\psi'(\bar{x}_j)} \frac{\partial \psi_j}{\partial \psi_k} = \frac{1}{-\frac{2\mu}{\alpha}\bar{x}_j + \frac{\mu}{\alpha}(\alpha + 2\beta)} \frac{\partial \psi_j}{\partial \psi_k} \, .$$

## 6.7 Symmetry of bifurcation manifolds in input space

In section 5.1 we already considered the symmetry of activation functions within class $\mathcal{S}_0$ and we have seen that this implies a symmetry in input space around the point $-W c$. If we consider arbitrary activation functions, a similar result follows as well. All extremal points of $\sigma'$, i.e. points $\xi_c$ such that $\sigma''(\xi_c) = 0$, induce a (local) symmetry around the point

$$u(\xi_c) = \xi_c - W\sigma(\xi_c)$$

in input space as can be seen easily from (6.12). To illustrate the behaviour, consider fig. 6.3 and imagine a given point $\psi$ on a single branch of the bifurcation manifold in $\psi$-space. At the global maximum $\psi_{\max} = \sigma'(\xi_1)$ there exists only a single preimage of $\sigma'$, namely $\xi_1$. Thus in input
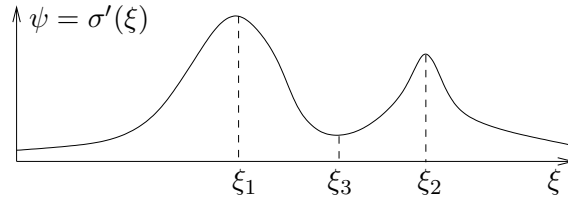
PSfrag replacements
$u_1$
$u_2$
$\psi_1$
$\psi_2$



Figure 6.3: Derivative of activation function with multiple critical points $\xi_i$.



(a) $\boldsymbol{\psi}$-space
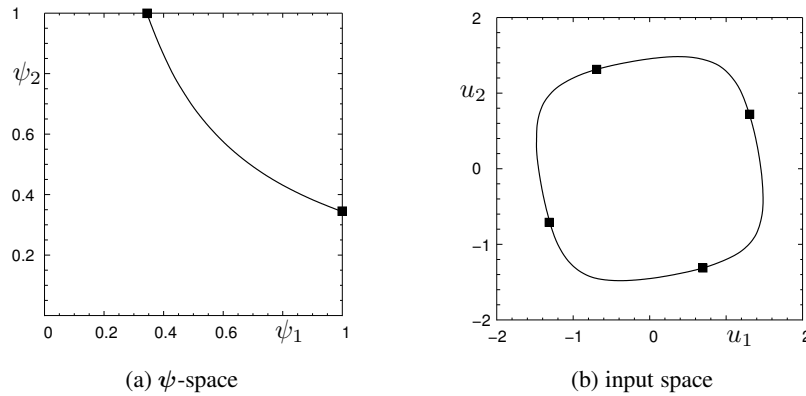
(b) input space

Figure 6.4:  Bifurcation manifolds of a two-neuron network in $\boldsymbol{\psi}$-space and in input space.  The four branches of the bifurcation curve in input space are glued together at points $\xi_i = 0$ as indicated by black boxes, which correspond to extremal points of the hyperbolic tangent.

space two different branches of the associated bifurcation manifold originate at $\xi_1 - W\sigma\xi_1$. Both branches will have a similar shape on both sides of the symmetry point due to the similarity of $\sigma'$ and smoothness of $\sigma$. At the point $\psi = \sigma'(\xi_2)$ another pair of branches of the bifurcation manifold appears in input space and one of these branches joins with one branch of the first pair at $u(\xi_3)$.

As a concrete example consider figure 6.4, which shows the bifurcation curves in $\boldsymbol{\psi}$-space (a) and in input-space (b) for a two-neuron network with the hyperbolic tangent as activation functions. Here the unique extremal point of $\tanh'$ is $\xi = 0$, which corresponds to the maximal gain $\psi = 1$. Although there exists a single bifurcation curve in $\boldsymbol{\psi}$-space, the associated curve in input space splits into four different branches in principle. Nevertheless these branches are glued together at input points, which correspond to values $\psi_i = 1$ resp. $\xi_i = 0$, and consequently form a single bifurcation curve as well.

## 6.8 Periodic orbits

In principle the presented method to detect local bifurcations of *fixed points* can be applied to detect bifurcations of periodic orbits as well. Actually we encounter so many difficulties, that an analytical solution is not feasible in practice. To discuss this, consider a $T$-periodic orbit $\bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_T$. Note that the subscripts do not denote vector components here but different points along the orbit. According to (2.20) the Jacobian of the $T$-th iterate of the network's system mapping is given as

$$J = D(\bar{\boldsymbol{x}}_T)W \, \cdots \, D(\bar{\boldsymbol{x}}_2)W \, D(\bar{\boldsymbol{x}}_1)W \,. \tag{6.31}$$

If we denote by $\psi_i$ the derivatives evaluated along the orbit, we obtain a system of equations similar to (6.4) and (6.5):

$$\psi_i = \sigma'(W\bar{x}_i + u) =: \sigma'(\xi_i)$$
$$\bar{x}_{i+1} = \sigma(W\bar{x}_i + u) = \sigma(\xi_i)$$

which imply the $T - 1$ equations:

$$\psi_i = \sigma'\left(W\sigma(\sigma'^{-1}(\psi_{i-1})) + u\right) \qquad i = 2, \ldots, T. \qquad (6.32)$$

The bifurcation conditions are identical to those in table 6.1 if the complex Jacobian (6.31) is used. At first glance these are determinant expressions in $nT$ variables $\psi_1, \ldots, \psi_T$. But according to (6.32) all $\psi_i$ with $i > 1$ can be expressed in terms of $\psi_1$ by multiple nested application of $\sigma'$, $\sigma$ and $\sigma'^{-1}$, which makes the bifurcation condition a very complicated nonlinear equation. Further more – because $\sigma'$ generally has multiple preimages for a given value $\psi$ – we actually get a whole bunch of bifurcation conditions defining a distinct branch of the bifurcation manifold in $\psi$-space each. For example for activation functions within class $\mathcal{S}_0$ equation (6.32) reads

$$\psi_i = \sigma'(W\bar{x}_i + u) = \sigma'(W\psi^{-1}(\psi_{i-1}) + u),$$

where the components of the vector function $\psi^{-1}$ are defined by (5.11). Thus we get two different solution equations per nesting level resulting in a total of $2^{T-1}$ different bifurcation conditions. As we have seen in the analysis of fixed point bifurcations before, each of them may produce several branches of the bifurcation manifold in $\psi$-space again. Altogether it is not feasible anymore to compute the bifurcation manifolds for $T$-periodic orbits, even if $T$ is small.

## 6.9 Summary

Before we continue with the discussion of several bifurcation diagrams in the next chapters, we shortly summarise the general results obtained so far and outline the algorithmic procedure, which is necessary to compute the bifurcation manifolds for purposes of visualisation.

We propose to compute bifurcation manifolds initially in the abstract space of activation function derivatives $\psi = \sigma'(\xi)$, which has a much simpler structure than the input space itself. Subsequently the bifurcation manifolds can be transformed into the input space employing equation (6.12), which generates $2^n$ different branches for each solution branch in $\psi$-space if typical sigmoid activation functions are used. If the bifurcation manifold in $\psi$-space contains submanifolds, along which some $\psi_i$ reach extremal values – e.g. their maximal values $\psi_i = \frac{1}{4}\mu_i\alpha_i$ – appropriate branches in the input space can be glued together along the associated submanifolds in input space.

In order to reduce the computational effort, in the following we restrict ourselves to activation functions within the class $\mathcal{S}_0$. As we have seen in section 5.1, all functions within this class lead to an equivalent dynamical behaviour, if the weight and input parameters are adjusted appropriately. Hence, we can consider a single representative of this class, and we choose the hyperbolic tangent due to its high symmetry.

For all local fixed point bifurcations considered here, the bifurcation manifolds in $\psi$-space are implicitly defined by the test functions of table 6.1. These equations have to be solved for an arbitrary $\psi_i$ in dependence of all other variables $\psi_j$, $j \neq i$. For this purpose we sample the $n-1$-dimensional hypercube $\mathcal{H}_i = \prod_{j \neq i}(0, 1]$ along a regular grid, whose spacing into the direction $\psi_j$ is computed by averaging all directional derivatives of the solution manifold along the hyperplane,
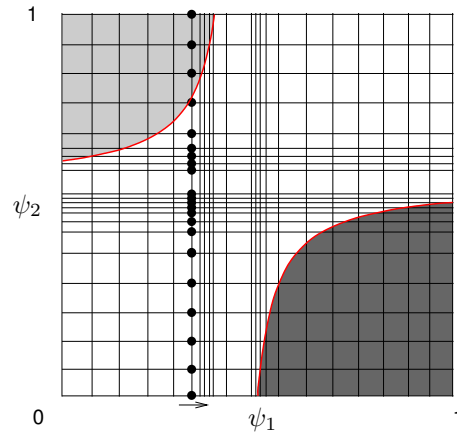
Figure 6.5:   Computation of the bifurcation manifolds in $\boldsymbol{\psi}$-space along a regular grid in $\psi_1$-$\psi_2$-space. The grid spacing into an arbitrary direction $\psi_j$ depends on the corresponding directional derivatives of the solution manifold, averaged along the hyperplane (here along the line), which is perpendicular to the direction $\psi_j$.

which is perpendicular to the direction $\psi_j$. Figure 6.5 shows this situation in case of a three-neuron network, where we solve the bifurcation conditions for $\psi_3$ in dependence of $\psi_1$ and $\psi_2$. The spacing into the direction of $\psi_1$ at the value indicated by the small arrow, is computed from the derivatives averaged over all explicitly marked grid points along the $\psi_2$-direction.

In order to circumvent wall-like artefacts during the visualisation process, which stem from discontinuities of the solution manifolds, we have to assign each solution point to a unique solution branch. As we have seen during the derivation of the solution equations in the previous sections, we obtain up to $2^{n-1}$ disjoint solution branches, whose boundaries are defined by complex implicit conditions in turn. Hence the effort to assign solution points to different branches increases rapidly with the number of neurons, which makes it not feasible to compute bifurcation manifolds of networks with more than three neurons analytically. It has to be continued to study these large network by numerical methods, e.g. continuation techniques as described in the introduction of this chapter.

# 7 Discussion of Bifurcation Curves of Two-Neuron Networks

In the following we apply the analytical expressions, developed in the previous chapter, to discuss some interesting bifurcation diagrams of two-neuron networks. A similar analysis was performed before for single-neuron networks [Pasemann, 1993b, 1997], two-neuron networks [Pasemann, 1993a, 2002; Beer, 1995] and ring networks [Pasemann, 1995], where Pasemann studied these networks by numerical simulations only. To the best of our knowledge analytical expressions for bifurcation manifolds of fully recurrent discrete-time RNNs have not been published before. Our expressions can be evaluated fast enough to allow an interactive study of the bifurcation manifolds under varying connection weights. This allows us to explore the dependency of the manifolds on the weight matrix online. By means of such an interactive exploration we obtained new results and interesting bifurcation diagrams, which we will discuss in the subsequent sections.

In order to simplify the assignment of the various bifurcation curves to their corresponding bifurcation types, we utilise a colouring scheme. In all presented bifurcation diagrams the four possible branches of the saddle-node bifurcation manifold are coloured in red, green, blue and yellow. The Neimark-Sacker bifurcation curve, which is always a closed curve in input space, is coloured black, and all branches of the period-doubling bifurcation manifold are coloured magenta.

## 7.1 Bifurcation manifolds in different spaces

As we have seen before, besides the input space the $\psi$-space is an interesting space to study bifurcations of fixed points of RNNs. For general two-dimensional dynamical systems the stability analysis of fixed points proceeds in the simplest way if studied in a third space, which is spanned by the trace $T = \operatorname{tr} J(\bar{\boldsymbol{x}})$ and the determinant $D = \det J(\bar{\boldsymbol{x}})$ of the corresponding Jacobian matrix. In this case the domain of stability is given by a triangle in the $T$-$D$-plane bounded by the three straight lines $T - D = 1$, $T + D = 1$ and $D = 1$, which correspond to the bifurcation boundaries of saddle-node, period-doubling and Neimark-Sacker bifurcation respectively (see fig. 7.1a). This result can be summarised in the following theorem [Thompson and Stewart, 1986]:



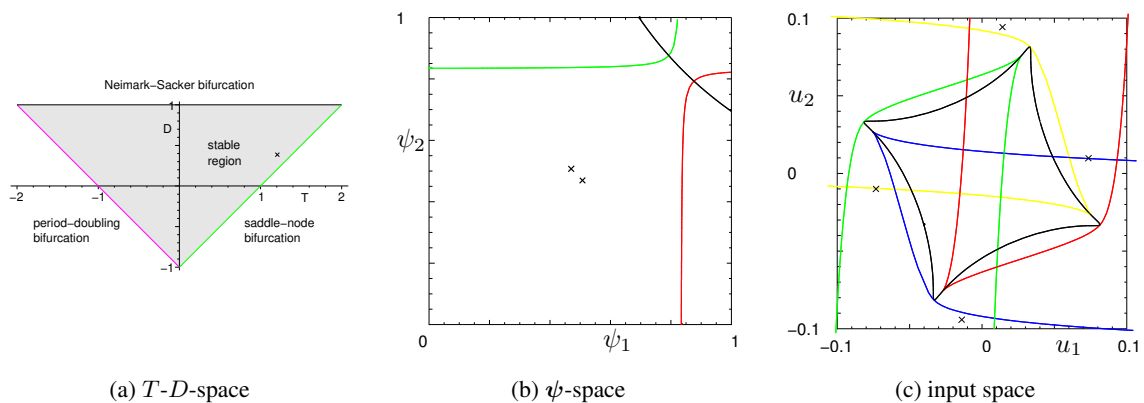(a) $T$-$D$-space    (b) $\psi$-space    (c) input space

Figure 7.1:   Bifurcation curves of fixed points in different spaces. Corresponding points in these spaces are marked by crosses.

**Theorem 7.1** Let $\bar{x}$ be a fixed point of a two-dimensional dynamical system and denote with $T$ and $D$ the trace and the determinant of its corresponding Jacobian respectively. Then it holds:

(i) The fixed point is nonhyperbolic if and only if $(T, D)$ lies on the border of the triangle, which is bounded by the lines $T - D = 1$, $T + D = 1$ and $D = 1$.

(ii) The hyperbolic fixed point is asymptotically attracting if and only if $(T, D)$ lies inside this triangle.

A subset of the $T$-$D$-space can be mapped back to the $\psi$-space by inversion of the following mapping from $\psi$-space to $T$-$D$-space:

$$
\begin{aligned}
T &= \psi_1 w_{11} + \psi_2 w_{22} \\
D &= \psi_1 \psi_2 \det W \, .
\end{aligned}
\tag{7.1}
$$

Given a specific trace and determinant within the stable triangular region of the $T$-$D$-plane, we find at most two corresponding points in $\psi$ space, for which we find in turn multiple associated fixed points according to (6.13). Of course, these fixed points can be observed only if the associated input vector is chosen, which can be computed with equation (6.12). These different points are visualised by cross marks within figures 7.1a-c. The intercorrelation between the different spaces is summarised by the following diagram:

$$
\boldsymbol{x} \xrightarrow{\;(6.9)\;} \boldsymbol{\xi} \xrightarrow{\;(6.4)\;} \boldsymbol{\psi} \xrightarrow{\;(7.1)\;} T, D
$$
$$
\Big\downarrow {\scriptstyle (6.12)}
$$
$$
\boldsymbol{u}
$$

Each arrow indicates a proper mapping between the adjacent spaces, but all mappings are many-to-one, i.e. have multiple preimages. Due to the resulting ambiguities the $\psi$-space as well as the $T$-$D$-space are not suitable to study bifurcation diagrams, because there exists no proper mapping back to the input values, which are the actual bifurcation parameters. Hence points within these spaces do not fully determine the dynamical behaviour of the RNN – not even that of all fixed points. This becomes possible within the input space only.

As an example consider figure 7.1 again. There exist four different points in input space, which correspond to a single point in $\psi$-space. But, the points in input space belong to different regions of dynamical behaviour: while two points lie exactly on a bifurcation curve and correspond to the associated bifurcation type (saddle-node), the other two input values do not represent bifurcation points. Hence, we cannot reason from a point in $\psi$-space to unique dynamical behaviour of the RNN. The underlying reason for the qualitative difference of the various spaces with respect to a dynamical analysis is that the different branches in the input space, which correspond to a single solution branch in $\psi$-space, can cause multiple intersections and consequently produce additional regions of different dynamical behaviour.

Nevertheless the $\psi$-space and especially the $T$-$D$-space facilitate the detection of fixed point bifurcations, due their easily checkable bifurcation conditions. Further, the use of the $\psi$-space for the computation of bifurcation manifolds actually makes the handling of the various solution branches feasible.

In order to classify the different regions in input space according to their corresponding dynamical behaviour, we need a reference point within the input bifurcation diagram, for which the dynamical behaviour can be deduced analytically. Varying the inputs along an arbitrary path starting at this reference point, bifurcations of fixed points occur, if a bifurcation boundary is crossed. Knowing

the properties of the corresponding bifurcation type we can predict parts of the dynamical behaviour within the newly entered dynamical regime. In this manner we can predict the number and stability properties of all fixed points observable within different regions of the bifurcation diagram. Note that we cannot predict the whole repertoire of dynamical behaviour, because we do not know the bifurcation manifolds corresponding to bifurcations of periodic orbits or to global bifurcations.

## 7.2 Existence of a unique fixed point

A particularly suitable reference point is an input vector, which makes the RNN globally asymptotically stable. In contrast to the more elaborate methods, which we presented in chapter 4, here we employ the conservative but simple small gain theorem to prove GAS. Consider an arbitrary fixed point $\bar{x}$ of the dynamical system, which is locally stable if the spectral radius of the associated Jacobian matrix $J(\bar{x})$ is strictly less than one. The spectral radius can be bounded from above by any induced matrix norm $\|\cdot\|$ according to the inequality:

$$\rho(J(\bar{x})) \leq \|J(\bar{x})\| \leq \|D(\bar{x})\| \, \|W\| \, .$$

Thus the fixed point $\bar{x}$ is locally asymptotically stable if $\|D(\bar{x})\| \, \|W\| < 1$. Hence, if the derivatives $D(\bar{x}) = \mathrm{diag}(\psi)$ at the fixed point or the norm of the weight matrix are small enough the fixed point will be stable. Particularly, if $\|W\|$ grows without bounds, all stable fixed points will have vanishing derivatives, or equivalently will be driven into the saturation regime. This result was strictly proven for multi-sigmoid activation functions by Hirsch [1994].

If the small gain condition $\|D(x)\| \, \|W\| < 1$ holds within the whole state space, i.e.

$$\sup_{\substack{x \in X \\ i=1,\dots,n}} \sigma_i'([Wx + u]_i) = \sup_{x \in X} \|D(x)\| < \|W\|^{-1} \tag{7.2}$$

the system's mapping is a contraction and the fixed point is globally asymptotically stable [Cessac, 1994]. This is an even stronger condition then **D**-stability of $W$, because the spectral radius of general matrices is usually smaller than any matrix norm. [1] Given a fixed weight matrix, equation (7.2) will be satisfied if all input components $u_i$ are large enough in magnitude, thus defining a suitable reference point in the bifurcation diagrams.

Now we can turn to the discussion of concrete bifurcation diagrams. Remember, that we use the hyperbolic tangent as activation function in all examples. All results easily translate to other sigmoid functions from class $\mathcal{S}_0$, and if we denote the activation function with $\sigma$ we actually refer to a function from this class.

## 7.3 Rotation matrices

First we consider weight matrices which represent scaled rotations in the plane. Parameterising the rotation angle $\theta$ and the scaling factor $r$ we obtain:

$$W(r, \theta) = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \, . \tag{7.3}$$

---

[1] An exception are hermitian matrices in conjunction with the spectral norm, i.e. the matrix norm induced by the euclidian norm. In this case the spectral radius equals the matrix norm.

Due to the nature of this weight matrix we can expect to observe stable periodic or quasi-periodic orbits if $r$ is larger than one. Otherwise we observe a globally attracting fixed point only, as the following theorem shows:

**Theorem 7.2** We consider the two-neuron network $\boldsymbol{x} \mapsto \mathbf{tanh}(W\boldsymbol{x} + \boldsymbol{u})$, where the weight matrix $W$ is given by (7.3). This RNN is absolutely stable if and only if $r \leq 1$, where the origin is the unique attractor.

**Proof**    We first prove the "if"-case.  Obviously the hyperbolic tangent satisfies the standard incremental sector condition (4.8).  Further the weight matrix (7.3) is diagonally Schur stable if $r < 1$, because we can use $P = \mathbf{1}$ and obtain:

$$W^t P W - P = (r^2 - 1)\,\mathbf{1} < 0 \quad \Leftrightarrow \quad |r| < 1\,.$$

Consequently, theorem 4.3 can be applied, which proves absolute stability of the RNN. The case $r = 1$ can be included because the inequality

$$\mathbf{tanh}(\boldsymbol{x})^t \mathbf{tanh}(\boldsymbol{x}) < \boldsymbol{x}^t \boldsymbol{x}$$

holds strictly for all $\boldsymbol{x} \neq 0$, such that $V(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{x}$ is a Lyapunov function in this case as well.

It remains to prove the "only-if"-case, where we simply give a counterexample.  For $r > 1$ the Neimark-Sacker bifurcation condition $\psi_1 \psi_2 \det W = \psi_1 \psi_2\, r^2 = 1$ can be satisfied for values $\psi_i \in (0, 1]$. Consequently the associated fixed point is not stable anymore, which is in conflict to its global asymptotic stability. $\square$
This theorem extends a statement given by Tonnelier et al. [1999], who proved the local asymptotic stability of the origin for values $r < 1$.

In the following we discuss the development of the bifurcation diagram in input space while fixing $r = 1.2$ and varying $\theta$ within the interval $[0..\pi]$. Other values of $r > 1$ produce qualitatively similar behaviour.  The remaining quantitative differences will be discussed at the end of this section, such that we obtain a complete bifurcation diagram of this two-neuron network, where the inputs as well as $r$ and $\theta$ are considered as bifurcation parameters, resulting in a four-dimensional bifurcation diagram.

**Remark:**  Due to the periodicity of the trigonometric functions it suffices to consider values of $\theta$ within the interval $(-\pi, \pi]$.  We can reduce this interval further to $[0, \pi]$, if we take into account that a sign change of $\theta$ transposes the weight matrix, or equivalently reorders the neurons.

### 7.3.1  $\theta = 0$

At $\theta = 0$ the neurons are decoupled such that the system becomes a direct product of two one-dimensional RNNs, each having the bifurcation diagram of example 3.1 with $w = r$. Hence we observe up to $2^n$ coexisting stable fixed points. Figure 7.2 shows the corresponding bifurcation diagram, which partitions the input space into nine regions.  Within regions I exists a globally asymptotically stable fixed point, while regions II represent Cusp regions, i.e. there exist three fixed points simultaneously – two stable nodes and one saddle. Crossing the saddle-node bifurcation curves, which separate both regions, the typical hysteresis effect can be observed. Within the intersection of the horizontal and vertical region-II-stripes both independent sub-networks are within their Cusp domain in parallel and we observe four stable nodes, four saddle and one unstable node. The location of the stable fixed points and the extension of their bassins of attraction depend on the concrete input vectors in any case.
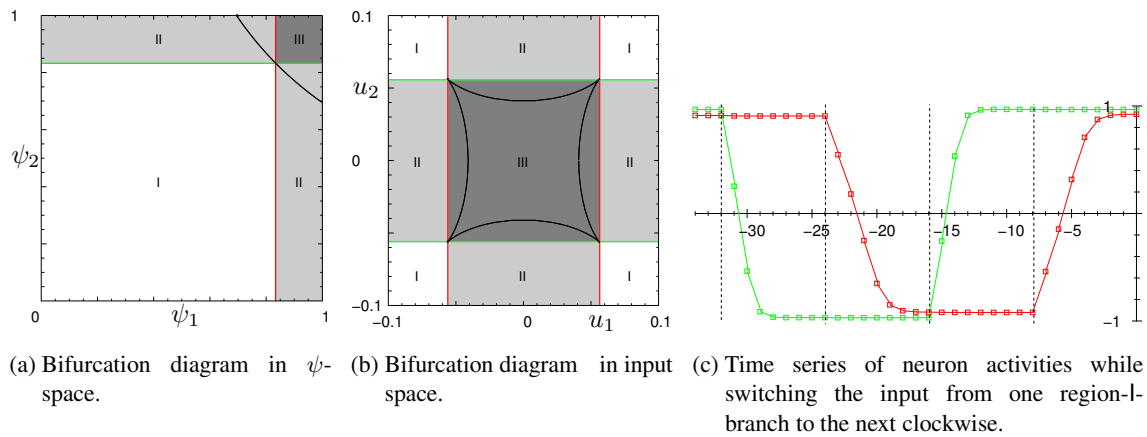
(a) Bifurcation diagram in $\psi$-space.

(b) Bifurcation diagram in input space.

(c) Time series of neuron activities while switching the input from one region-l-branch to the next clockwise.

Figure 7.2: Dynamical behaviour of RNN with weight matrix $W(r = 1.2, \theta = 0)$.

Obviously at the intersections of the bifurcation curves at the corners of region III both neurons undergo a saddle-node bifurcation simultaneously. Hence these intersections correspond to double saddle-node bifurcations. Varying the inputs around these points does not lead to noticeable changes of the actual dynamical behaviour due to the hysteresis effect. Although new fixed points are created and destroyed while crossing the bifurcation curves, the old fixed points keep stable and a nearby state will not leave their vicinity. Only if an opposite bifurcation curve is crossed, i.e. the hysteresis gap is covered, these stable fixed points disappear and all trajectories are attracted by the remaining fixed point.

This bifurcation diagram is complete, because bifurcations of periodic points cannot occur due to the fact that the network splits into a direct product of one-dimensional sub-systems, each having a strictly monotonic system mapping. The Neimark-Sacker bifurcation curves – drawn in figure 7.2 as thin lines – are not relevant, because they correspond to real eigenvalues $\lambda$ and $\lambda^{-1}$ and thus do not produce any bifurcations. Remember, that the test function (3.19) indicates only that the product of two eigenvalues equals one — it does not detect a conjugate complex pair of eigenvalues per se.

### 7.3.2 $\theta > 0$, no oscillatory orbits

If $\theta$ becomes nonzero the situation changes slightly. Now there exists a nonempty region $C$ in $\psi$-space, which corresponds to complex eigenvalues of the Jacobian. This region is bounded by two straight lines through the origin, which indicate the occurrence of double real eigenvalues (see fig. 7.3a). A fraction of the Neimark-Sacker bifurcation curve lies within region $C$ and leads to a supercritical Neimark-Sacker bifurcation and its associated attractive invariant curve.

Nevertheless, as long as $\theta$ is small, in simulations we observe attractive oscillatory orbits within tiny regions of the input space only, although the Neimark-Sacker curve encloses a much larger region. Rather almost all trajectories are attracted to a fixed point, if an input vector within the oscillatory region is applied. The key to explain this unexpected behaviour is a global bifurcation, as we will see later in this section.

But first, consider the (coloured) saddle-node bifurcation curves in input space (figure 7.3b). They enclose four Cusp regions, denoted by II. Further, there exist regions in input space where two, three or all four Cusp regions partially overlap. These regions correspond to the existence of five,
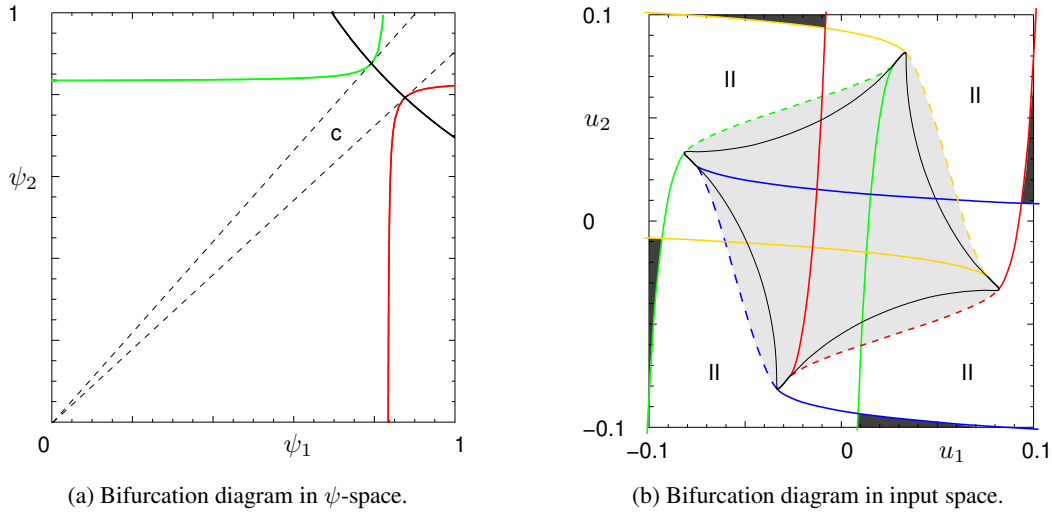
(a) Bifurcation diagram in $\psi$-space.

(b) Bifurcation diagram in input space.

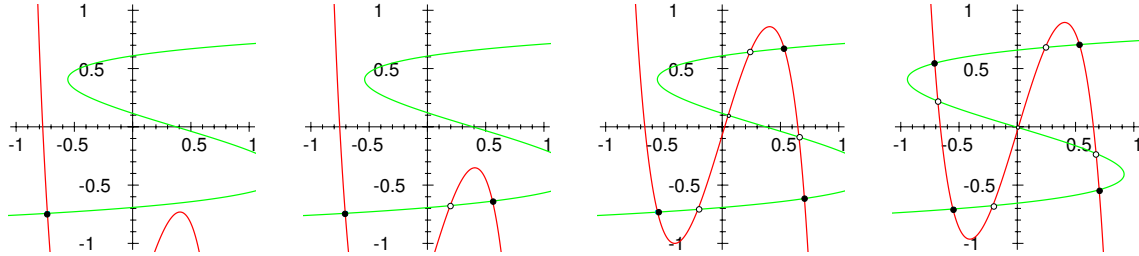Figure 7.3: Bifurcation curves of RNN with weight matrix $W(r = 1.2, \theta = 0.05)$.



Figure 7.4:   Graphs of the nullclines (in state space) indicating fixed points of the RNN. While the shape of the nullclines is determined by the weights, their location is given by the input vector. Stable nodes are marked by solid circles while saddles are marked by open circles.

seven or nine fixed points respectively.[2]  Outside all Cusp regions, i.e. within the dark shaded regions, which extend infinitely into the input space, a globally asymptotically stable fixed point exists. At any intersection of two saddle-node bifurcation curves occurs a double saddle-node bifurcation, i.e. at two *different* points in state space occur saddle-node bifurcations simultaneously.

Figure 7.4 illustrates the location of all existing fixed points for several choices of the input vector. Visualised are the nullclines, which are defined by the fixed point conditions $\bar{x}_1 = \varphi_1(\bar{\boldsymbol{x}})$ (red curve) and $\bar{x}_2 = \varphi_2(\bar{\boldsymbol{x}})$ (green curve). Clearly their intersections correspond to fixed points of the RNN. Solving these equations we obtain the following expressions for the nullclines in state space:

$$
\begin{aligned}
x_2(x_1) &= \frac{\sigma_1^{-1}(x_1) - w_{11}x_1 - u_1}{w_{12}} \qquad \text{(red curve)} \\
x_1(x_2) &= \frac{\sigma_2^{-1}(x_2) - w_{22}x_2 - u_2}{w_{21}} \qquad \text{(green curve)} .
\end{aligned}
\tag{7.4}
$$

The function $\sigma^{-1}(x) = \sigma_{\alpha,\beta,\mu}^{-1}(x)$ is monotonically increasing with

$$
\lim_{x \searrow \beta} \sigma_{\alpha,\beta,\mu}^{-1}(x) = -\infty \quad \text{and} \quad \lim_{x \nearrow \alpha+\beta} \sigma_{\alpha,\beta,\mu}^{-1}(x) = +\infty .
$$

---

[2]  Borisyuk and Kirillov [1992] and Beer [1995] presented a similar bifurcation diagram for a two-neuron continuous-time RNN, but they obtained the bifurcation curves by numerical methods.
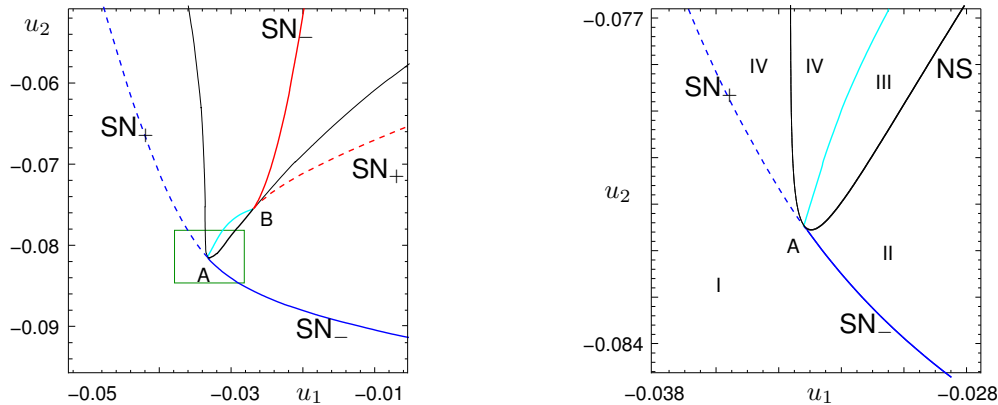
Figure 7.5: Magnifications of the bifurcation diagram depicted in figure 7.3. Along the dashed curves an unstable node and a saddle appear or disappear through saddle-node bifurcations. The bold black curve represents the Neimark-Sacker bifurcation curve, while the thin black curves is a non-bifurcation curve, representing neutral saddles. The endpoints A and B of the Neimark-Sacker bifurcation curve mark 1:1 resonances of different saddle-node pairs.

The linear term $-w_{ii}x_i$ in (7.4) leads to a bending of the graph of the nullclines if $\frac{1}{4}\mu\alpha\,w_{ii} > 1$, enabling more than one fixed point for certain inputs. While the shapes of the nullclines are determined by the weights only, the inputs $u_1$, $u_2$ shift them vertically and horizontally respectively, producing the different fixed point patterns. Note that stable nodes and saddles alternate along the nullclines.

Consider once more the bifurcation diagram in figure 7.3 and its magnifications in figure 7.5. The Neimark-Sacker curve is completely located within Cusp regions and touches their boundaries at eight points, which are located pairwise in the vicinity of the four corners of the diamond shaped region – in figure 7.5 this pair is denoted by $A$ and $B$ respectively. A numerical analysis [3] reveals that these intersections correspond to a 1:1 resonance, i.e. fixed points having double real eigenvalues $\lambda_1 = \lambda_2 = 1$. As discussed in section 3.2.4 these codim-2 bifurcation points partition the Neimark-Sacker curve into regions, which correspond to a true Neimark-Sacker bifurcation (bold sections) and into regions, which indicate a neutral saddle only, but no true bifurcation (thin sections).

In order to discuss the dynamical behaviour in a neighbourhood of the 1:1 bifurcation point, consider the magnification in figure 7.5b. Note that the magnified region is located within the green Cusp region of figure 7.3. Thus there exist at least three fixed points in state space. To get an overview of the phase portrait have a look at figure 7.6a. The three fixed points located on the lower branch of the (green) $x_2$-nullcline originate from the green Cusp region. They are far away from the other two fixed points in the upper right corner, which in turn are located within a neighbourhood of the 1:1 resonance point. This neighbourhood is magnified in figure 7.6b for several choices of the input vector. In the following we will concentrate on this neighbourhood only and ignore other existing fixed points. Our discussion of the 1:1 resonance follows the one presented in Kuznetsov [1995] for general discrete-time dynamical systems.

For inputs in region I of figure 7.5, there exist no fixed points and all trajectories leave the considered neighbourhood in state space. If the inputs are varied along a path counter-clockwise around the 1:1 resonance point (A) in parameter space, first a pair of fixed points (a saddle and a stable node) appears, when the blue solid curve $SN_-$ is crossed from region I to region II. Next, the stable fixed point becomes unstable undergoing a Neimark-Sacker bifurcation along the bold black

---

[3] We use the software package Content [Kuznetsov and Levitin] to study high-codim bifurcations along codim-1 bifurcation curves.

(a) overview of the state space, input $u$ in region IV

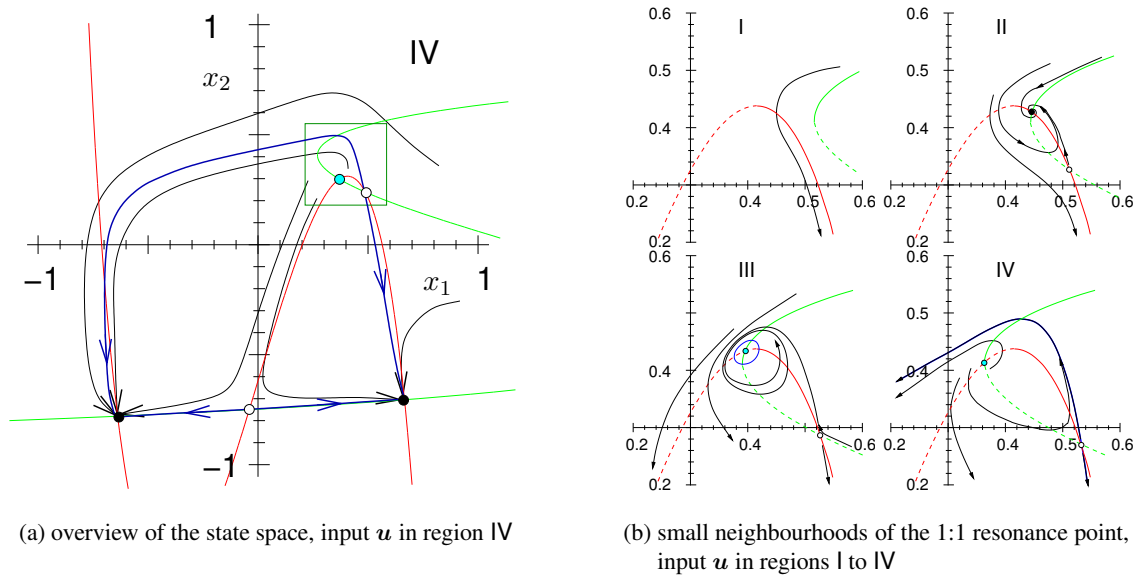(b) small neighbourhoods of the 1:1 resonance point, input $u$ in regions I to IV

Figure 7.6: Phase portraits near the 1:1 resonance bifurcation. The stable invariant curve is shown blue and connects stable nodes and saddles, if the input is chosen from region IV. Nullclines are drawn as green and red curves – stable branches as solid curves and unstable branches as dashed curves. Black curves visualise exemplary trajectories.

curve NS. Consequently, for inputs within region III of the input space exists an attracting closed invariant curve in phase space, which is depicted as a bold blue curve in figure 7.6b. The diameter of this invariant curve rapidly increases as the input departs from the Neimark-Sacker bifurcation curve (fig. 7.7a), until it finally touches the saddle point (fig. 7.7b). At this input value a global bifurcation happens, where the stable and unstable manifolds of the saddle coincide, or intersect transversally giving rise to a homoclinic structure. This type of global bifurcation is called homoclinic loop or saddle separatix loop bifurcation [Hale and Koçak, 1991] and is indicated in figure 7.5 by the cyan bifurcation curve originating from the 1:1 resonance points. The actual location of this bifurcation curve cannot be computed analytically, but simulations suggest, that it runs very close to the Neimark-Sacker curve, such that the oscillatory parameter domain is very small. This explains, why it is very difficult to observe stable oscillatory orbits in practice.

What happens to the blue invariant curve in region IV? Both branches of the unstable manifold of the saddle leave the considered neighbourhood and finally run into a stable node – in our example initially into the lower left one in fig. 7.5a. Particularly, we cannot observe orbits anymore, which are oscillatory in the limit $t \to \infty$. After a further global bifurcation, when the upper branch of the unstable manifold gets captured within the bassin of attraction of the opposite stable fixed point, we finally obtain the phase portrait depicted in 7.5a. Actually, the boundary between both bassins of attractions is formed by the stable manifold of the lower saddle point. Hence, at this secondary global bifurcation, a branch of the stable manifold of the distant saddle and a branch of the unstable manifold of the saddle, which is involved in the 1:1 resonance, coincide. As a consequence, an invariant curve of large diameter occurs, which connects all stable nodes and saddles via unstable manifolds of the saddles. The dynamics on this curve attracts every trajectory to the nearest stable node – unless the trajectory starts at a saddle point itself.

As we already mentioned, the thin black curve is a non-bifurcation curve, which represents neutral saddles only. Thus, the next real bifurcation occurs if the blue dashed bifurcation curve $SN_+$ is crossed back into region I. Here the unstable node and saddle coalesce and disappear through a saddle-node bifurcation. Simultaneously the closed invariant curve disappears, because such an

PSfrag replacements

$u_1$
$u_2$
$\psi_1$
$\psi_2$

(a) swelling invariant curve

PSfrag replacements

$u_1$
$u_2$
$\psi_1$
$\psi_2$

(b) bifurcation point

PSfrag replacements

$u_1$
$u_2$
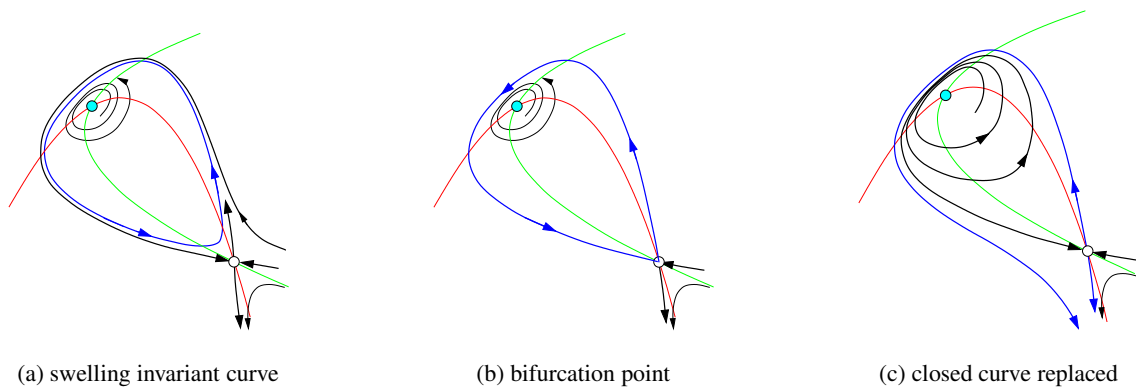$\psi_1$
$\psi_2$

(c) closed curve replaced

Figure 7.7: Homoclinic bifurcation near the 1:1 resonance point. Stable and unstable manifolds of saddles are shown as bold black curves, while normal black curves represent other exemplary trajectories. The swelling invariant curve (a) touches the saddle at the bifurcation point and the stable and unstable manifolds of the saddle coincide (b). After the bifurcation the closed invariant curve vanishes in its original form and all trajectories leave the considered neighbourhood (c). Simultaneously the stable manifold of the attracting fixed point outside this neighbourhood and the unstable manifold of the saddle coalesce and form a new invariant curve, which replaces the previous one.

attracting curves requires an unstable node inside, which repels all trajectories towards the curve.

Summarising, we can conclude that there exists a closed invariant curve in phase space as long as the input to the RNN is located within the shaded region in input space. This region is bounded on the one hand by the dashed saddle-node bifurcation curves, along which an unstable node and a saddle are created, and on the other hand by the bold Neimark-Sacker bifurcation curves, along which a stable node becomes an unstable node (compare fig. 7.3b once more). Notice, that there can exist a single unstable node only. This can be seen for example from the intersection patterns of the nullclines in figure 7.4, where nodes and saddles alternate along each nullcline. All saddle-node bifurcations, which occur within the shaded region, create or destroy a stable node and a saddle *on* the closed invariant curve.

### 7.3.3 $\theta > 0$, oscillatory orbits

Based on the observations of the previous section, we can expect to observe oscillatory orbits within wide regions of the parameter space only if the closed invariant curve, which exists for inputs within the light shaded region of the parameter space, does not include stable nodes and saddles, i.e. if the shaded region is not completely covered by Cusp regions. This is ensured above a critical parameter value $\theta_{\mathrm{osc}}(r)$, where all four branches of the saddle-node bifurcation manifold meet at the origin. Above this critical value we obtain the bifurcation diagram which is shown in figure 7.8. The only qualitative difference to the bifurcation diagram in figure 7.3b is the dark shaded area in the centre of the input space, which corresponds to the existence of oscillatory orbits in phase space.

We shortly describe the change in dynamical behaviour when the saddle-node bifurcation curve is crossed from the light to the dark shaded region. As we have seen in the previous section, there exists an attracting closed invariant curve for input vectors within the light shaded region, but stable nodes on this curve attract almost all trajectories[4] (fig. 7.9a). If the saddle-node bifurcation

---

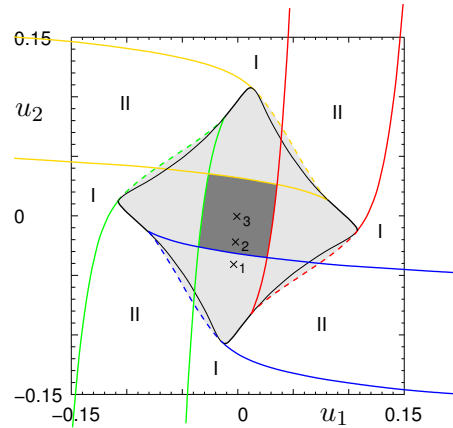[4] Trajectories starting at any unstable fixed point obviously do not leave this fixed point anymore.

Figure 7.8: Bifurcation diagram of RNN with weight matrix $W(r = 1.2, \theta = 0.05)$. Within the dark shaded area oscillatory orbits can be observed. The numbered crosses mark input vectors used to generate the plots in figure 7.9.

curve is crossed, the stable node and the saddle on the closed curve coalesce and finally disappear for inputs within the dark shaded region. Only the unstable node and the closed invariant curve remain, such that almost all trajectories are attracted by the invariant curve.[4]

The saddle-node bifurcation on the limit cycle has two interesting properties. First, the amplitude of the appearing oscillatory orbit is large, because the closed invariant curve exists before the bifurcation occurs. This is in contrast to Neimark-Sacker bifurcations, where the invariant curve is created with small diameter during the bifurcation and swells slowly with increasing distance of the parameters to the bifurcation manifold. Second, the frequency of the oscillatory orbit is arbitrarily small. If the input vector stays close to the saddle-node bifurcation curve, the vector field $(x, \varphi(x))$ is still small in the vicinity of the former stable node. Thus the state variable stays for a long time close to this location during its motion along the closed curve (fig. 7.9c). Consequently, the corresponding time series exhibits short bursts of neural activity between long periods of nearly constant activity.

The bifurcation diagram in figure 7.8 shows, that there do not exist any other fixed points besides the unstable node, if input parameters are chosen within the dark shaded region. Simulations further suggest, that there do not exist any other periodic orbits apart from those on the closed invariant curve. Hence, the invariant curve is asymptotically attracting for almost all initial states[4]. The dynamics on the closed invariant curve itself can be very complex. If the inputs are varied within the dark shaded region, various phase-locking phenomena occur as reported by many authors [Chapeau-Blondeau and Chauvet, 1992; Doyon et al., 1993; Haschke et al., 2001].

### 7.3.4 $\theta \approx \pi$

If $\theta$ increases further and becomes close to $\pi$ we observe period-doubling bifurcations. Obviously the neurons are decoupled at $\theta = \pi$, obeying the evolution law $x_i \mapsto \sigma(-rx_i + u_i)$ each. These mappings have a single fixed point for all choices of inputs $u_i$, which can be easily seen from figure 7.10: there exists a single intersection of the graph of the monotonically decreasing activation function and the line $y = x$.

Considering both, the self-connection weight $w = -r$ and the input $u$ as parameters, we can compute the period-doubling bifurcation manifold in $(u, w)$-space shown in figure 7.11. This completes the bifurcation diagram of figure 3.2a, where we already displayed the saddle-node
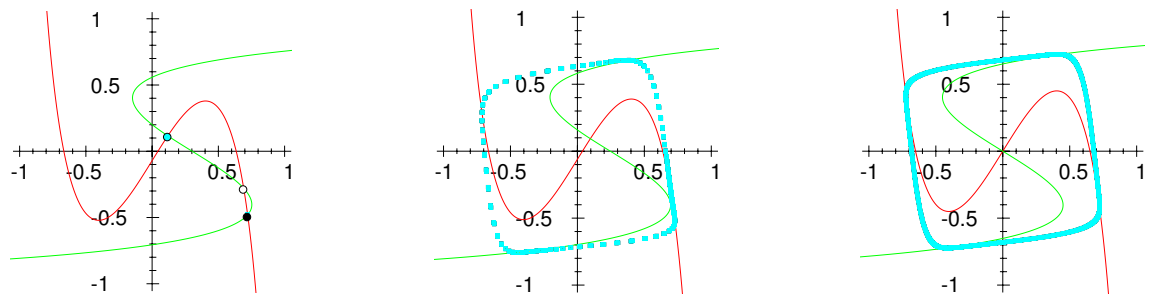
(a) Nullclines in state space, indicating fixed points of the RNN. An exemplary oscillatory orbit, which emerges for input vectors chosen from the dark shaded region in input space, is depicted by cyan-coloured boxes. The high density of points in the vicinity of the former saddle-node pair indicates that the time series stays for a long time at this location.
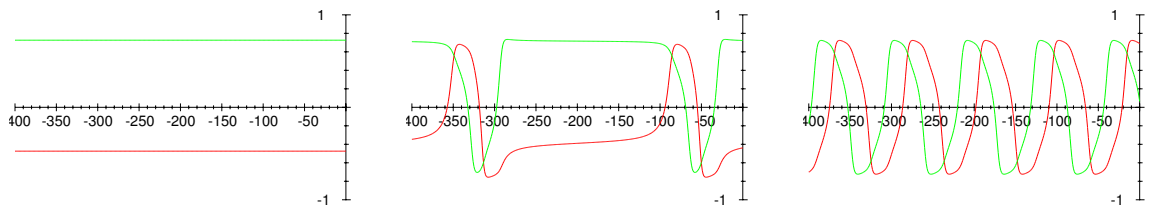


(b) fixed point                    (c) long-period orbit                    (d) short-period orbit

Figure 7.9:  Bifurcation sequence, which is obtained when the input vector is varied from inside the blue Cusp region, into the dark shaded region. Shown are the nullclines (a) and the time series (b-d) at the three different input points marked in figure 7.8.
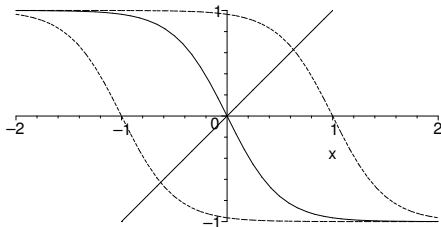


Figure 7.10:  Graphs of the map $x \mapsto \tanh(-2x + u)$ for several $u$. Obviously only a single fixed point, i.e. intersection with the line $y = x$, can exist.
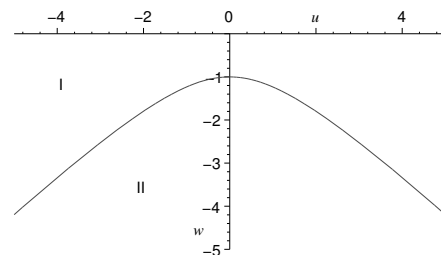
Figure 7.11:  Bifurcation manifold of the period-doubling bifurcation of the one-dimensional RNN $x \mapsto \tanh(wx + u)$.
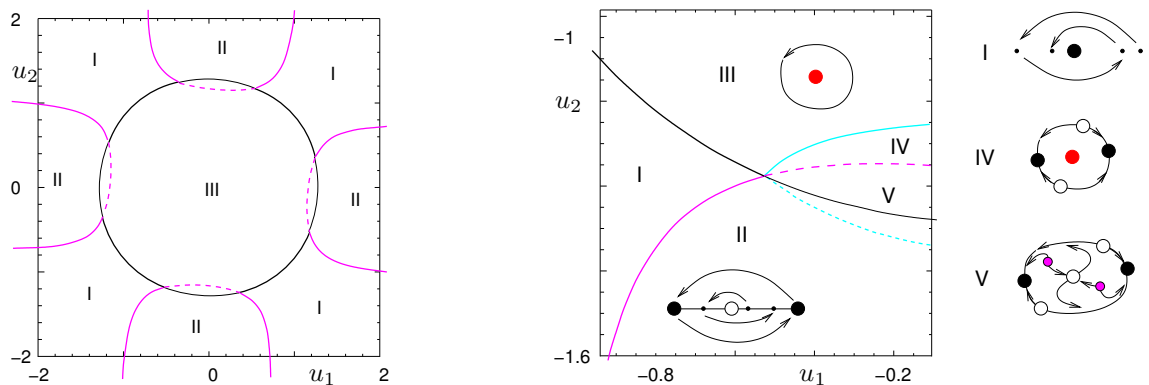


Figure 7.12: Bifurcation diagram of RNN with weight matrix $W(r = 1.2, \theta = 3)$.

bifurcation manifold of a single neuron. In region I a global asymptotically stable fixed point exists, which becomes unstable when the bifurcation curve is crossed. In region II a stable period-2 orbit exists besides the unstable node. As can be seen from both bifurcation diagrams, saddle-node bifurcations occur for weights $w \geq \frac{4}{\mu\alpha}$ only and period-doubling bifurcations occur for weights $w \leq -\frac{4}{\mu\alpha}$ only.

If $\theta$ is decreased below $\pi$, Neimark-Sacker bifurcations become possible, because in $\psi$-space exists a region $C$ again, which corresponds to fixed points with complex eigenvalues. A typical bifurcation diagram is shown in figure 7.12. Intersections between the black Neimark-Sacker curve and the magenta period-doubling curve correspond to 1:2 resonances, which are codim-2 bifurcation points. This can be seen without numerical simulations this time, because there exists a single fixed point only, which has a pair of eigenvalues satisfying $\lambda_1 \lambda_2 = 1$ along the Neimark-Sacker curve and which has a single eigenvalue $\lambda = -1$ along the period-doubling curve. Obviously at the intersection of these two curves, both eigenvalues become $-1$ simultaneously. Like for the 1:1 resonance, the 1:2 resonance points partition the Neimark-Sacker curve into sections actually corresponding to a Neimark-Sacker bifurcation (bold sections) and into sections corresponding to a neutral flip saddle (thin sections).

As in the previous cases, region I of the bifurcation diagram corresponds to a unique stable fixed point, which is a double flip node. Simulations suggest that it is even globally attracting. Crossing the solid magenta curve into region II a supercritical period-doubling bifurcation occurs, generating a stable period-2 orbit. The former stable fixed point becomes a saddle and the two branches of its unstable manifold end up at the two periodic points respectively. The corresponding phase portrait is shown in figure 7.13. Note, that the trajectories actually flip around the saddle. Because these trajectories are difficult to visualise, in fig. 7.13 the orbits of the *second iterate* of the system's mapping are shown as coloured curves, while corresponding curves have the same colour. The second iterate is embedabble into a continuous-time system, which allows us to draw continuous trajectories.

In contrast, if an input path crosses the bold black curve from region I into region III, a supercritical Neimark-Sacker bifurcation occurs, which produces an attracting closed invariant curve. The single fixed point becomes completely unstable, which is indicated by the red circle within the corresponding phase portrait. The diameter of the closed curve slowly increases as the input parameter departs from the bifurcation boundary. This is a typical property of Neimark-Sacker bifurcations (compare (3.14)). Once more, on the closed invariant curve various phase-locking phenomena occur, if the inputs are varied within region III.

Close to the dashed magenta period-doubling curve all orbits become phase-locked to an period-2 orbit. This means, that there happens a double saddle-node bifurcation of the second iterate *on* the limit cycle, which is indicated by the solid cyan curve. The two created saddle-node pairs of the second iterate correspond to a stable period-2 orbit and an unstable (saddle) period-2 orbit of the system's mapping itself. These period-2 points alternate along the closed invariant curve as shown in the phase portrait corresponding to region IV in input space. If the inputs are varied further to cross the dashed magenta curve from region IV to region V, the central flip node undergoes a subcritical period-doubling bifurcation, such that this completely unstable node becomes a flip saddle and another unstable period-2 cycle appears – indicated by the two magenta dots in the phase portrait of region V. Obviously the phase portraits of regions II and V do not match yet, particularly the disappearance of the closed invariant curve is not explained. Varying the inputs from region V to region II different global bifurcations occur, which are roughly indicated by the dotted cyan curve. A more detailed description of the 1:2 resonance is given by Kuznetsov [1995], but the complete bifurcation picture still seems to be unknown. Nevertheless, the description given here is sufficient to explain the observable dynamical behaviour. The global bifurcation
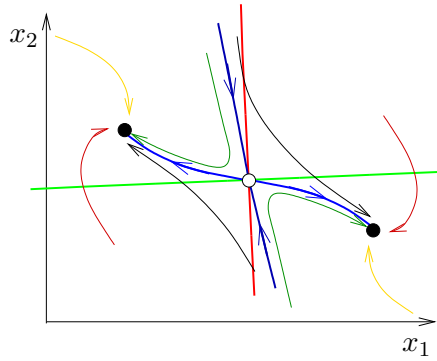
Figure 7.13: Phase portrait obtained for inputs within region II of the bifurcation diagram in figure 7.12. Note, that the orbit actually flips around the saddle fixed point – corresponding trajectories have same colours.
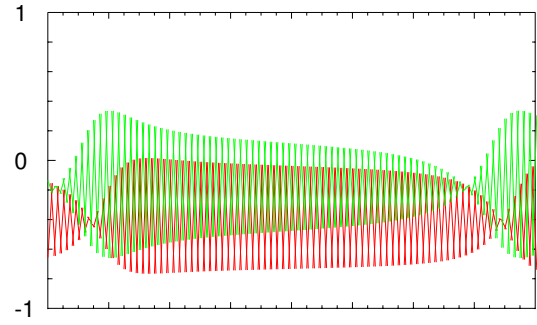
Figure 7.14: Beats occurring near the phase-locking curve.

curves seem to be very close to each other, such that they are not relevant in practice.

Finally we just mention, that we observe beats of the neural activity – as shown in figure 7.14 – if the input vector becomes close to the cyan bifurcation curve. Similarly to the observed long-period orbits near a saddle-node bifurcation on a limit cycle, the orbit spends a lot of time near the upcoming period-2 orbits on the limit cycle. As we have seen, the second iterate of the system's mapping actually undergoes a saddle-node bifurcation on the limit cycle.

### 7.3.5 Qualitative changes of bifurcation diagrams, while varying $\theta$

After the discussion of the bifurcation diagrams for some specific rotation matrices $W(r, \theta)$, we can give now a complete discussion of the overall bifurcation diagram in the parameter space $u_1$-$u_2$-$r$-$\theta$. According to the theorem proven at the beginning of this chapter, $r > 1$ is a necessary condition to observe interesting dynamical behaviour – otherwise the RNN is absolutely stable. As we already mentioned, for all $r > 1$ the sequence of different bifurcation diagrams in input space is qualitatively identical – an overview of this sequence is summarised in figure 7.15.

At the parameter value $\theta = 0$ the neurons are completely decoupled. For slightly larger parameters $\theta > 0$ we can observe orbits, which are oscillatory in the asymptotic limit $t \to \infty$, only for carefully chosen input parameters. Hence, it is difficult to observe such oscillatory orbits in practice, as long as $\theta$ is below the critical limit $\theta_{\mathrm{osc}}(r)$. The (global) bifurcation point $\theta_{\mathrm{osc}}(r)$ is defined by the condition, that all saddle-node bifurcation curves meet at the origin, i.e. the following equations are satisfied:

$$1 - r \cos(\theta_{\mathrm{osc}})(\psi_1 + \psi_2) + r^2 \psi_1 \psi_2 = 0 \,, \qquad \text{(saddle-node bifurcation condition)}$$
$$\boldsymbol{\xi} = W \boldsymbol{\sigma}(\boldsymbol{\xi}) + 0 \,, \qquad \text{(fixed point condition, } \boldsymbol{u} = 0\text{)}$$
$$\boldsymbol{\psi} = \boldsymbol{\sigma}'(\boldsymbol{\xi}) \qquad \text{(definition of } \boldsymbol{\psi}\text{)}$$

This system of nonlinear equations can be solved numerically and the resulting solution curve is shown in figure 7.16. Its slowly increasing shape guarantees a wide range of oscillatory behaviour for moderate values of $r$. In the asymptotic limit, i.e. if $r$ or all gains of the activation functions tend to infinity, the critical value $\theta_{\mathrm{osc}}(r)$ approaches $\frac{1}{4}\pi$. In any case, quasiperiodic orbits can be observed only for values $\theta > \theta_{\mathrm{osc}}$. Such orbits can only occur on the closed invariant curve, but stable nodes on this curve – which exist for values $\theta \le \theta_{\mathrm{osc}}$ – attract all trajectories starting on

(a) $\theta = 0.0$

(b) $\theta = 0.05$

(c) $\theta = 0.1$

(d) $\theta = 0.2$

(e) $\theta = 0.3$

(f) $\theta = 0.6$
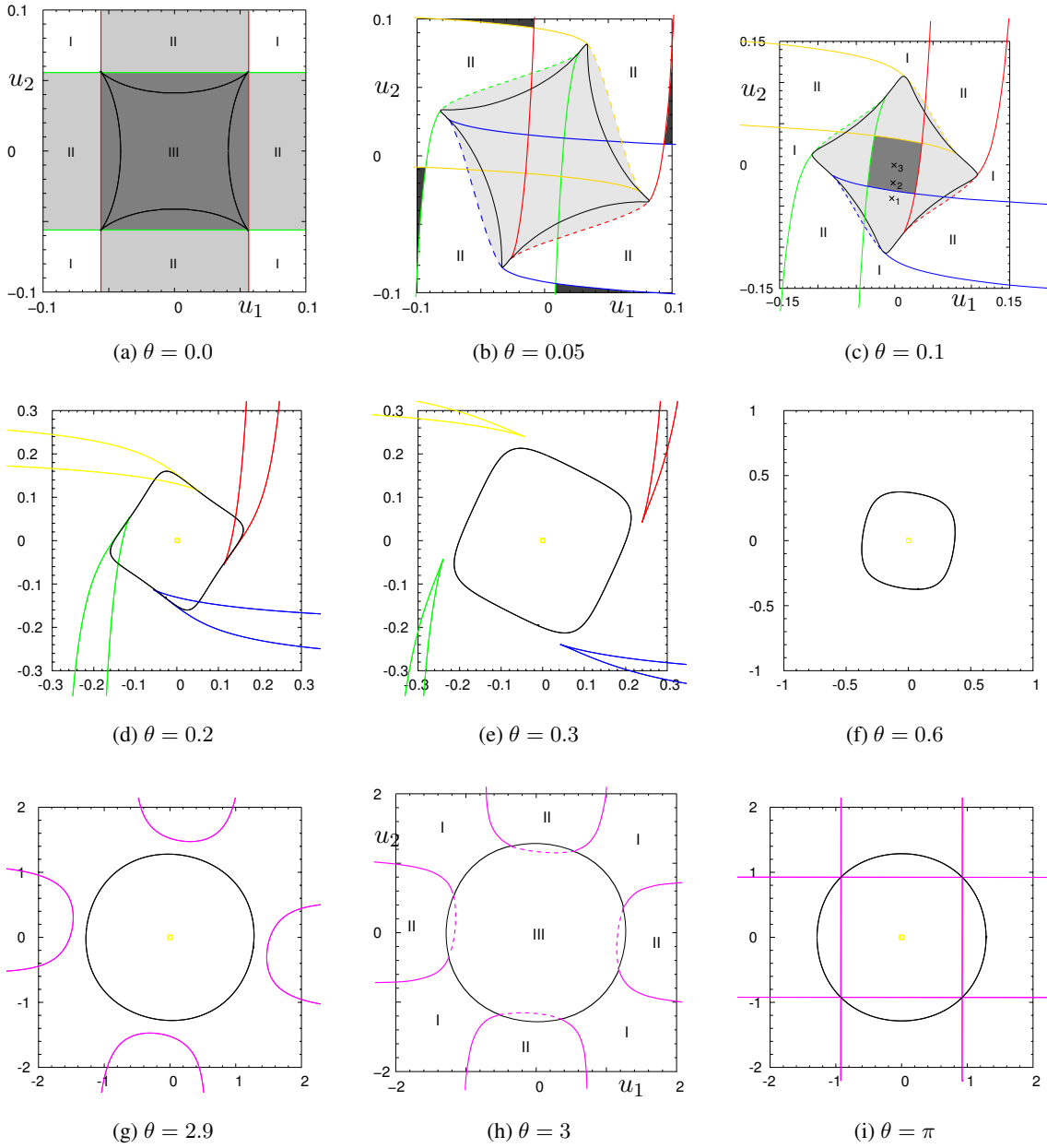
(g) $\theta = 2.9$

(h) $\theta = 3$

(i) $\theta = \pi$

Figure 7.15:  Development of the bifurcation diagrams in input space during variation of the weight matrix of the RNN: $W(r = 1.2, \theta)$, $\theta \in [0, \pi]$.
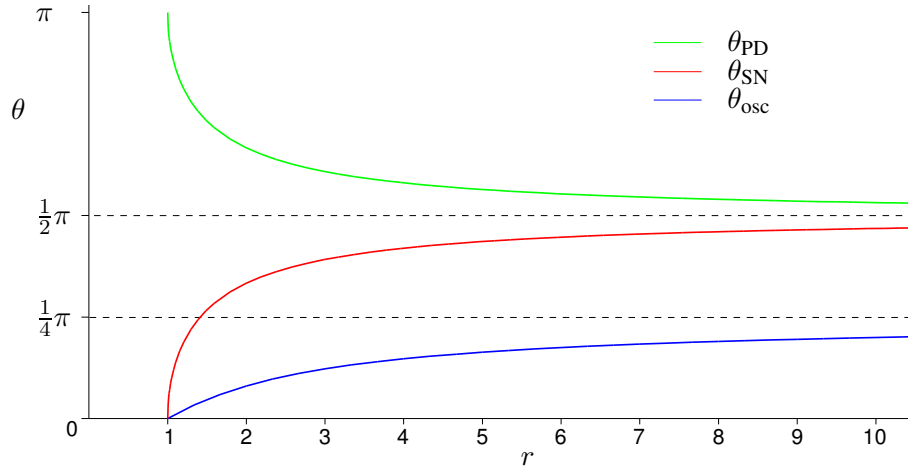
Figure 7.16: Bifurcation diagram in $(r, \theta)$-space, defining the weight matrix (7.3) of the RNN.

the curve, and thus impede the appearance of quasiperiodic orbits. Nevertheless, it is possible to observe purely periodic orbits on the closed invariant curve, which may have an arbitrary large integer period. Indeed such periodic orbits appear and disappear on the limit cycle via saddle-node bifurcations of higher iterates of the system's mapping in the same manner as stable nodes and saddles do. In figure 3.4 an example of such a pair of an unstable and stable periodic orbit on the invariant curve was presented already. Notice, that the use of the hyperbolic tangent heavily simplifies the formulation of the above implicit equations, because all saddle-node bifurcation curves meet at the origin and not at any other point in input space.

If $\theta$ is increased, the Cusp regions withdraw from the region enclosed by the Neimark-Sacker curve, such that above some threshold $\theta_{\mathrm{NS}}(r)$, there exist sections of the Neimark-Sacker curve, which directly connect region I, which corresponds to a unique stable fixed point, with region III in the centre of the input space, which corresponds to oscillatory activity of almost all trajectories (fig. 7.15d). If $\theta$ is further increased, the Cusp regions do not intersect with the Neimark-Sacker region at all (fig. 7.15e). Both regimes allow the occurrence of oscillatory orbits with arbitrary small amplitude, if input vectors are chosen, which are located arbitrarily close to the appropriate sections of the Neimark-Sacker bifurcation curve.

Increasing $\theta$ above a threshold $\theta_{\mathrm{SN}}(r)$ the Cusp regions totally disappear (fig. 7.15f), because the saddle-node bifurcation condition does not have admissible solutions for $\psi$ anymore, i.e. solutions within the range $[0, \frac{1}{4}\mu\alpha]^2$. This threshold can be computed analytically from the saddle-node bifurcation condition (6.14):

$$\theta_{\mathrm{SN}}(r) = \arccos\left(\frac{1}{r}\right) .$$

At a further critical angle $\theta_{\mathrm{PD}}(r)$, curves corresponding to period-doubling bifurcations occur in the bifurcation diagram (fig. 7.15g). They intersect with the Neimark-Sacker region for even larger $\theta$ (fig. 7.15h). Finally at $\theta = \pi$ the neurons become decoupled again, and exhibit period-doubling bifurcations only (fig. 7.15i).

The most important bifurcation points $\theta_{\mathrm{bif}}$ are shown in figure 7.16 in dependence on $r$. Note, that these curves are bifurcation curves in $(r, \theta)$-space. This figure completes the bifurcation analysis for two-dimensional RNNs, parameterised by the set $(r, \theta, u_1, u_2)$. The overall bifurcation diagram in four dimensions composes from the cross product of figure 7.16 and the corresponding input diagrams from figure 7.15.

PSfrag replacements
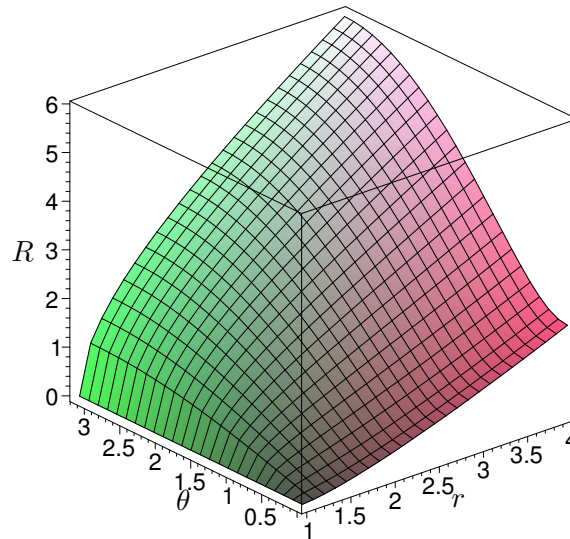$R$
$u_1$
$u_2$
$\psi_1$
$\psi_2$

Figure 7.17:   Dependence of the approximate radius $R$ of the region in input space, which corresponds to potentially oscillatory dynamical behaviour, on the weight matrix parameters $r$ and $\theta$.

As can be seen from figure 7.16 both $\theta_{\mathrm{SN}}$ and $\theta_{\mathrm{PD}}$ converge to $\frac{1}{2}\pi$ as $r$ tends to infinity. Hence, for increasing gain of the activation functions, the parameter interval of the purely oscillatory regime decreases and tends to zero, if the gain approaches infinity. Nevertheless it is possible to observe quasiperiodic as well as periodic orbits in the whole interval $[\theta_{\mathrm{osc}}(r), \pi)$. Only if the activation functions become step functions with infinite gain (e.g. Heaviside functions) quasiperiodic orbits totally disappear. As was shown by McGuire et al. [2002], in this case all oscillatory orbits have a finite integer period.

Notice in figure 7.15, that the input region enclosed by the Neimark-Sacker curve, i.e. the region which corresponds to potentially oscillatory dynamical behaviour, continuously increases with $\theta$. It turns out, that its diameter increases almost linearly with $r$. The dependence of this diameter on the parameters $r$ and $\theta$ is shown in figure 7.17.

Further notice, that the number of regions corresponding to different dynamical behaviour is maximal in the neighbourhood of the origin. Moreover, input values near zero lead most probably to oscillatory dynamics, because the Neimark-Sacker region is centred at the origin. This explains an observation made by Beer [2003]. He compared the performance of an evolution algorithm applied to randomly initialised RNNs and to center-crossing RNNs, which are obtained from randomly initialised networks by an adjustment of the biases (a fixed input contribution), such that the bifurcation manifolds become centred around the origin again. Beer employed the Fermi function as the activation function of the investigated networks, such that his center-crossing condition actually resembles equation (5.4), which transforms the inputs for an appropriate network, which employs the hyperbolic tangent as its activation functions. The evolution algorithm, which should find a specific oscillatory behaviour, performed much better when it was initialised with a population, which was seeded with these center-crossing networks.

### 7.3.6  Different types of Spiking

During the detailed bifurcation analysis in this chapter we have studied several mechanisms, how oscillatory behaviour can emerge. Such oscillatory dynamics of neurons can be understood as spiking behaviour, and in this section we want to summarise the properties of the bifurcations, which lead to such a behaviour. We will see that the different involved bifurcation types can be

employed to control different properties of the emerging oscillations – in particular their frequency and amplitude.

There exist many mathematical models to explain neural spiking and spike propagation [Hodgkin and Huxley, 1954; FitzHugh, 1969; Morris and Lecar, 1981, to mention a few]. While most of these models are formulated as differential equations and include many biological and electrophysiological details, we concentrate on the essential mechanisms, which are necessary to evoke oscillatory activity. With this approach we follow the presentation of Hoppensteadt and Izhikevich [1997], who investigated spiking in continuous-time systems. As we will see, the same mechanisms apply in discrete-time systems as well.

Hodgkin [1948] performed experiments on excitable neurons, where he applied an external current to a neuron or group of neurons and measured their spiking frequency. He found two types of excitable neurons[5], later classified by Rinzel and Ermentrout [1989]:

- Type I neural excitability: Spikes can be generated with arbitrary low frequency, depending on the strength of the applied current.

- Type II neural excitability: Spikes are generated in a certain frequency range, which is relatively insensitive to the applied current.

The external stimulus in Hodgkin's experiments acts as a bifurcation parameter, which toggles between the oscillatory and quiescent dynamical regime. Furthermore the strength of this stimulus influences the properties of the spiking activity. In our RNN model the external stimulus is represented by the external inputs $u$. A short comparison of the properties of type I and type II neural excitability with the discussed bifurcation types, shows that:

- Type I neural excitability corresponds to the occurrence of a saddle-node bifurcation on a limit cycle, while

- Type II neural excitability corresponds to the occurrence of a supercritical Neimark-Sacker bifurcation apart from Cusp regions.

As we have seen, the saddle-node bifurcation on a limit cycle allows for large amplitude oscillations with a frequency that varies smoothly over a range from zero to $\theta$ approximately, where small frequencies are observed close to the saddle-node bifurcation curve and maximal frequencies are observed far away from it, i.e. at the origin in input space if the hyperbolic tangent is used as activation function (see fig. 7.18a).

In contrast, a classical Neimark-Sacker bifurcation immediately leads to an oscillation with a high frequency, which is given by the argument of the complex eigenvalues of the Jacobian at the nonhyperbolic fixed point. In the example studied within this section, this frequency equals approximately $\theta$ and varies with the distance of the Neimark-Sacker bifurcation curve to the origin in input space [Haschke et al., 2001]. With this bifurcation type the amplitude of the oscillation smoothly increases from zero and the frequency keeps relatively constant (fig. 7.18b).

Remember, that the observed low frequency with Type I spiking originates from a nearly constant orbit, which is intermitted by short spikes (compare fig. 7.9c). Thus the corresponding oscillation is far away from being harmonically. If we want to smoothly control the frequency of a harmonic oscillation, we have to modify the weight matrix itself, particularly the parameter $\theta$, which denotes the average rotation angle during one iteration.

---

[5] Hodgkin also identified a third class of neural excitability: those neurons which could create action potentials with difficulty or not at all. These neurons are far away from a bifurcation point and are not considered here.

(a) Type I excitability: saddle-node bifurcation on a limit cycle ($r = 4.2$, $\theta = 1$)



(b) Type II excitability: classical Neimark-Sacker bifurcation ($r = 2.8$, $\theta = 1$)
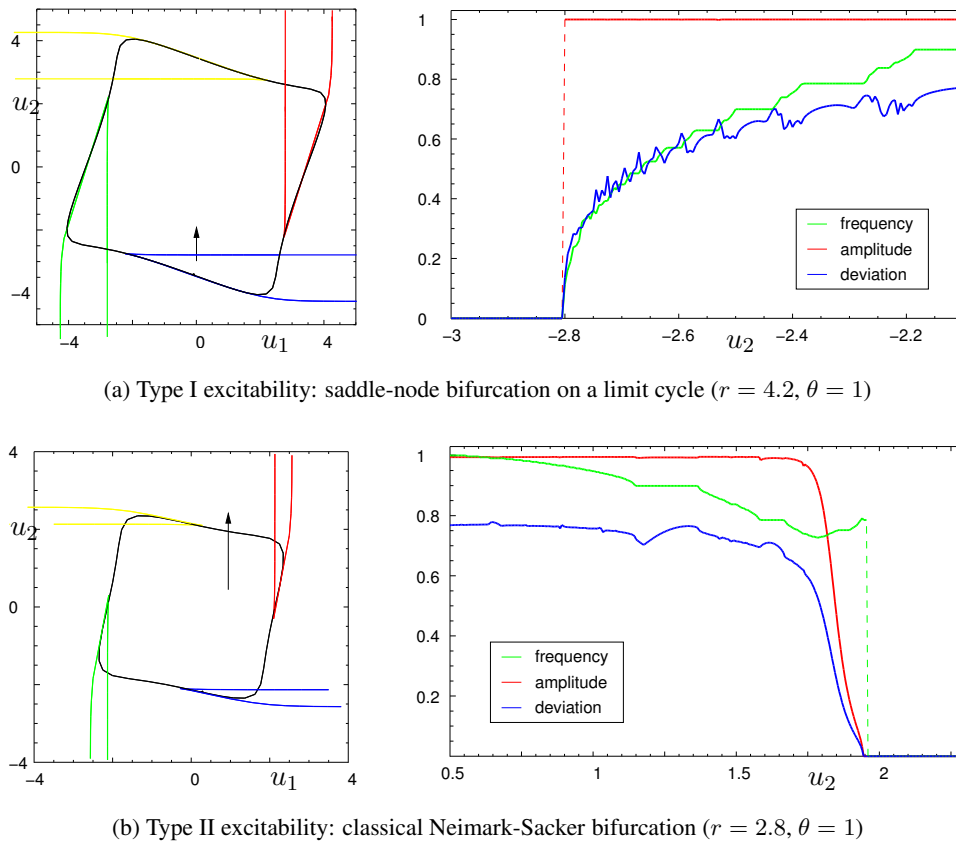
Figure 7.18: Evolution of the oscillation frequency and amplitude along the paths shown in the bifurcation diagrams (left column). Notice the sudden jump of amplitude (a) and frequency (b) in dependence of the occurring bifurcation type. We have plotted the amplitude as well as the standard deviation of the time series from its average value. This latter value decreases to zero for type I excitability as well, because the trajectory becomes nearly constant in the vicinity of the bifurcation point.

## 7.4 Bifurcations of periodic orbits

We already mentioned, that it is not possible to predict the whole repertoire of dynamical behaviour on the basis of fixed point bifurcations alone, because in this case we do not take into account bifurcation manifolds, which correspond to bifurcations of periodic orbits or to global bifurcations. These bifurcation manifolds cannot be computed analytically yet. Nevertheless, simulations of all networks, which we considered in the previous section, suggest that there do not exist any additional bifurcation manifolds in these examples.

Now we want to present an example, where bifurcations of 2-cycles occur besides normal fixed point bifurcations. The weight matrix under consideration is shown in figure 7.20 together with its bifurcation diagram in input space. The only fixed point bifurcations, which are exhibited by the corresponding network, are of the period-doubling type and the associated bifurcation curve is coloured magenta. Additionally we observe saddle-node bifurcations of the second iterate, i.e. saddle node bifurcations, which generate a stable and an unstable period-2 orbit. The corresponding cyan coloured bifurcation curves additionally separate the regions III from the globally stable region I.

We know that the system is globally asymptotically stable in region I and the corresponding unique fixed point can loose its stability along the magenta curve only, because there do not occur any
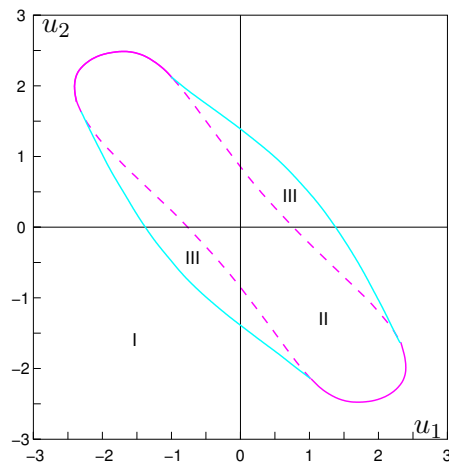
(a) bifurcation diagram

$$W = \begin{pmatrix} -0.8 & 1.3 \\ 1.4 & -0.5 \end{pmatrix}$$

(b) weight matrix

Figure 7.20: Bifurcation diagram and weight matrix of a network, which exhibits a fixed point bifurcation of the second iterate of the system's mapping. Additionally to the fixed point bifurcation curve (magenta) bifurcation manifolds of periodic orbits (cyan) are shown.

other fixed point bifurcations. Nevertheless we cannot conclude global stability outside region II, because there exists the additional cyan saddle-node bifurcation curve of the second iterate. It is even possible, that bifurcation manifolds of higher-period orbits further restrict the globally stable region. This example should clarify, that the bifurcation manifolds of fixed points form a small, but important part of the overall bifurcation diagram. They cannot explain the whole repertoire of the dynamics in general.

The cyan saddle-node bifurcation curves meet the period-doubling curves at codim-2 bifurcation points and the corresponding bifurcation type is called degenerate or generalised period-doubling bifurcation. The dynamical behaviour in the vicinity of such points is explained by Kuznetsov [1995] and looks as follows: In region I the network is globally asymptotically stable. Crossing the cyan curve into region III a stable and an unstable period-2 orbit are born through a saddle-node bifurcation of the second iterate, while the unique fixed point remains stable. If the inputs vary further from region III to region II and cross the dashed magenta curve, the stable node undergoes a subcritical period-doubling bifurcation, where the unstable period-2 orbit is destroyed again (compare with section 3.2.3). The remaining unstable fixed point and the stable period-2 orbit coalesce along the solid magenta curve, along which a supercritical period-doubling bifurcation occurs.

If the inputs are varied along a path from region I to region II and back again, crossing region III each time, we observe a hysteresis effect again. The unique stable node existing in region I looses its stability along the dashed magenta curve and the trajectory becomes attracted to the 2-cycle. Going back to region I this oscillatory behaviour remains stable until the cyan bifurcation curve is crossed. Hence region III is an hysteresis domain.

# 8 Bifurcations in High-Dimensional Neural Networks

As we have seen in the previous sections, it is not easily possible to compute bifurcation manifolds for networks of more than three neurons – mainly due to the exponentially increasing number of different branches. Nevertheless we can try to infer dynamical properties of a large network by studying its smaller subnetworks. We consider two different approaches to this reduction of dimension:

- low-dimensional cross sections in the high-dimensional space and

- cascaded, i.e. feedforward structured RNNs.

Both approaches aim at a simplification of the bifurcation conditions, i.e. the test functions in table 6.1, which are determinant equations involving the Jacobian matrix and which yield polynomials in $\psi$, whose coefficients in turn are polynomials in terms of the weight connections $w_{ij}$. While the first approach corresponds to fixing some components of $\psi$ to constant values, the second approach employs appropriately structured weight matrices, such that many coefficients of the polynomials in $\psi$ vanish.

## 8.1 Cross sections

A first attempt to reduce the complexity of the test function polynomials is to regard only a few of the $\psi$-components as free parameters and to fix all other components to some desired value. As we already pointed out in section 6, there exists a direct relation (the one-to-many correspondence (6.13)) between the *components* of $\psi$ and those of the fixed point vector $\bar{x}$. Thus fixing $\psi$-components is equivalent to fixing $\bar{x}$-components and vice versa. Such a fixing of components induces a cross section within the appropriate spaces, which is defined by a hyperplane through the fixed components.

According to (6.12) the corresponding input vector undergoes the following nonlinear transformation with respect to $\psi$ or rather $\xi$ which depends component-wise on $\psi$:

$$\boldsymbol{u} = \boldsymbol{\xi} - W\boldsymbol{\sigma}(\boldsymbol{\xi}) \quad \text{with } \boldsymbol{\xi} \in Z(\boldsymbol{\psi}) = \boldsymbol{\sigma'}^{-1}(\boldsymbol{\psi}). \tag{6.12'}$$

Consequently, for general weight matrices $W$ there exists neither a component-wise relationship between $\psi$ and $\boldsymbol{u}$ nor a hyperplane in $\boldsymbol{u}$-space, which corresponds to the cross sections in $\psi$-space or in state space. The component-wise relationship is annihilated by the linear mixing matrix $W$ and the nonlinearity of $\boldsymbol{\sigma}$ causes a deformation of the hyperplane, such that input vectors corresponding to a hyperplane in $\psi$-space are located on a curved manifold in $\boldsymbol{u}$-space.

**Example 8.1** For illustration consider the three-neuron network having the weight matrix

$$W = \begin{pmatrix} 0.7 & -1.1 & 0.7 \\ 1.1 & 0.7 & -0.25 \\ -1.40 & -0.6 & 0.4 \end{pmatrix} \tag{8.1}$$

and the hyperbolic tangent as activation function for all neurons. Further, let $\psi_3$ be fixed at $\psi_3 = 0.94$, which corresponds to fixing $x_3 = \pm\sqrt{1 - \psi_3} = \pm 0.25$. The resulting bifurcation manifolds

are depicted in figure 8.1. The bifurcation curves obtained from fixing $\psi_3$ are visualised as magenta boxes along the two-dimensional bifurcation manifolds. Notice that these curves lie at planes in $\psi$- and $x$-space but not in $u$-space.

In the following, we investigate the bifurcation conditions in more detail if some $\psi$-components are fixed. Particularly we answer the question, whether any conclusions can be drawn for the whole RNN from the dynamical behaviour of subnetworks, which are obtained from the original network by removing single neurons and their connections, i.e. if we consider submatrices of the original weight matrix.

For this purpose we split all vectors into a fixed part – indicated by a subscript $c$ – and a variable part – indicated by a subscript $v$. With appropriate splittings of the weight matrix $W$ we obtain the dynamics:

$$
\begin{bmatrix} \boldsymbol{x}_v \\ \boldsymbol{x}_c \end{bmatrix} \mapsto \boldsymbol{\sigma} \left( \begin{bmatrix} W_{vv} & W_{vc} \\ W_{cv} & W_{cc} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_v \\ \boldsymbol{x}_c \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_v \\ \boldsymbol{u}_c \end{bmatrix} \right) = \boldsymbol{\sigma} \left( \begin{bmatrix} W_{vv}\boldsymbol{x}_v + W_{vc}\boldsymbol{x}_c + \boldsymbol{u}_v \\ W_{cv}\boldsymbol{x}_v + W_{cc}\boldsymbol{x}_c + \boldsymbol{u}_c \end{bmatrix} \right) . \tag{8.2}
$$

Here $W_{vv}$ is the weight matrix of the resulting low-dimensional subnetwork. In order to answer the question, whether the knowledge of bifurcation manifolds of this subnetwork allows to draw conclusions for the whole network as well, we have to consider the Jacobian matrix $J(\bar{\boldsymbol{x}})$, which is needed for all bifurcation conditions. Using the notation $D_{v,c}(\bar{\boldsymbol{x}}_{v,c}) = \mathrm{diag}\big(\boldsymbol{\psi}_{v,c}(\bar{\boldsymbol{x}}_{v,c})\big)$ the Jacobian can be written as

$$
J(\bar{\boldsymbol{x}}) = D(\bar{\boldsymbol{x}})W = \begin{bmatrix} D_v W_{vv} & D_v W_{vc} \\ D_c W_{cv} & D_c W_{cc} \end{bmatrix} . \tag{8.3}
$$

In theorem A.4 of appendix A.3 we prove, that the bialternate product preserves the block structure of its multipliers to some degree, such that we obtain the following expression for $J(\bar{\boldsymbol{x}}) \odot J(\bar{\boldsymbol{x}})$:

$$
J(\bar{\boldsymbol{x}}) \odot J(\bar{\boldsymbol{x}}) = P \begin{bmatrix} D_v \odot D_v & & \\ & D_c \otimes D_v & \\ & & D_c \odot D_c \end{bmatrix} \begin{bmatrix} W_{vv} \odot W_{vv} & X_{\hat{v},cv} & W_{vc} \odot W_{vc} \\ X_{cv,\hat{v}} & X_{cv,cv} & X_{cv,\hat{c}} \\ W_{cv} \odot W_{cv} & X_{\hat{c},cv} & W_{cc} \odot W_{cc} \end{bmatrix} P^{-1} ,
$$
$$
\tag{8.4}
$$

where $P$ is a permutation matrix and $X_{a,b}$ denote matrices of dimension $a \times b$ with $\hat{v} = \frac{1}{2}v(v-1)$ and $\hat{c} = \frac{1}{2}c(c-1)$. The numbers $v$ and $c$ denote the number of free resp. fixed $\psi$-components. Employing the specific structure of these matrices we prove the following theorem:

**Theorem 8.2** Consider the structured RNN (8.2) with a general weight matrix $W$ and fixed state components $\boldsymbol{x}_c$. The bifurcation manifolds of the lower dimensional RNN

$$
\boldsymbol{x}_v \mapsto \boldsymbol{\sigma}(W_{vv}\boldsymbol{x}_v + \boldsymbol{u}_v') \tag{8.5}
$$

resemble appropriate submanifolds of the original RNN's bifurcation manifolds if and only if the fixed state components $\boldsymbol{x}_c$ are clamped at values $\boldsymbol{x}_c = \boldsymbol{\sigma}(\boldsymbol{\xi}_c)$ where $\boldsymbol{\xi}_c$ are critical values of $\boldsymbol{\sigma}$, i.e. values $\boldsymbol{\xi}_c$ such that $\boldsymbol{\sigma}'(\boldsymbol{\xi}_c) = 0$. Particularly this means that $\boldsymbol{\psi}_c$ equals zero.

In this case, the corresponding submanifold on the original RNN's bifurcation manifold is given by

$$
\boldsymbol{u}_v = \boldsymbol{u}_v' - W_{vc}\bar{\boldsymbol{x}}_c \qquad\qquad = \boldsymbol{u}_v'(\boldsymbol{\psi}_v) - W_{vc}\bar{\boldsymbol{x}}_c(\boldsymbol{\psi}_c) \tag{8.6}
$$

$$
\text{and} \quad \boldsymbol{u}_c = \boldsymbol{\xi}_c - W_{cv}\bar{\boldsymbol{x}}_v - W_{cc}\bar{\boldsymbol{x}}_c = f(\boldsymbol{\psi}_c) - W_{cv}\bar{\boldsymbol{x}}_v(\boldsymbol{\psi}_v) . \tag{8.7}
$$

PSfrag replacements $z$ 0.500

$u_1$
$u_2$
$\psi_1$
$\psi_2$

PSfrag replacements

$u_1$
$u_2$
$\psi_1$
$\psi_2$

(a) $\boldsymbol{\psi}$-space

(b) state space

PSfrag replacements

$u_1$
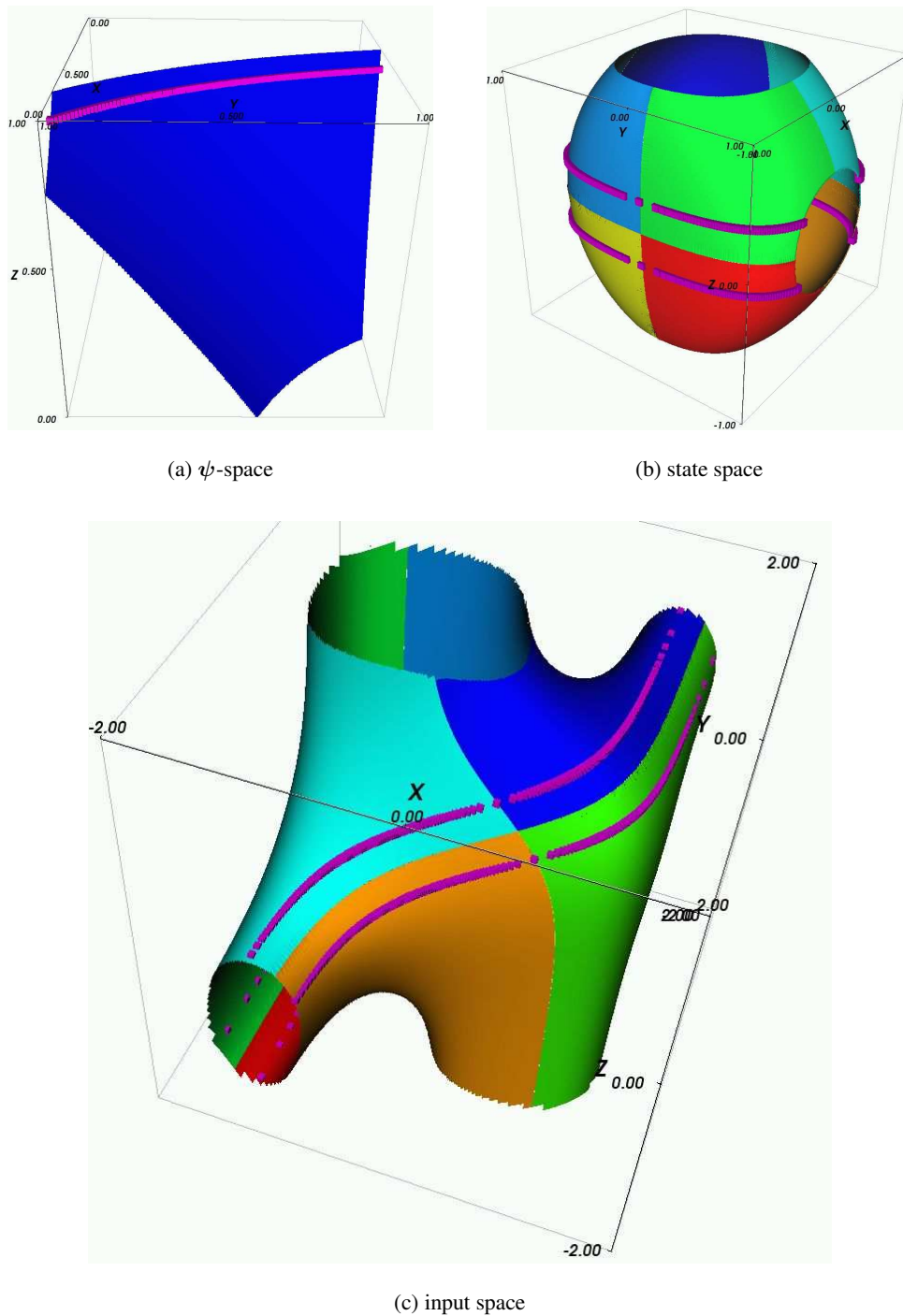$u_2$
$\psi_1$
$\psi_2$

(c) input space

Figure 8.1: Bifurcation manifolds of a three-neuron network with weight matrix (8.1). Bifurcation curves obtained from fixing $\psi_3 = 0.94$ (or equivalently $x_3 = \pm 0.25$) are shown as magenta boxes along the bifurcation manifolds. Notice that these curves are located along planes in $\boldsymbol{\psi}$- and $\boldsymbol{x}$-space but not in $\boldsymbol{u}$-space.

**Proof**    The bifurcation conditions for saddle node and period doubling bifurcations have the form

$$\det(J \mp \mathbf{1}) = 0 \qquad \text{and} \qquad \det(J_{vv} \mp \mathbf{1}) = 0 \,,$$

while the bifurcation condition for a Neimark-Sacker bifurcation has the form

$$\det(J \odot J - \mathbf{1}) = 0 \qquad \text{and} \qquad \det(J_{vv} \odot J_{vv} - \mathbf{1}) = 0$$

for the complete RNN and the subnetwork respectively. Inserting the expressions for $J$ and $J \odot J$ from (8.3) resp. (8.4) and employing the determinant expansion theorem it can be easily seen, that the bifurcation conditions become identical if and only if all but the first rows or all but the first columns of $J$ resp. $J \odot J$ are zero, i.e.

$$J \mp \mathbf{1} = \begin{pmatrix} J_{vv} \mp \mathbf{1} & 0 \\ D_c W_{cv} & \mp \mathbf{1} \end{pmatrix} \qquad \text{or} \qquad J \mp \mathbf{1} = \begin{pmatrix} J_{vv} \mp \mathbf{1} & D_v W_{vc} \\ 0 & \mp \mathbf{1} \end{pmatrix}$$

in case of saddle-node and period-doubling bifurcations and

$$J \odot J - \mathbf{1} = P \begin{bmatrix} D_v \odot D_v\, W_{vv} \odot W_{vv} - \mathbf{1} & D_v \odot D_v\, X_{\hat{v},cv} & D_v \odot D_v\, W_{vc} \odot W_{vc} \\ 0 & -\mathbf{1} & D_c \otimes D_v\, X_{cv,\hat{c}} \\ 0 & 0 & -\mathbf{1} \end{bmatrix} P^{-1}$$

or

$$J \odot J - \mathbf{1} = P \begin{bmatrix} D_v \odot D_v\, W_{vv} \odot W_{vv} - \mathbf{1} & 0 & 0 \\ D_c \otimes D_v\, X_{cv,\hat{v}} & -\mathbf{1} & 0 \\ D_c \odot D_c\, W_{cv} \odot W_{cv} & D_c \odot D_c\, X_{\hat{c},cv} & -\mathbf{1} \end{bmatrix} P^{-1}$$

in case of a Neimark-Sacker bifurcation. Performing a comparison of coefficients with respect to the free variables $\psi_v$, it can be seen that these conditions are fulfilled if and only if $D_c$ becomes identical to zero, which means that all fixed $\psi$-components $\psi_c$ equal zero. Employing the definition $\psi := \sigma'(\xi)$ and equations (6.6–6.12) the assertion follows immediately. $\square$

**Remarks:**

- Typical sigmoidal activation functions – for example functions from $\mathcal{S}_0$ – have exactly two critical points: $\xi = \pm\infty$. Thus the state components $\bar{x}_c$ have to take their saturation values in order to obtain the same bifurcation manifolds for the considered low dimensional system and the complete system. According to (8.7) this means that the bifurcation manifold of the low dimensional subsystem resembles the cross sections of the high dimensional bifurcation manifold, which are obtained if the input components $u_c$ corresponding to fixed components tend to infinity in absolute value, i.e. if $u_c \to \pm\infty$.

- The following naive approach to draw conclusions from low dimensional subnetworks for the complete network turns out to be wrong: In equation (8.2) we might fix $\bar{x}_c$ at some arbitrary values, i.e. consider the appropriate cross section. Then we can consider bifurcation manifolds of the low dimensional subsystem (8.5). Employing (8.7), the corresponding fixed points $\bar{x}_v$ can be used to find an appropriate input vector $u_c$ for the second part of the RNN, such that $(\bar{x}_v, \bar{x}_c)$ actually *is* a fixed point of the complete RNN.

  Thus, by construction the state vector $\bar{x} = [\bar{x}_v^t, \bar{x}_c^t]^t$ is a fixed point of the complete RNN, if the correct input vector is applied. But this fixed point is not necessarily hyperbolic any

more, i.e. it does not undergo a bifurcation. As theorem 8.2 reveals this is the case for those special values of $\boldsymbol{x}_c$ only, which correspond to critical values of $\boldsymbol{\sigma}$. Indeed this naive approach ignores the more complex bifurcation condition of the high dimensional network, but meets the appropriate condition for the subnetwork only.

From theorem 8.2 we can immediately derive the following corollary, which highlights the special importance of the self feedback in RNNs.

**Corollary 8.3** If a self-feedback weight $w_{ii}$ of an arbitrary RNN with sigmoid activation functions within class $\mathcal{S}_0$, satisfies $|w_{ii}| > \frac{4}{\mu\alpha}$, the corresponding state component $x_i$ undergoes a saddle-node ($w_{ii} > 0$) or period-doubling ($w_{ii} < 0$) bifurcation under variation of $u_i$, if all other input components are fixed at large values, i.e. $u_{j \neq i} \to \pm\infty$.

**Proof** If $|u_{j \neq i}|$ tend to infinity, the appropriate state components are driven into the saturation regime and the derivatives $\psi_{j \neq i}$ vanish. Hence the networks dynamics reduces to the one-dimensional dynamics

$$x_i \mapsto \sigma\left(w_{ii}x_i + u_i + u_{\text{ext}}\right),$$

where $u_{\text{ext}} = W_{vc}\bar{\boldsymbol{x}}_c$ summarises all external influences from other neurons. According to figures 3.2 and 7.11 the single neuron undergoes a saddle-node or period-doubling bifurcation under the given assumptions. $\square$

The corollary implies, that a network with a self-coupling weight $|w_{ii}| > \frac{4}{\mu\alpha}$ cannot be absolutely stable – independently of all other weights. Particularly, the weight matrix itself might be Schur stable, but it cannot be **D**-stable in this case, which would be a contradiction to theorem 4.3 of chapter 4.

In a similar manner we can proof the existence of Neimark-Sacker bifurcations and their corresponding oscillations, if the interconnection scheme of any two neurons of a large RNN resembles any appropriate rotation matrix, which we considered in chapter 7.

Notice, that the width of the Cusp domain or the oscillation domain in input space increases with the magnitude of the self-coupling weight according to figures 3.2a and 7.11. Particularly interesting for applications are very thin as well as very broad Cusp regions. Thin cusp regions on the one hand cause a large state change (from one saturation level to the opposite) due to small input variations, i.e. they act as an amplifier. On the other hand, broad cusp regions are characterised by wide hysteresis domains, such that they can be employed as working memory, which is very robust to small input fluctuations, but has small transition times from one state to the other as soon as the input variation becomes large enough. The sensitivity to the input variations can be adjusted by the strength of the self-connection weight, which in turn controls the width of the cusp and hysteresis domain. In evolutionary programming experiments Nakahara and Doya [1998] have shown, that networks, which are optimised for a task that requires efficient use of memory, indeed work on the edge of a cusp domain.

## 8.2 Cascaded RNNs

Even if it is not possible to draw conclusions from low to high dimensional systems in general cases, there exist *special* weight matrices, such that $\psi_c$ can be fixed at *arbitrary* values as well. As we see in the following, such weight matrices have a *block triangular shape* and the corresponding networks are called *cascaded* networks. The notion of cascades was introduced by Hirsch [1989]

in order to study global asymptotic stability of complex networks in terms of smaller subnets. We will extend this analysis to all types of fixed point bifurcations of such reducible networks, more concretely we prove the following theorem:

**Theorem 8.4** Given a cascaded RNN, the bifurcation conditions of fixed points for saddle-node, period-doubling and Neimark-Sacker bifurcations split into the corresponding conditions of its recurrent subnets along the block diagonal.

**Proof**    Without loss of generality it suffices to consider a cascaded network composed of two recurrent networks with weight matrices $A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{m \times m}$, which are interconnected by the third matrix $B \in \mathbb{R}^{n \times m}$. Hence, the cascaded network possesses the weight matrix

$$W = \begin{pmatrix} A & B \\ & C \end{pmatrix}.$$

Using equation (8.3), the necessary conditions for a saddle-node or period-doubling bifurcation at a fixed point $\bar{x}$ can be written as:

$$0 = \det(J(\bar{x}) \mp \mathbf{1}) = \det(D(\bar{x})W \mp \mathbf{1}) = \det(D_n(\bar{x})A \mp \mathbf{1}) \, \det(D_m(\bar{x})C \mp \mathbf{1}), \quad (8.8)$$

where $D_n(\bar{x}) = \mathrm{diag}(\psi_1(\bar{x}), \ldots, \psi_n(\bar{x}))$ and $D_m(\bar{x}) = \mathrm{diag}(\psi_{n+1}(\bar{x}), \ldots, \psi_{n+m}(\bar{x}))$ denote the restriction of the diagonal matrix $D(\bar{x})$ of derivatives at the fixed point to its $n$- and $m$-dimensional submatrices respectively. Obviously these bifurcation conditions are fulfilled if and only if the appropriate bifurcation condition is fulfilled for one of the subnetworks.

The proof of the Neimark-Sacker case is slightly more involved because of the more complex structure of the bialternate product matrix (8.4), which in this case has the form:

$$W \odot W = P \begin{pmatrix} A \odot A & X_{\hat{n},mn} & B \odot B \\ & C \otimes A & X_{mn,\hat{m}} \\ & & C \odot C \end{pmatrix} P^{-1}.$$

Again $X_{a,b}$ denote matrices of dimension $a \times b$ with $\hat{n} = \frac{1}{2}n(n-1)$ and $\hat{m} = \frac{1}{2}m(m-1)$. Consequently, the Neimark-Sacker bifurcation condition becomes:

$$
\begin{aligned}
0 &= \det\big((D \odot D)(W \odot W) - \mathbf{1}\big) \\
&= \det\big(P^{-1} D \odot D P P^{-1} W \odot W P - \mathbf{1}\big) \\
&= \det\big(D_n \odot D_n A \odot A - \mathbf{1}\big) \, \det\big(D_m \odot D_m C \odot C - \mathbf{1}\big) \, \det\big(D_m \otimes D_n C \otimes A - \mathbf{1}\big) \\
&= \det\big((D_n A) \odot (D_n A) - \mathbf{1}\big) \, \det\big((D_m C) \odot (D_m C) - \mathbf{1}\big) \, \det\big((D_m C) \otimes (D_n A) - \mathbf{1}\big).
\end{aligned}
$$

Once more, the bifurcation condition composes from the product of the Neimark-Sacker conditions corresponding to the subnetworks and the additional condition

$$\det\big((D_m C) \otimes (D_n A) - \mathbf{1}\big) = 0.$$

It remains to prove that this additional condition does not introduce new bifurcation manifolds. Remember, that the Neimark-Sacker condition may introduce spurious manifolds due to neutral saddles, which do not correspond to actual bifurcation points. A true Neimark-Sacker bifurcation is only observable if a pair of conjugate-complex eigenvalues of the Jacobian crosses the unit circle simultaneously. Hence, we have to study the eigenvalues satisfying the additional condition. Let us denote the eigenvalues of the Jacobians $J_n = D_n A$ and $J_m = D_m C$ with $\lambda_i$ and $\mu_j$ respectively. Then the additional condition is equivalent to [Lancaster and Tismenetsky, 1985]:

$$\det\big((D_m C) \otimes (D_n A) - \mathbf{1}\big) = \prod_{i,j}(\lambda_i \mu_j - 1) = 0. \quad (8.9)$$

Assume a pair of eigenvalues $\lambda_i$ and $\mu_j$ satisfy this condition, i.e. $\lambda_i \mu_j = 1$. Assume further, that $\lambda_i$ and $\mu_j$ is indeed a pair of complex-conjugate eigenvalues, which crosses the unit circle, i.e. $\lambda_i = \mathrm{e}^{\mathrm{i}\varphi}$ and $\mu_j = \mathrm{e}^{-\mathrm{i}\varphi}$, where $\varphi \in (0, \pi)$. Only in this case, they cause an additional Neimark-Sacker bifurcation. Because the weight matrices and thus the Jacobians are real matrices, there have to exist conjugate eigenvalues $\bar{\lambda}_i = \mu_j$ and $\bar{\mu}_j = \lambda_i$ of the Jacobians $J_n$ and $J_m$ respectively, such that the Neimark-Sacker bifurcation conditions of the subnetworks are already satisfied for the pairs $(\lambda_i, \bar{\lambda}_i)$ and $(\mu_j, \bar{\mu}_j)$ respectively. Consequently, if the additional condition (8.9) is satisfied for a pair of complex-conjugate eigenvalues, we have a *double* Neimark-Sacker bifurcation at hand. If the eigenvalues $\lambda_i$ and $\mu_j$ are real, they do not cause a bifurcation at all. In any case, the additional condition does not introduce new solutions. $\square$

**Remarks:**

- The independence of the bifurcation conditions corresponding to single subnetworks implies the independence of their corresponding $\boldsymbol{\psi}$-components ($D_n$ and $D_m$ in the proof). This in turn implies that we can fix $\boldsymbol{\psi}$-components corresponding to single subnetworks at arbitrary values and thus study the associated cross sections in $\boldsymbol{\psi}$- resp. $\bar{\boldsymbol{x}}$-space. Resuming to the notation from the previous section, i.e. denoting the fixed and variable vector components by a subscript $c$ resp. $v$, this means that each cross section yields the same low dimensional bifurcation manifolds in $\boldsymbol{\psi}$-space for any given $\boldsymbol{\psi}_c$. Furthermore these manifolds are defined by the low dimensional subsystem $W_{vv}$ only.

  We can distinguish two possibilities of subnetwork coupling:

  – feedforward, i.e. from $\boldsymbol{x}_v$ to $\boldsymbol{x}_c$ ($W_{vc} = 0$, $W_{cv} \neq 0$), and
  – feedback, i.e. from $\boldsymbol{x}_c$ to $\boldsymbol{x}_v$ ($W_{vc} \neq 0$, $W_{cv} = 0$).

  According to (8.6) we have $\boldsymbol{u}_v = \boldsymbol{u}'_v$ in the first case, i.e. each cross section along $\boldsymbol{u}_c$ in input space yields exactly the same bifurcation manifolds as the low-dimensional subnetwork $W_{vv}$. Note, that these cross sections in input space have no relation to any cross section in $\boldsymbol{\psi}$-space, because $\boldsymbol{u}_c$ depends on $\boldsymbol{x}_v$ which in turn depends nonlinearly on $\boldsymbol{u}_v$.

  In the second case the low dimensional bifurcation manifolds – represented by $\boldsymbol{u}'_v$ – are shifted in dependence of the fixed point components $\boldsymbol{x}_c$. Now $\boldsymbol{u}_c$ is a function of $\boldsymbol{x}_c$ only, such that cross sections along $\boldsymbol{u}_c$ do relate to cross section along $\boldsymbol{\psi}_c$ at appropriate values. Note, that there might exist several fixed points $\boldsymbol{x}_c$ at a given input vector $\boldsymbol{u}_c$, which would imply several copies of the low-dimensional bifurcation manifold within the cross section at $\boldsymbol{u}_c$.

- Notice, that the off-diagonal matrices do not enter the bifurcation conditions at all. Hence, the weights interconnecting single subnetworks do not influence the bifurcation manifolds in $\boldsymbol{\psi}$-space – but they distort the corresponding manifolds in input space according to (8.6) and (8.7). Only if block matrices both below and above the diagonal become nonzero, i.e. if the subnets are recurrently coupled, the bifurcation conditions change, as we have seen in the previous section. This allows us to prove the following corollary:

**Corollary 8.5** Given a cascaded RNN, the bifurcation conditions for saddle-node, period-doubling and Neimark-Sacker bifurcations for all *periodic* orbits split into the corresponding conditions of its recurrent subnets along the block diagonal.

**Proof**   As we discussed in section 2.2.4 the bifurcation conditions of a periodic orbit of period $T$ are determined by the Jacobian of the $T$-th iterate of the system's map:

$$J = D(\bar{\boldsymbol{x}}_T)W \, \cdots \, D(\bar{\boldsymbol{x}}_2)W \, D(\bar{\boldsymbol{x}}_1)W \, .$$

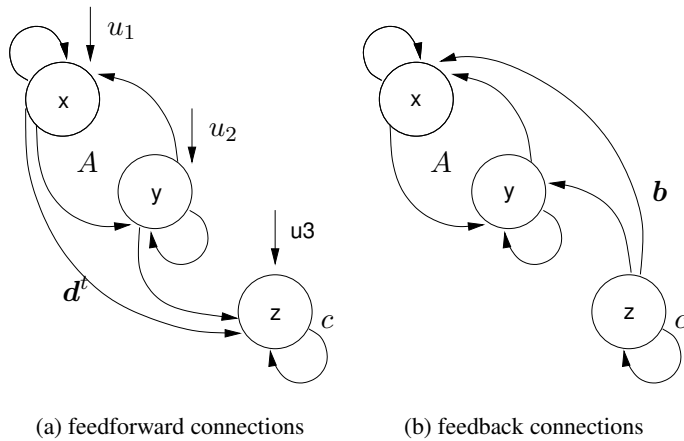(a) feedforward connections  (b) feedback connections

Figure 8.2: Cascaded networks composed of a two-neuron network and a single neuron, which are strongly coupled through feedforward (a) or feedback (b) connections.
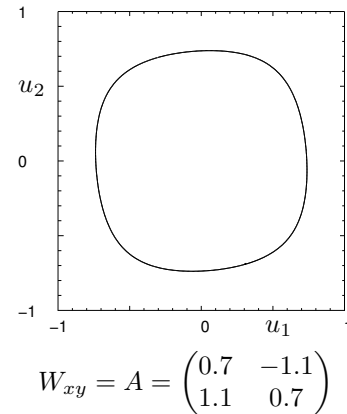


$$W_{xy} = A = \begin{pmatrix} 0.7 & -1.1 \\ 1.1 & 0.7 \end{pmatrix}$$

Figure 8.3: Neimark-Sacker bifurcation curve of the subnetwork with weights $A$.

The bifurcation conditions themselves remain unchanged. Clearly the block structure of $W$ is inherited by all matrices $D(\bar{x}_i)W$, $i = 1, \ldots, T$ and hence by $J$ as well. Applying theorem 8.4 the assertion follows immediately. $\square$

## 8.3 Example: two-neuron rotation matrix interconnected to a single neuron

In order to illustrate the consequences of theorem 8.4 and its remarks, we consider the three-neuron network with the weight matrix

$$W = \begin{pmatrix} A & b \\ d^t & c \end{pmatrix} \qquad \text{where} \qquad A = \begin{pmatrix} 0.7 & -1.1 \\ 1.1 & 0.7 \end{pmatrix}, \tag{8.10}$$

which contains two subnetworks defined by the rotation matrix $A$ and the simple self-coupling term $c$. Both subnetworks are strongly coupled with either submatrix $d^t$ or submatrix $b$ in a feedforward resp. feedback manner (fig. 8.2). For simplicity, the weight matrix $A$ is chosen such that the corresponding two dimensional subnetwork exhibits Neimark-Sacker bifurcations only (fig. 8.3). Remember, that this bifurcation curve resembles the cross sections of the high-dimensional bifurcation diagrams at least for $u_3 = \pm\infty$ resp. $\psi_3 = 0$ (according to theorem 8.2).

In the following qualitative discussion of the dynamical behaviour we consider the self-coupling weight $c$ as another free parameter – additionally to the external inputs $\bar{u}$. According to the analysis in examples 3.1 and 3.5 the single neuron undergoes saddle-node bifurcations at two critical input points $u_3^{\pm}$ as soon as $c > 1$. Within the interval spanned by these bifurcation points, three fixed points exist, each having a different activity $\bar{x}_3$.

### No interaction between subnetworks: $b = 0, d = 0$

In the simplest case, if both subnetworks operate independently of each other, the resulting bifurcation manifolds of the composed network are tensor products of the bifurcation manifolds of the subnets, i.e. the circle corresponding to the Neimark-Sacker bifurcation of the two-neuron network becomes a tube along the $x_3$-axis, and the two bifurcation points $u_3^{\pm}$ become bifurcation planes

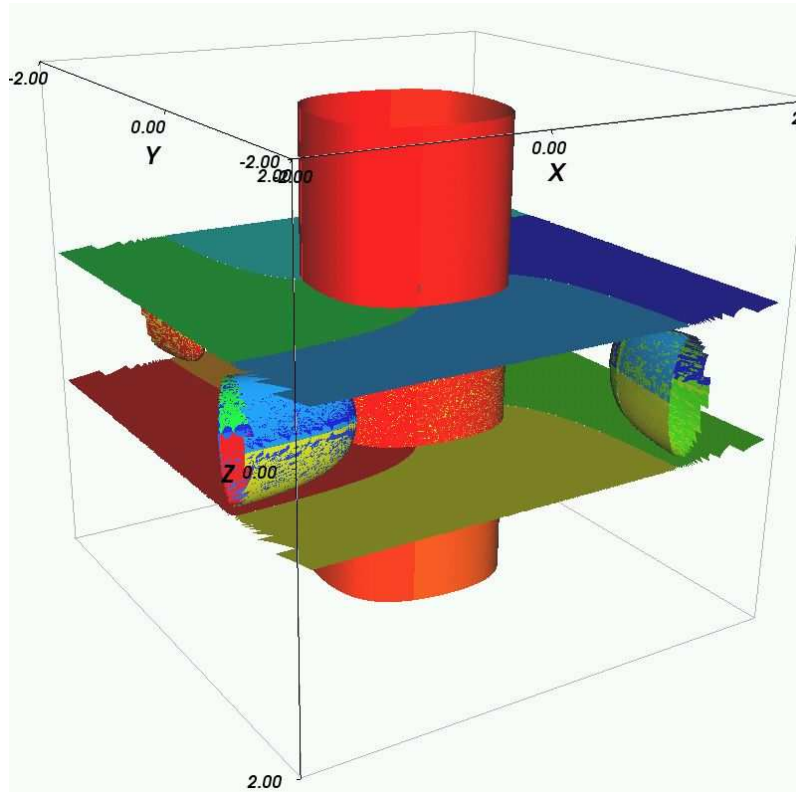PSfrag replacements

$u_1$

$u_2$

$\psi_1$

$\psi_2$

Figure 8.4: Bifurcation manifolds obtained by composing a three-neuron network (8.10) from two simple subnets without interactions between them.

defined by $u_3 = u_3^\pm$ (fig. 8.4). For inputs within the tube, the two-neuron subnetwork oscillates and outside it is globally asymptotically stable. The single-neuron subnetwork possesses three fixed points for all inputs between the bifurcation planes and is g.a.s. otherwise. Naturally the composed network possesses two stable and one unstable invariant manifolds for all input vectors between the planes *and* within the tube. The orbits along these manifolds are oscillatory in $x_1$, $x_2$ but approach a constant value in $x_3$ (see fig. 8.7). The unstable manifold is actually a saddle-manifold, which is attracting within the $x_1$-$x_2$-plane but repelling along the $x_3$-axis.

The bifurcation diagram in fig. 8.4 shows additionally four other manifolds having the shape of truncated ellipsoids. They originate from the *additional* condition introduced by the necessary Neimark-Sacker bifurcation condition:

$$\det(c \otimes A - \mathbf{1}) = c^2 \det A = 0$$

and correspond to neutral saddles. Although we already proved in theorem 8.4 that manifolds originating from this additional condition do not represent true bifurcation manifolds, we want to prove here that they correspond to neutral saddles in case of this example. Simultaneously, this points out an interesting connection to the results of the previous section.

The additional ellipsoid manifolds form straight tubes, if $|u_{1,2}|$ tend to infinity. As we have seen in the previous section, the cross section at $u_i = \pm\infty$ corresponds to the cross section $\psi_i = 0$ in $\psi$-space. Thus the cross section of the tubes in the limit $u_{1,2} \to \pm\infty$ is given by the bifurcation manifolds of the two-dimensional system shown in fig. 8.6. Because the Jacobians of this system can exhibit real eigenvalues only, the Neimark-Sacker curve in fig. 8.6 represents neutral saddles. This argument remains true for the whole bifurcation manifold, because there occur no other fixed point bifurcations on the manifold. Remember from the discussion of the two-dimensional networks, that an intersection of a Neimark-Sacker and a saddle-node manifold could lead to a
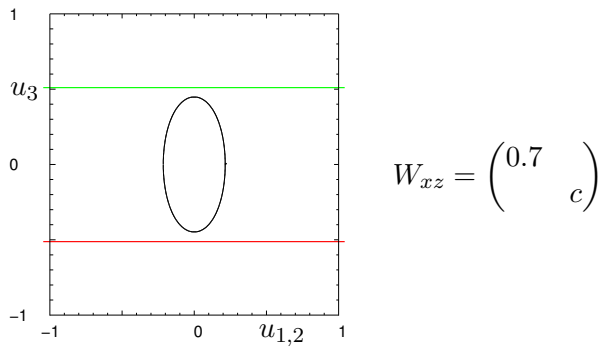
frag replacements

PSfrag replacements

$$W_{xz} = \begin{pmatrix} 0.7 & \\ & c \end{pmatrix}$$

Figure 8.6:  Bifurcation manifolds obtained at cross sections $\psi_1 = 0$ or $\psi_2 = 0$. Because the corresponding weight matrix $W_{xz}$ has only real eigenvalues, the circular Neimark-Sacker curve represents neutral saddles only but no true bifurcation points.
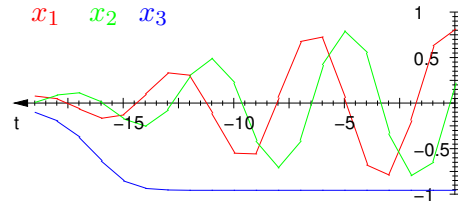
Figure 8.7:  Time series plot of a trajectory starting near the unstable origin of the network with weight matrix $W = \begin{pmatrix} A & \\ & 2 \end{pmatrix}$.

switching of the meaning of the Neimark-Sacker manifold from true bifurcations to neutral saddles and vice versa.

**Feedforward interaction between subnetworks: $b = 0, d \neq 0$**

If feedforward connections from the oscillatory to the single-neuron subnetwork are introduced (see fig. 8.2a), the otherwise stable activity $x_3$ becomes oscillatory as soon as $x_1$, $x_2$ begin to oscillate, i.e. as soon as the input vector enters the oscillatory tube. While its oscillation frequency depends only on the properties of the two-neuron oscillatory subnetwork, the amplitude of the $x_3$-oscillation can be varied smoothly in dependence of the input component $u_3$ (fig. 8.8). If it is large in magnitude, the effect of forwarding the oscillatory components $x_1$, $x_2$ to neuron three is relatively small, such that its oscillation amplitude keeps small. Actually a large external input $u_3$ drives the neuron into the saturation regime of the activation function, such that the relatively small changes of its overall input due to oscillations of $x_1$ and $x_2$ lead to small amplitude oscillations. If the overall input to the neuron becomes smaller in magnitude and departs from the saturation regime, oscillations of $x_1$ and $x_2$ are amplified and the oscillation amplitude increases.

The vertical dashed lines in fig. 8.8 indicate saddle-node bifurcation points $u_3^\pm$ of the single-neuron subnetwork. Obviously the largest amplitudes can be achieved within the Cusp region, i.e. the region containing two stable oscillatory orbits simultaneously. Hence an observed attractor within this region is not globally asymptotically stable, but susceptible to noise. Furthermore we observe the appropriate hysteresis effect indicated by the short dashed lines, while varying the input $u_3$ from negative to positive values and back: If the oscillation amplitude of $x_3$ becomes too large, the trajectory switches to the other stable attractor, which possesses a smaller oscillation amplitude. Actually here a higher-order bifurcation of the oscillatory attractor occurs. From a physicist's point of view we could argue, that the dynamics switches to the attractor with a lower energy level.

From another point of view, we can regard the single-neuron network as a dynamical system with time-dependent, i.e. oscillatory input. This naturally results in an oscillatory trajectory of the neuron's state. If the overall input continuously crosses the opposite Cusp bifurcation boundaries (compare fig. 3.2a) we observe transients from one saturation point to the other. A typical example for this behaviour is shown in fig. 8.9. Here the driving oscillatory network oscillates with low
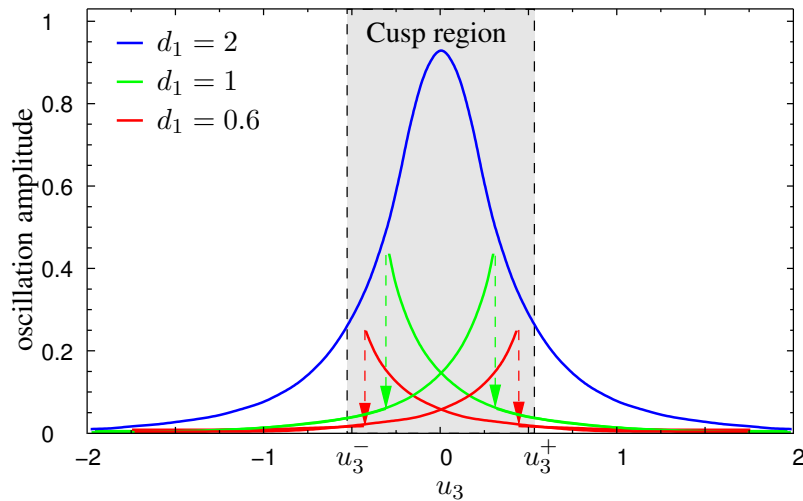
Figure 8.8: Dependence of the oscillation amplitude of $x_3$ in dependence of the input component $u_3$. Obviously the amplitude can be varied smoothly from zero to a certain maximum, which depends on the actually used weight matrix. Shown are three different curves corresponding to different strengths of the feedforward connection.
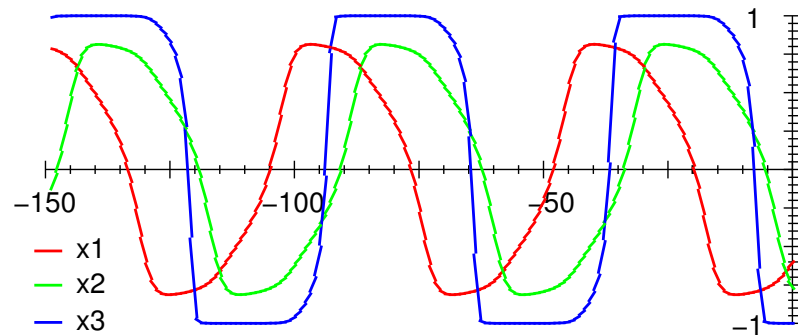


Figure 8.9: The time series shows rectangular oscillations of $x_3$ according to the switching between both saturation values with intermediate transients.

frequency, implying slow changes of the external input and of the bassin boundaries of the fixed points of the third neuron. Consequently we observe oscillations of nearly rectangular shape. The actually observed behaviour depends on several factors simultaneously:

- the frequency of the oscillatory external input relative to the convergence speed of the transients of the single Cusp neuron,

- the amplitude and mean value of the time-dependent input relative to the width of the Cusp region of the single-neuron network and

- the actual shape of the bassin boundaries of the fixed points, which naturally change with the time-dependent external input.

Notice, that together with the frequency control introduced for two-neuron oscillatory networks in section 7.3.6, we attain full control about the main properties of an oscillation: its frequency and its amplitude. Further we do not need to change the weight matrix itself – the control is achievable by two input components only. The weight matrix just defines the range of accessible oscillatory modes.

The introduction of feedforward weights naturally changes the bifurcation manifolds in input space according to the equations (8.6 – 8.7). Because the weights $W_{vc}$ from the single-neuron
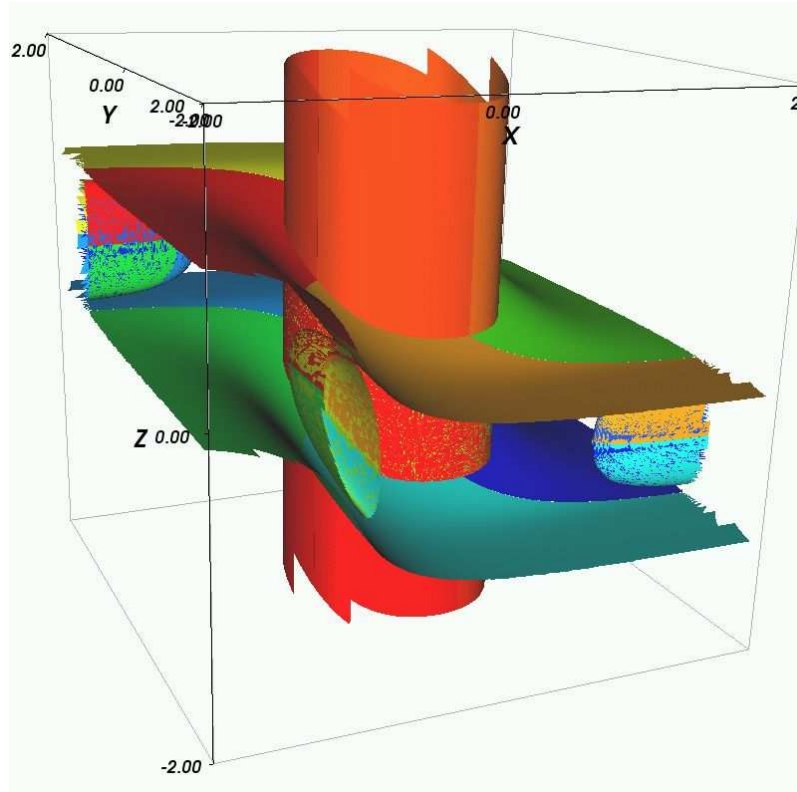
PSfrag replacements

$u_1$

$u_2$

$\psi_1$

$\psi_2$

Figure 8.10: Bifurcation manifolds obtained from network (8.10) with feedforward interactions ($\boldsymbol{d} = [0.6, 0]^t$).

network to the oscillatory two-neuron network vanish, the oscillatory tube still extends along the $u_3$ axis. Nevertheless the saddle-node bifurcation manifolds, which were planes originally, curve now as shown in figure 8.10. Remember, that the bifurcation manifolds in $\boldsymbol{\psi}$-space are independent of the interconnections of the subnetworks and hence do not change at all.

### Feedback interaction between subnetworks: $b \neq 0, d = 0$

If we introduce feedback connections instead of feedforward connections the situation reverses: The saddle-node bifurcation manifolds corresponding to bifurcations of the single-neuron subnetwork remain planes, but the oscillatory tube will be transformed (fig. 8.11a) into a $S$-shaped tube. Actually, its shape becomes similar to the bifurcation diagram of a single-neuron network, shown in figures 3.1 and 8.11b.

An explanation for this behaviour can be given easily, if we consider the corresponding connection structure depicted in fig. 8.2b. The single-neuron is completely decoupled from the rest of the network and hence behaves like a single neuron. Between the bifurcation points $u_3^{\pm}$ three different fixed points exist. Their activity $x_3$ acts as a further external input to the oscillatory subnetwork and thus shifts its Neimark-Sacker bifurcation curve (fig. 8.3) in the input space $u_1$-$u_2$ according to (8.6). Consequently each cross section in input space at an arbitrary value $u_3$ contains up to three (within the interval $(u_3^-, u_3^+)$) identical copies of the Neimark-Sacker bifurcation curve of fig. 8.3. Their displacement from the origin in the $u_1$-$u_2$-space is given by $-\boldsymbol{c}\bar{x}_3(\psi_3)$ which is a function of $\psi_3$ only.

To discuss the observed dynamical behaviour in more detail, consider a cross section in input space near one of the bifurcation points $u_3^{\pm}$ and within the $u_3$-interval corresponding to three fixed
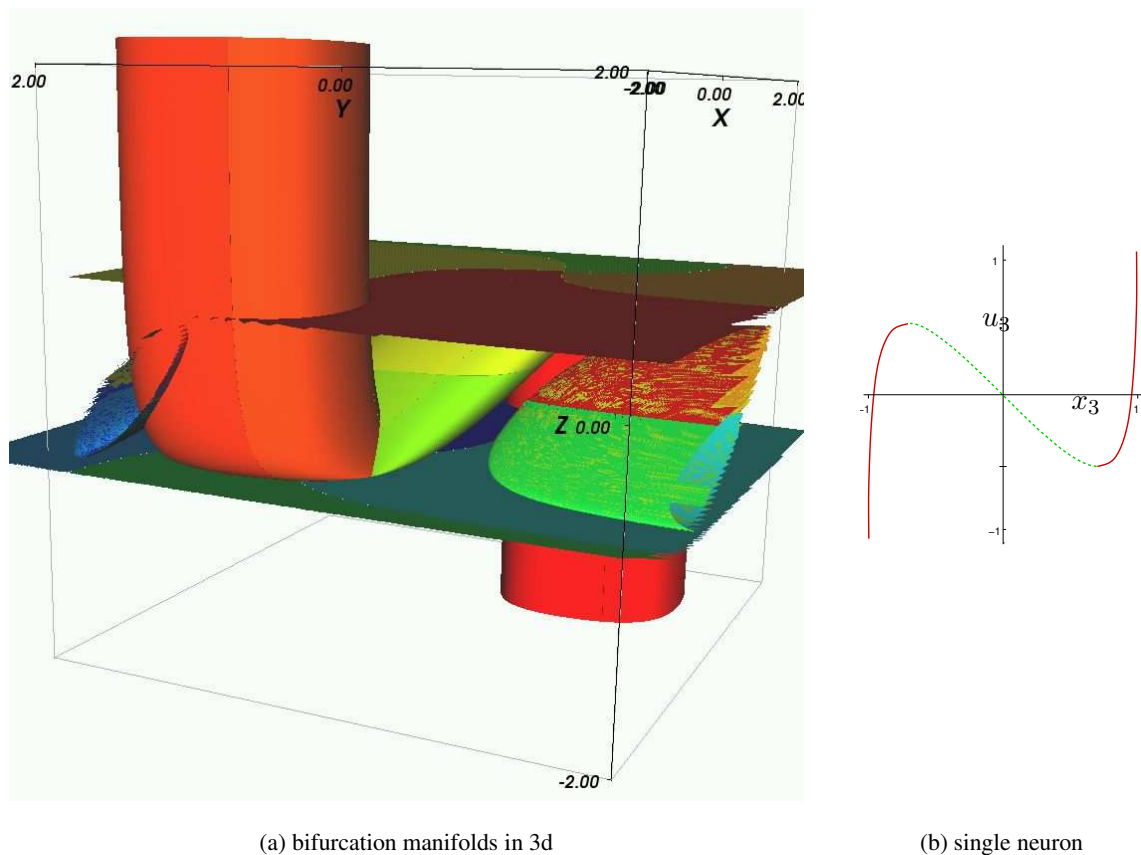
PSfrag replacements

$u_1$

$u_2$

$\psi_1$

$\psi_2$

$u_3$

$x_3$

(a) bifurcation manifolds in 3d                     (b) single neuron

Figure 8.11:   Bifurcation manifolds obtained from network (8.10) with feedback interactions ($\boldsymbol{b}$ = $[1.6, -0.7]^t$). The oscillatory tube curves like the bifurcation diagram of the single-neuron subnetwork (b).

point of the single-neuron dynamics. An appropriate bifurcation diagram is shown in fig. 8.12. Obviously the Neimark-Sacker curves overlap within a neighbourhood of the bifurcation points $u_3^{\pm}$ due to their non-zero radius. Outside the Neimark-Sacker regions the dynamics is convergent. Each Neimark-Sacker region corresponds to a specific fixed point of the single-neuron dynamics, such that crossing a bifurcation boundary leads to an oscillation within the $x_1$-$x_2$-components of the corresponding fixed point. Consequently a previously stable fixed point becomes a saddle – attracting only along the $x_3$-axes. The saddle corresponding to the middle-branch of the single-neuron bifurcation diagram becomes completely unstable. The invariant closed manifold corresponding to this unstable fixed point is attracting within the $x_1$-$x_2$-plane only but repelling along the $x_3$-axes.

Notice, that it is possible to observe two stable invariant manifolds with oscillatory dynamics, if the feedback weights $\boldsymbol{c}$ are small enough, such that the Neimark-Sacker regions I and II overlap. The three parts of the oscillatory tube are coloured red and green in fig. 8.11a. Practically this type of feedback connections are less useful than the feedforward connections discussed previously.

## 8.4  Circular Networks

Circular networks have a weight matrix connecting all neurons in a circular way:

$$x_i(t+1) = \sigma(w_i x_{i-1}(t) + u_i) \qquad i = 1, \ldots, n \quad x_0 \equiv x_n \,.$$
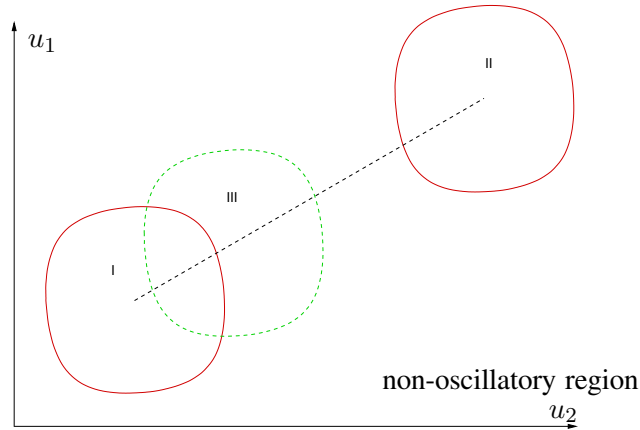
Figure 8.12:   Cross section of the bifurcation diagram in fig. 8.11a at a value $u_3$ in the neighbourhood of one of the bifurcation points $u_3^{\pm}$.

Using the notation $\mathsf{W} = \prod w_i$ and $\Psi(\bar{\boldsymbol{x}}) = \prod \psi_i$ the corresponding Jacobian matrix at a fixed point $\bar{x}$ has eigenvalues

$$\lambda(J) = (\Psi\mathsf{W})^{\frac{1}{n}} .$$

These eigenvalues are distributed equidistantly on a circle in the complex plane with radius $\sqrt[n]{|\Psi\mathsf{W}|}$. Thus all eigenvalues leave the unit circle simultaneously if and only if $|\Psi\mathsf{W}| = 1$, which causes multiple bifurcations and the creation of several periodic orbits. Depending on whether there exist real positive, real negative and complex eigenvalues the corresponding bifurcation types, i.e. saddle-node, period-doubling and Neimark-Sacker, can be observed.

Notice, that all bifurcation types, if they occur at all, possess the same bifurcation condition and the same bifurcation manifolds, given by

$$\Psi = \prod \psi_i = |\mathsf{W}|^{-1} .$$

A detailed discussion of the created periodic orbits was given first by Blum and Wang [1992] for some special circular matrices and was extended later by Nelle and Pasemann [1993] to general circular weight matrices. The key idea to understand the bifurcation behaviour, is to consider the $n$-th iterate of the system map. Because $x_i(t+1)$ is a function of $x_{i-1}(t)$ only, we obtain:

$$x_i(t+n) = \sigma(w_i\sigma(w_{i-1}\sigma(\ldots) + u_{i-1}) + u_i) = \phi_i(x_i(t)) \qquad i = 1, \ldots, n ,$$

i.e. $x_i(t+n)$ is a function of $x_i(t)$ only. Thus the $n$-th iterate of the system map decomposes into $n$ *independent* nonlinear equations. All nonlinear mappings $\phi_i$ are strictly increasing or strictly decreasing sigmoid functions for positive resp. negative $\mathsf{W}$. This follows easily from the fact that their derivatives equal $\Psi\mathsf{W}$ and $\Psi > 0$.

Thus, all $n$ independent one-dimensional subsystems simultaneously undergo a bifurcation if $\Psi(\bar{\boldsymbol{x}})\mathsf{W} = 1$, namely a saddle-node bifurcation if $\mathsf{W} > 0$ and a period-doubling bifurcation if $\mathsf{W} < 0$. Furthermore they cannot exhibit other bifurcation types, for example Neimark-Sacker bifurcations required for quasiperiodic oscillations.

We consider both cases separately in the following. Let be $\mathsf{W} > 0$ first. The corresponding network is called an even loop network. The saddle-node bifurcations of the independent subsystems of the $n$-th iterate, create two new stable fixed points, where the old fixed point looses its stability. Because the composed system is the direct product of the subsystems these three fixed points can be arbitrarily combined to produce a total of $3^n$ different fixed points of the composed system.

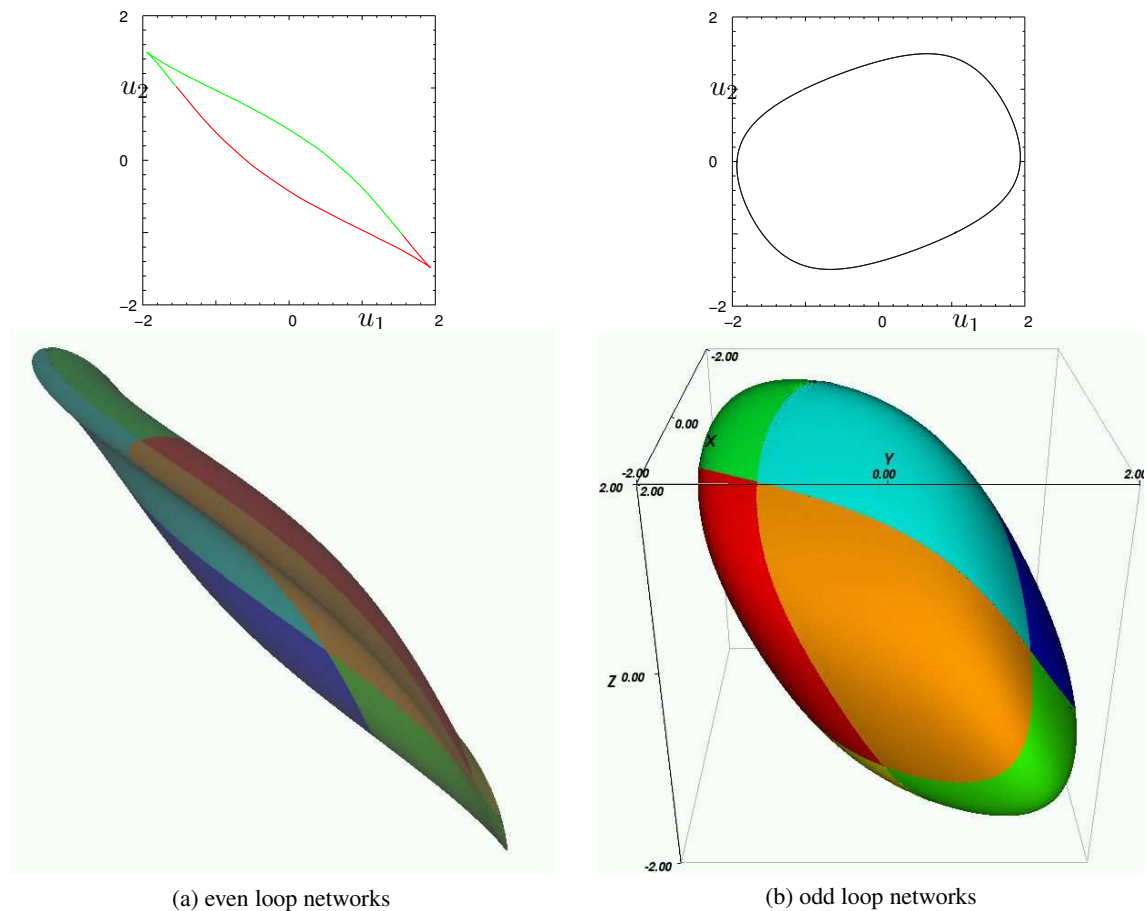(a) even loop networks

(b) odd loop networks

Figure 8.13: Typical bifurcation manifolds of two-neuron (upper row) and three-neuron (lower row) circular networks.

Obviously $2^n$ of these fixed points are stable, one is totally unstable and the other ones are saddles. Because they are fixed points of the $n$-th iterate of the original network, they correspond to periodic orbits of period $p$ with $p \mid n$. In fact there exist two stable fixed points of the original network, corresponding to the single Jacobian eigenvalue $\lambda = 1$. All other stable periodic points can be derived from these fixed points by combination [Blum and Wang, 1992; Nelle and Pasemann, 1993].

The case $\mathsf{W} < 0$, corresponding to so called odd loop networks, can be handled similarly. After the bifurcations has taken place, all one-dimensional subsystems possess a stable 2-cycle and an unstable fixed point. Thus the original network possesses a single unstable fixed point only and $2^n$ stable periodic points, all of even period.

The discussion of circular networks gets finished by an comprehensive characterisation of all observed periodic attractors, done by Pasemann [1995]. It remains to show typical bifurcation manifolds of circular networks. They are visualised for an even and an odd loop circular network in figures 8.13a,b respectively. Notice the sharp boundaries in case of the even loop networks in contrast to the smooth boundaries in case of the odd loop networks. This is a result of the saddle-node bifurcation, implying a sharp Cusp bifurcation manifold of dimension $n - 2$. The enclosed compact region in input space corresponding to complex dynamical behaviour, is deformed according to the relative magnitude of the weights. The bifurcation manifolds in $\psi$- and in state space are defined by $\mathsf{W} = \prod w_i$ only, but do not depend on the magnitudes of the single weights.

# 9 Summary and Discussion

This thesis provides a general approach to analytically compute bifurcation manifolds in the input space of additive recurrent neural networks. We propose to study bifurcation manifolds in the abstract space of activation function derivatives $\boldsymbol{\psi} = \boldsymbol{\sigma}'$ instead of considering the parameter space directly. This approach yields easily solvable bifurcation conditions and the resulting solution manifolds in $\psi$-space are independent of the actually chosen activation function.

Nevertheless, the study of different dynamical regimes in dependence of the external inputs applied to the neurons requires a transformation of the bifurcation manifolds from the abstract $\psi$-space to the parameter space itself. Only this transformation is dependent on the properties of the nonlinear activation functions. The resulting bifurcation diagram in input space is much more complicated than that in $\psi$-space, because each solution branch in $\psi$-space splits typically into $2^n$ branches during the transformation.

We discussed the bifurcation diagram of the two-neuron network, whose weight matrix represents a scaling and rotation in the plane. We have seen, that both the Neimark-Sacker bifurcation and the saddle-node bifurcation on the limit cycle are responsible for emerging oscillatory behaviour. Their different properties facilitate either an amplitude or a frequency control, where the inputs rather than the weights are used as control variables. The control of both properties simultaneously becomes possible, if a cascaded three-neuron network is considered.

Although we developed our approach to compute bifurcation manifolds for discrete-time neural networks only, it easily translates to continuous-time networks as well. If we consider the commonly used continuous-time model

$$\dot{\boldsymbol{x}} = -\boldsymbol{x} + \boldsymbol{\sigma}(W\boldsymbol{x} + \boldsymbol{u})\,,$$

the corresponding conditions for saddle-node bifurcations of fixed points do not differ from those of the discrete-time network. While the period-doubling bifurcation does not occur at all, the Hopf bifurcation condition, which is the continuous-time equivalent to the Neimark-Sacker bifurcation, differs and takes the form

$$\det(2(D(\boldsymbol{x})W - \mathbf{1}) \odot \mathbf{1}) = 0\,.$$

As we have seen, the bifurcation manifolds of high-dimensional RNNs are very complex in general. The exponentially increasing number of their different branches makes it infeasible to compute and visualise them efficiently, although an analytical solution is possible in case of saddle-node and period-doubling bifurcations of fixed points. Actually this increasing number of branches, resulting in an appropriate fine-grained partition of the input space into regions of different dynamical behaviour, accounts for the rich dynamical repertoire of RNNs, which makes them so attractive in engineering applications.

In an empirical study of large networks of McCulloch-Pitts neurons with random connectivity we found an exponentially increasing number of different, mostly oscillatory, dynamical regimes, if the external inputs to the neurons were subject to an increasing level of noise [McGuire et al., 2002]. Below a certain level of noise the observed trajectory was robust against perturbations. In this study we normalised the biases such, that the overall input to each neuron became zero in the average, such that the possibility for a neuron to fire was $\frac{1}{2}$. This corresponds to a centring of the

bifurcation diagram at the origin. From the discussions of the bifurcation manifolds in this thesis, we can conclude, that near the origin there exists the highest density of bifurcation manifolds, which explains the observation of a rich repertoire of dynamical behaviour in the vicinity of this point. Golubitsky and Schaefer [1985] call such regions in parameter space organising centres and Beer [2003] proposes to seed a search in parameter space with appropriately normalised networks.

It is known, that already small networks of two or three neurons show complex dynamical behaviour and our discussion of bifurcation diagrams confirmed this fact. In a large network composed of such small modules, which are interconnected weakly, higher order effects can be observed, such as phase locking or different types of chaos [Wang, 1991b; Pasemann, 1999; Bersini, 2002]. Although it is difficult to draw some general conclusions for the whole network from the dynamics of its components, we proved that this becomes possible for cascaded networks. If the recurrent interconnection of the submodules among each other remains weak, it should be possible to prove similar results for those networks as well, e.g. employing perturbation techniques.

The independence of the bifurcation manifolds in $\psi$-space casts a new light on the absolute stability conditions for recurrent neural networks of chapter 4. As we have seen, the existence of bifurcation manifolds, which correspond to local fixed point bifurcations, depends on the clipping hypercube of admissible solutions in $\psi$-space. This hypercube is defined by the bounds determined by the incremental sector conditions which are fulfilled by the activation functions — the particular shape of these activation functions is not relevant. Consequently, the occurrence of fixed point bifurcations indeed depends on the incremental sector condition and the weight matrix only, such that it is not possible to find stricter absolute stability conditions in many cases. Indeed, in theorem 7.2, we used this property to prove absolute stability of the two-neuron network, which we investigated in detail in chapter 7. Only if the absolute stability condition of theorem 4.3 and the no-bifurcation-condition differ, there is room for further improvements of the absolute stability bounds, because absolute stability implies that there cannot occur any bifurcation.

It is interesting, that we predominantly observe supercritical bifurcations in simulations of RNNs, i.e. most bifurcation are soft, besides those of saddle-node type. Presumably this is the reason, why weight adaption of typical learning methods navigates relative robustly through the huge parameter space. As we have seen, the parameter space is heavily cluttered by bifurcation manifolds. Thus, online parameter changes will often cross bifurcation boundaries, which would lead to catastrophic behaviour of the trajectory if the associated bifurcations would be hard.

As a recent study suggests, learning never really leaves the stability domain if certain learning tasks are considered [Schiller and Steil, 2003; Schiller, 2003]. Rather, the network approaches the edge of the stability domain, such that small perturbations cause a bifurcation with an appropriate change of dynamical behaviour. A deeper study of this phenomenon is desirable in order to understand the underlying principles of learning algorithms. Such a study can heavily benefit from our approach to compute bifurcation manifolds, such that trajectories in parameters space – as they occur during learning – can be projected to the underlying bifurcation diagram, which in turn allows conclusions about the change of the dynamical behaviour.

In this work, we concentrated on the study of bifurcation diagrams in parameter space. It should be emphasised, that there might exist simultaneously several different dynamical regimes in *state space* for one and the same parameter vector. Indeed, during the discussion of the phase portraits corresponding to various input regions, we have seen the simultaneous existence of several attractors. Which one of these attractors is finally observed, depends crucially on the initial state of the network. Because this state can be changed by an external intervention only, we concentrated on the analysis of bifurcation diagrams, which allow to control the dynamical behaviour with external inputs, which do not influence the state of the network directly. Nevertheless, the computation of the bassin boundaries of existing attractors is as much interesting as the computation of bifurcation

manifolds in parameter space. Currently, there do not exist methods to compute these boundaries analytically, but some authors compute them by costly simulations [Pasemann, 2002].

Another important subject for future research is the computation of bifurcation manifolds of higher period orbits and other global bifurcations. Their knowledge would improve the bifurcation diagrams discussed in this work, which considered local fixed point bifurcations only. Sometimes we can conclude from the occurring local bifurcations to the existence of such global bifurcations, but we are not able to compute their associated manifolds analytically yet. From simulations we conjecture, that the bifurcation manifolds corresponding to higher iterates of the system's mapping are typically located within regions in parameter space, which are enclosed by fixed point bifurcation manifolds already. Otherwise we would have observed more complex phase portraits in our simulations.

We should emphasise, that the whole bifurcation picture of RNNs still remains to be unknown – even in the case of a two-neuron network – if all weight and input parameters are taken into account. Despite our efforts to outline parts of this picture, there is much room for further research. We want to finish with a striking quotation of Beer [1995]:

*"The range of phase portraits and bifurcations that can
occur in higher dimensions is bewildering."*

# A Preliminaries from Linear Algebra and Analysis

## A.1 Jordan Normal Form

Each quadratic matrix $A \in \mathbb{R}^{n \times n}$ is similar to a matrix $J = P^{-1}AP$, which is called *Jordan normal form* of $A$ and which has the form

$$
J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix} . \tag{A.1}
$$

Thus, it splits into quadratic Jordan blocks $J_i = J_{r_i}(\lambda_i)$ along the diagonal and all other entries of $J$ are zero. The *Jordan blocks* to the eigenvalue $\lambda$ have the form

$$
J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix} = \lambda \mathbf{1} + N \in \mathbb{C}^{r \times r} , \tag{A.2}
$$

where the eigenvalues $\lambda$ might be complex as well. Because $A$ is real, complex eigenvalues occur as complex conjugate pairs always. The columns of the matrix $P$ are called *generalised eigenvectors of $A$*, i.e. they are elements of $\ker(A - \lambda\mathbf{1})^i$. If $A$ is diagonalisable all Jordan blocks have dimension $r = 1$ and $P$ forms a basis of eigenvectors. The matrix $A$ is diagonalisable if and only if all eigenvalues are simple, i.e. their algebraic and geometric multiplicities are equal.

$N^j$ is a nilpotent matrix. All its entries equal to zero, except those with a distance $j$ above the diagonal. More precisely it holds $(N^j)_{k,l} = \delta_{k+j,l}$. For example, for an Jordan block of dimension $r = 3$ we obtain:

$$
N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad N^2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad N^3 = 0 .
$$

## A.2 Spectral Properties of Matrices

**Definition A.1** Given a matrix $A \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_i$, $i = 1, \ldots, m$ $(m \leq n)$ we define the *spectral radius* of $A$ as

$$
\rho(A) = \max_i |\lambda_i|
$$

**Lemma A.2 (Stuart and Humphries [1996])** If all eigenvalues of a matrix $A \in \mathbb{R}^{n \times n}$ are simple there exists a norm $\|\cdot\|$ on $\mathbb{R}^n$ such that

$$
\|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \rho(A) .
$$

Otherwise, for an arbitrary small $\varepsilon > 0$ there exists a norm $\|\cdot\|$ on $\mathbb{R}^n$, such that it holds

$$\rho(A) \le \|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| \le \rho(A) + \varepsilon\,.$$

**Theorem A.3 (Lancaster and Tismenetsky [1985])**
Let $A, M \in \mathbb{R}^{n \times n}$ with $m_{ij} \ge |a_{ij}| \ge 0$. Then $\rho(M) \ge \rho(A)$.

## A.3  Bialternate Product

The Neimark-Sacker bifurcation condition requires the computation of the bialternate product matrix $J \odot J$. Given a specific block structure of the matrix $J$, the bialternate product preserves this structure to some degree, as we prove in the following theorem. We utilise this result in chapter 8 to derive several decomposition theorems.

**Theorem A.4** Given a matrix $M \in \mathbb{R}^{(n+m) \times (n+m)}$ with the following block structure

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},\tag{A.3}$$

where $A \in \mathbb{R}^{n \times n}$, $D \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{m \times n}$. Then its bialternate product $M \odot M$ is similar to the matrix

$$\begin{pmatrix} A \odot A & X_{\hat{n},mn} & B \odot B \\ X_{mn,\hat{n}} & X_{mn,mn} & X_{mn,\hat{m}} \\ C \odot C & X_{\hat{m},mn} & D \odot D \end{pmatrix},\tag{A.4}$$

where $X_{a,b}$ denote matrices of dimension $a \times b$ and $\hat{n} = \frac{1}{2}n(n-1)$, $\hat{m} = \frac{1}{2}m(m-1)$.

**Proof** The definition of the bialternate product (3.18) yields:

$$(M \odot M)_{(k,l),(i,j)} = \begin{vmatrix} m_{ki} & m_{li} \\ m_{kj} & m_{lj} \end{vmatrix} \quad k,l,i,j = 1\dots n+m \text{ and } l<k, j<i\tag{A.5}$$

In order to infer the block structure of the bialternate product matrix, we reorder its rows – specified by indices $k, l$ – and its columns – specified by indices $i, j$ – into three partitions according to the original block structure of the matrix $M$:

- $1 \le k \le n$ and $1 \le l < k \le n$          - $1 \le i \le n$ and $1 \le j < i \le n$

- $n < k \le n+m$ and $1 \le l \le n$  and     - $n < i \le n+m$ and $1 \le j \le n$

- $n < k \le n+m$ and $n < l < k$          - $n < i \le n+m$ and $n < j < i$

This yields the following matrix for the bialternate product:

| | $1 \le i \le n$ $1 \le j < i$ | $n < i \le n+m$ $1 \le j \le n$ | $n < i \le n+m$ $n < j < i$ |
|---|---|---|---|
| $1 \le k \le n$ $1 \le l < k$ | $\begin{vmatrix} a_{k,i} & a_{l,i} \\ a_{k,j} & a_{l,j} \end{vmatrix}$ | $\begin{vmatrix} b_{k,i-n} & b_{l,i-n} \\ a_{k,j} & a_{l,j} \end{vmatrix}$ | $\begin{vmatrix} b_{k,i-n} & b_{l,i-n} \\ b_{k,j-n} & b_{l,j-n} \end{vmatrix}$ |
| $n < k \le n+m$ $1 \le l \le n$ | $\begin{vmatrix} c_{k-n,i} & a_{l,i} \\ c_{k-n,j} & a_{l,j} \end{vmatrix}$ | $\begin{vmatrix} d_{k-n,i-n} & b_{l,i-n} \\ c_{k-n,j} & a_{l,j} \end{vmatrix}$ | $\begin{vmatrix} d_{k-n,i-n} & b_{l,i-n} \\ d_{k-n,j-n} & b_{l,j-n} \end{vmatrix}$ |
| $n < k \le n+m$ $n < l < k$ | $\begin{vmatrix} c_{k-n,i} & c_{l-n,i} \\ c_{k-n,j} & c_{l-n,j} \end{vmatrix}$ | $\begin{vmatrix} d_{k-n,i-n} & d_{l-n,i-n} \\ c_{k-n,j} & c_{l-n,j} \end{vmatrix}$ | $\begin{vmatrix} d_{k-n,i-n} & d_{l-n,i-n} \\ d_{k-n,j-n} & d_{l-n,j-n} \end{vmatrix}$ |

$$\tag{A.6}$$

Comparing the corners of this table with definition (A.5) obviously yields the structure of (A.4).
□

**Remark:** The needed similarity transform $P$ is a permutation matrix, i.e. swaps rows resp. columns of the original bialternate matrix, such that the form of (A.6) is achieved. The original definition of the bialternate matrix with lexicographic ordering of the double indices intermingles rows and columns between the second and third blocks in (A.6). Notice, that the actually employed ordering is irrelevant if the determinant of the bialternate product is considered. Each permutation introduces a factor of $-1$ to the determinant, which can be neglected in our cases, because the determinant shall equal zero.

If we set $C = 0$ in the previous equations, we immediately obtain the following corollary.

**Corollary A.5** Given the block matrix

$$M = \begin{pmatrix} A & B \\ & D \end{pmatrix} , \tag{A.7}$$

i.e. $C = 0$ in (A.3), its bialternate product is similar to the matrix:

$$\begin{pmatrix} A \odot A & X_{\hat{n},mn} & B \odot B \\ & D \otimes A & X_{mn,\hat{m}} \\ & & D \odot D \end{pmatrix} ,$$

where $D \otimes A$ denotes the tensor product of the matrices $D$ and $A$[1].

This result is especially useful for the bifurcation analysis of fixed points of *cascaded* RNNs. These types of RNNs introduced by Hirsch [1989] have a weight matrix $W$ which takes the form of equation (A.7) – eventually after permuting the order of neurons. Thus a condition involving the determinant of the bialternate product of its Jacobian splits into a product of determinants – including those of the subnets (compare theorem 8.4).

## A.4 Solutions of a quadratic equation

The solution formula for the roots of the quadratic equation $ax^2 + bx + c = 0$ is well known:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} . \tag{A.8}$$

However this formula is numerically unstable: If $a$ or $c$ are small, then one of the roots will involve subtraction of $b$ from a very nearly equal quantity (the square root of the discriminant), which leads to a large numerical error for this root. A better way to compute the roots is [Press et al., 1992]:

$$x'_1 = \frac{q}{a} \qquad x'_2 = \frac{c}{q} \qquad \text{where } q = -\frac{1}{2}\left(b + \operatorname{sgn}(b)\sqrt{b^2 - 4ac}\right) , \tag{A.9}$$

which ensures that $q \to -b$ if $ac \to 0$.

Further we are interested in the continuity of each of the two possible *real* roots in dependence of the parameters $a$, $b$ and $c$. Because (A.9) introduces an artificial discontinuity through $q$, we consider (A.8) first. Clearly the square root of the discriminant introduces a discontinuity, if it

---

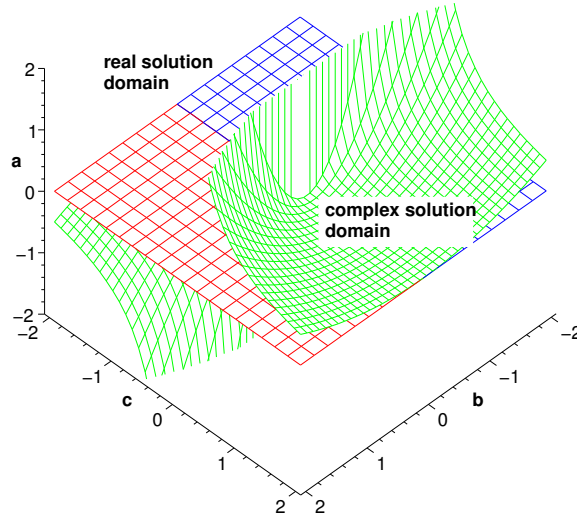[1] For an introduction to tensor products, see Lancaster and Tismenetsky [1985]

Figure A.1:  Discontinuity manifolds for the roots $x_1$, $x_2$ of the quadratic equation $ax^2 + bx + c = 0$ in $a$-$b$-$c$ parameter space.

looses its real solution (within domain $b^2 - 4ac < 0$). Another discontinuity is introduced by the zero of the denominator $2a$. The development of the solutions into a Taylor series

$$x_{1,2} = \frac{-b \pm |b| \mp 2a\frac{c}{|b|} + \mathcal{O}(a^2)}{2a} \tag{A.10}$$

shows that only one solution is discontinuous, namely $x_2$ if $b \geq 0$ and $x_1$ if $b \leq 0$. The other solution goes through $-\frac{c}{b}$, which is the solution of the remaining linear equation ($a = 0$). We excluded the special point $a = b = c = 0$, where any value $x$ is a solution. Altogether we get the discontinuity manifolds in $a$-$b$-$c$-space shown in fig. A.1, which separate this space into four regions. The green manifold separates the real solution domain from the complex solution domain, while the blue and red manifolds mark the discontinuous changes of $x_1$ and $x_2$ respectively. Along the line of their contact, i.e. for $a = b = 0$ no solutions exist at all, excepting the point $c = 0$.

In order to transfer these discontinuity results to the numerically stable solutions $x'_{1,2}$ of (A.9), we have to correlate both solutions with each other. As can be seen easily through comparison of both formulas it holds $x_1 = x'_1$ if $\mathrm{sgn}(b) = -1$. Thus:

$$x_1 = \begin{cases} x'_1 & \text{if } b < 0 \\ x'_2 & \text{else} \end{cases} \qquad x_2 = \begin{cases} x'_2 & \text{if } b < 0 \\ x'_1 & \text{else} \end{cases}, \tag{A.11}$$

which removes the artificial discontinuity introduced by $\mathrm{sgn}(b)$ in (A.9).

# B Matrix Stability

Many results of global asymptotic stability (GAS) of nonlinear systems are given in terms of GAS of a reference linear system or more precisely a whole set of corresponding linear systems. The nonlinear system can then be viewed as a perturbation of the reference linear system and its stability properties are assured by robust stability of the reference system or stability of the whole set of linear systems (compare section 4).

Thus it is important to characterise robust GAS of linear systems, i.e. stability of matrices and matrix polytopes. In the following we summarise some important results, which are particularly relevant for dynamical systems analysis. A deep introduction into the topic of matrix stability is given by Kaszkurewicz and Bhaya [2000].

We start with the definition of a number of matrix properties of real square matrices $A \in \mathbb{R}^{n \times n}$ and we cite some theorems characterising these properties.

**Definition B.1 (Definiteness)** The matrix $A$ is said to be *positive (semi)definite* if its dedicated quadratic form $v \mapsto v^t A v$ is strictly positive (non-negative) for all vectors $v \neq 0$. We write $A > 0$ $(A \geq 0)$ for short. Conversely $A$ is said to be *negative (semi)definite* if $-A$ is positive (semi)definite.

The definiteness of a matrix $A$ is determined by its symmetric part $A_{\text{sym}} = \frac{1}{2}(A^t + A)$ only. Actually it holds $v^t A_{\text{sym}} v = v^t A v$ for all $v$. For a symmetric real matrix $A$ we state the following equivalent characterisations of positive (semi)definiteness:

**Lemma B.2** Let $A \in R[n \times n]$ be symmetric. Then the following holds:

  (i) $A$ is positive (semi)definite if and only if all its (real) eigenvalues $\lambda_i(A)$ are positive (non-negative). [Ortega, 1987, p. 113]

  (ii) $A$ is positive definite if and only if all leading principal minors of $A$ are positive. [Zhang, 1999, p. 160]

  (iii) $A$ is positive semidefinite if and only if all principal minors (not only leading) are nonnegative. [Zhang, 1999, p. 160]

**Definition B.3** A real square matrix $A \in \mathbb{R}^{n \times n}$ is called *Schur stable* if there exists a positive definite matrix $P$ such that $A^t P A - P$ is negative definite. Analogously $A$ is said to be *Hurwitz stable* if $A^t P + P A$ is negative definite for some positive definite $P$.

The matrix is said to be Schur resp. Hurwitz *diagonally stable* if there exists a *diagonal* positive definite matrix $P$ with the above properties.

While Schur stability is equivalent to the proposition that the eigenvalue spectrum of $A$ lies completely in the interior of the unit disk of the complex plane (Stein's theorem) and thus ensures global asymptotic stability of the associated linear discrete-time dynamical system (2.2), Hurwitz stability is equivalent to the proposition that the eigenvalue spectrum lies completely in the open left half plane (Lyapunov equation) and ensures global asymptotic stability for the linear continuous-time dynamical system (2.1).

Note that given any positive definite matrix $P$, which satisfies Stein's resp. Lyapunov's equation for a matrix $A$, its symmetric part $P_{\mathrm{sym}}$ satisfies them too. Thus we will consider only symmetric matrices $P$ guaranteeing Schur resp. Hurwitz stability. Note further that a symmetric positive matrix $P$ defines a norm $\|\cdot\|_P = (\boldsymbol{x}^t P \boldsymbol{x})^{\frac{1}{2}}$ which can be used as a quadratic Lyapunov function as well. Diagonal stability thus defines a diagonal Lyapunov function and is clearly a stronger condition than stability.

Diagonal stability of matrices can be tested using linear matrix inequalities (LMI), which can be efficiently solved using convex optimisation tools [Boyd et al., 1993; Boyd, 1994; Balakrishnan, 1995]. This technique bases on the following equivalence for symmetric matrices $Q$, $R$ and an arbitrary matrix $S$. The LMI

$$\begin{bmatrix} Q & S \\ S^t & R \end{bmatrix} > 0$$

is equivalent to the set of nonlinear inequalities

$$R > 0 \quad \text{and} \quad Q - SR^{-1}S^t > 0 \,,$$

which resembles Stein's equation if one sets $Q = P$, $R = P^{-1}$ and $S = A^t$.

**Definition B.4** A matrix $A \in \mathbb{C}^{n \times n}$ is said to be normal if $A^* A = A A^*$.

This is equivalent to the condition, that a unitary matrix $U$ exists such that $U A U^*$ is diagonal.


## B.1 Matrix Polytopes and Interval Matrices

Now we introduce the concepts of matrix polytopes and interval matrices, which are sets of matrices frequently occurring during the analysis of nonlinear or uncertain dynamical systems. In the following, sets of matrices will be denoted by calligraphic letters $\mathcal{A}, \mathcal{K}, \mathcal{P}$ etc. while single matrices are further denoted by normal letters.

**Definition B.5** Given a set of vertex matrices $\mathcal{V} = \{V_1, \ldots, V_m\}, V_k \in \mathbb{R}^{n \times n}$ the convex *matrix polytope* $\mathcal{P}(\mathcal{V})$ is defined as

$$\mathcal{P}(\mathcal{V}) := \mathrm{conv}(\mathcal{V}) := \left\{ \sum \alpha_k V_k \,|\, \alpha_k \geq 0, \ \sum \alpha_k = 1 \right\} \,.$$

Using component-wise matrix inequalities, i.e. $A \leq B \ \Leftrightarrow \ a_{ij} \leq b_{ij}$, we define an *interval matrix* with boundaries $\underline{A}$ and $\overline{A}$ as the set of matrices

$$\mathcal{A} = [\underline{A}, \overline{A}] := \left\{ A \in \mathbb{R}^{n \times n} \,|\, \underline{A} \leq A \leq \overline{A} \right\} \,.$$

Notice, that an interval matrix is a hyperrectangle of maximal dimension $n^2$ (if all bounds satisfy $\underline{a}_{ij} < \overline{a}_{ij}$) and thus can be viewed as a special matrix polytope.

Hurwitz and Schur stability of *symmetric* interval matrices, which are given if both $\underline{A}$ and $\overline{A}$ are symmetric, is equivalent to the stability of a finite set of vertex matrices [Rohn, 1994b][1]. For arbitrary interval matrices Xiao and Unbehauen [2000] recently proved equivalence of Schur resp. Hurwitz stability of the whole interval matrix and its exposed 2-d hypersurfaces. Consequently, it

---

[1] In the case of Schur stability one has to restrict to the symmetric matrices within the interval.

does not suffice to check a finite number of vertex matrices anymore. Rather, the infinite number of matrices on these hypersurfaces have to be checked, which is not feasible in general.

Of particular interest in the stability analysis of recurrent neural networks (2.5) are matrix polytopes of the type $\mathcal{P}(W) = \{DW \,|\, D = [-\mathbf{1}, \mathbf{1}]\}$. These sets particularly include all possible Jacobian matrices of (2.5), if sigmoid activation functions are considered, which have a maximal slope equal to one.

**Definition B.6** Let $\mathcal{K} := [-\mathbf{1}, \mathbf{1}] = \mathrm{conv}(\{K \,|\, |K| = \mathbf{1}\})$ be the set of diagonal matrices with entries less than or equal 1 in absolute value. Given a matrix $A \in \mathbb{R}^{n \times n}$ we define

$$\mathcal{L}(A) := \{KA \,|\, K \in \mathcal{K}\}$$
$$\mathcal{R}(A) := \{AK \,|\, K \in \mathcal{K}\}.$$

Although both sets are included in $[-|A|, |A|]$ they significantly differ from this interval matrix, because rows resp. columns of $A$ are scaled *uniformly* in $\mathcal{L}(A)$ resp. $\mathcal{R}(A)$. In contrast, all entries of an interval matrix can vary independently. Hence it is possible to derive stricter stability theorems for the sets $\mathcal{L}(A)$ and $\mathcal{R}(A)$, which we cite from Bhaya and Kaszkurewicz [1993] in the following. To this end we need some further definitions. Because we will consider Schur stability only, we write stability for short.

**Definition B.7** We denote the sets of diagonally stable, **D**-stable, and stable matrices by

$$\mathcal{D} := \{A \,|\, A \text{ diagonally Schur stable}\},$$
$$\mathbf{D} := \{A \,|\, \mathcal{R}(A) \subset \mathcal{S}\} \text{ and}$$
$$\mathcal{S} := \{A \,|\, A \text{ Schur stable}\} \text{ respectively.}$$

A set $\mathcal{A}$ of matrices is called *simultaneously* diagonally stable, **D**-stable, or stable if there exists a *single* matrix $P$ such that all matrices $A \in \mathcal{A}$, are diagonally stable, **D**-stable, or stable employing this single matrix $P$ to satisfy Stein's condition. In this case we denote the appropriate matrix sets as $\mathcal{D}(P)$, $\mathbf{D}(P)$ and $\mathcal{S}(P)$ respectively in order to indicate, that we refer to a specific matrix $P$, for which Stein's equation is satisfied.

We have the following inclusions, where the first inclusion is proved in theorem B.10 below:

$$\mathcal{D} \subsetneq \mathbf{D} \subsetneq \mathcal{S} \qquad \text{and} \qquad \mathcal{D}(P) \subsetneq \mathbf{D}(P) \subsetneq \mathcal{S}(P).$$

Now we can state the following theorems:

**Theorem B.8 (Bhaya and Kaszkurewicz [1993])**
Let be given an arbitrary matrix $A$. Then $\mathcal{R}(A)$ is simultaneously stable if and only if all vertices of $\mathcal{R}(A)$ are simultaneously stable, i.e. $\mathcal{R}(A) \subset \mathcal{S}(P) \quad \Leftrightarrow \quad AK \in \mathcal{S}(P)$ for all $K$ with $|K| = \mathbf{1}$.

Simultaneous stability of the whole matrix polytope allows to define the global quadratic Lyapunov function $V(\boldsymbol{x}) = \boldsymbol{x}^t P \boldsymbol{x}$, which proves global asymptotic stability of the corresponding linear, but time-varying dynamical system. In particular, it can be used to prove absolute stability of (4.9), which implies absolute stability of the underlying nonlinear recurrent network. Hence, it is important to find classes of matrices $A$ such that the matrix polytopes $\mathcal{R}(A)$ or $\mathcal{L}(A)$ are simultaneously stable. The matrix $A$ is called simultaneously **D**-stable in this case.

**Theorem B.9 (Wang [1991a])**
Let $\mathcal{P}(\mathcal{V}) := \text{conv}(\{V_1, \ldots, V_m\})$ be a matrix polytope of normal matrices $V_i \in \mathbb{C}^{n \times n}$. Then the set of eigenvalues of all matrices $A \in \mathcal{P}(\mathcal{V})$ is a subset of the convex hull of all eigenvalues of the vertex matrices only:

$$\Lambda(\mathcal{P}) := \{\lambda_j(A) \,|\, A \in \mathcal{P}\} \subset \text{conv}\{\lambda_j(V_i) \,|\, V_i \in \mathcal{V}\}\,.$$

Thus given normal vertices, the stability of the whole matrix polytope is equivalent to the stability of its vertices.

**Theorem B.10 (Bhaya and Kaszkurewicz [1993])**
Given an arbitrary matrix $A$, the following equivalences hold:

$$\begin{aligned} A \in \mathcal{D}(P) \quad &\Leftrightarrow \quad \mathcal{L}(A) \subset \mathcal{D}(P) \quad \Leftrightarrow \quad \mathcal{R}(A) \subset \mathcal{D}(P) \qquad \text{and} \\ \mathcal{L}(A) \subset \mathcal{S} \quad &\Leftrightarrow \quad \mathcal{R}(A) \subset \mathcal{S}\,. \end{aligned}$$

According to this theorem, the two different matrix polytopes $\mathcal{L}(A)$ and $\mathcal{R}(A)$ are equivalent with respect to their stability properties. This agrees with the fact, that both types of discrete-time RNNs (2.11) and (2.12) exhibit equivalent dynamical behaviour.

Because diagonal stability of $A$ implies simultaneous diagonal stability of the whole polytope and thus absolute stability of (4.9), it is interesting to identify classes of matrices that are diagonally stable. Within such a class of diagonally stable matrices the terms diagonally stable, diagonally **D**-stable and stable become equivalent, because the following implications hold in general:

$$A \in \mathcal{D}(P) \quad \Rightarrow \quad A \in \mathbf{D}(P) \quad \Rightarrow \quad A \in \mathcal{S}(P)\,.$$

Obviously, a necessary prerequisite a diagonally stable matrix has to meet, is its stability. Bhaya and Kaszkurewicz [1993] identify the following classes of diagonally stable matrices:

- Stable, diagonally symmetrisable matrices,
  i.e. stable matrices $A$ for which a diagonal matrix $D$ exists such that $D^{-1}AD$ is symmetric.

- Stable positive matrices.

- **D**-stable matrices $A \in \mathbb{R}^{2 \times 2}$

- Stable triangular matrices

- Matrices $A$, such that $|A|$ is stable.

Employing the theory of M-matrices[2], Araki [1975] finds an additional class of diagonally stable matrices:

**Theorem B.11** Let $A$ be a nonnegative matrix, i.e. $a_{ij} \geq 0$. Then $A$ is diagonally stable if and only if $\mathbf{1} - A$ is an $M$-matrix.

Finally we contribute a theorem, which extends this class to stable normal matrices as well:

**Theorem B.12** For normal matrices stability implies diagonal stability as well.

---

[2] A comprehensive overview of non-negative and M-matrices is given by Berman and Plemmons [1979]

**Proof**    Let $A$ be normal. Then there exists an unitary matrix $U$ such that $D = U^*AU$ is the diagonal matrix of eigenvalues of $A$. Note that $U$ and $D$ are complex matrices in general. Because $A$ is stable it follows that $D^*D = \overline{D}D = \mathrm{diag}(|\lambda_i(A)|^2)$ has entries smaller than one. Thus we can conclude:

$$\overline{D}D - \mathbf{1} < 0$$
$$\Leftrightarrow \quad U\overline{D}U^* \, UDU^* - UU^* < 0 \quad \text{because } U^{-1} = U^*$$
$$\Leftrightarrow \quad A^*\mathbf{1}A - \mathbf{1} < 0$$

$\square$

# Bibliography

Mituhiko Araki. Application of M-matrices to the stability problems of composite dynamical systems. *Journal of Mathematical Analysis and Applications*, 52:309–321, 1975.

Amir F. Atiya and Alexander G. Parlos. New results on recurrent network training: Unifying the algorithms and accelerating convergence. *IEEE Trans. Neural Networks*, 11(3):697–709, May 2000.

Andrew D. Back and Tianping Chen. Universal approximation of multiple nonlinear operators by neural networks. *Neural Computation*, 14:2561–2566, 2002.

V. Balakrishnan. Linear matrix inequalities in robustness analysis with multipliers. *System & Control Letters*, 25:265–272, 1995.

Nikita E. Barabanov and Danil V. Prokhorov. Stability analysis of discrete-time recurrent neural networks. *IEEE Trans. Neural Networks*, 13(2):292–303, 2002.

P. H. Bauer, K. Premaratne, and J. Durán. A necessary and sufficient condition for robust asymptotic stability of time-variant discrete systems. *IEEE Trans. Automatic Control*, 38(9):1427–1430, 1993.

Randall D. Beer. On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behaviour*, 3(4):469–509, 1995.

Randall D. Beer. Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3): 91–99, 2000.

Randall D. Beer. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 2003. in press.

Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Computer science and applied mathematics. Academic Press, New York, 1979.

Pierre Bersini, Hugues; Sener. The connections between the frustrated chaos and the intermittency chaos in small hopfield networks. *Neural Networks*, 15(10):1197–1204, 2002.

Amit Bhaya and Eugenius Kaszkurewicz. On discrete-time diagonal and D-stability. *Linear Algebra and its Applications*, 187:87–104, 1993.

Amit Bhaya, Eugenius Kaszkurewicz, and Victor S. Kozyakin. Existence and stability of a unique equilibrium in continuous-valued discrete-time asynchronous hopfield neural networks. *IEEE Trans. Neural Networks*, 7(3), 1996.

E. K. Blum and Xin Wang. Stability of fixed points and periodic orbits and bifurcations in analog neural networks. *Neural Networks*, 5:577 – 587, 1992.

R. M. Borisyuk and A. Kirillov. Bifurcation analysis of a neural network model. *Biological Cybernetics*, 66:319–325, 1992.

S. Boyd, V. Balakrishnan, E. Feron, and L. El Ghaoui. Control system analysis and synthesis via linear matrix inequalities. In *Proc. ACC*, pages 2147–2154, 1993.

Stephen P. Boyd. *Linear matrix inequalities in system and control theory*. Number 15 in SIAM studies in applied mathematics. SIAM, 1994.

Shannon Campbell and DeLiang Wang. Synchronization and sesynchronization in a network of locally coupled wilson-cowan oscillators. *IEEE Trans. Neural Networks*, 7(3):541–553, 1996.

B. Cessac. Absolute stabiltity criterion for discrete time neural networks. *Journal Physics A: Math. Gen*, 27:L927–L930, 1994.

F. Chapeau-Blondeau and G. Chauvet. Stable, oscillatory and chaotic regimes in the dynamics of small neural networks with delay. *Neural Networks*, 5:735–743, 1992.

Tianping Chen and Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995.

Yun-Chung Chu and Keith Glover. Bounds of the induced norm and model reduction errors for systems with repeated scalar nonlinearities. In *IEEE Trans. Automatic Control* [Kaszkurewicz and Bhaya, 1993], pages 471–483.

A. Clark. *Being There: Putting Brain, Body, and World Together Again*. MIT Press, 1997.

A. H. Cohen, S. Rossignol, and S. Grillner, editors. *Neural Control of Rhythmic Movements in Vertebrates*. Wiley, New York, 1988.

M. A. Cohen and S. Grossberg. Absolute stability of global pattern information and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 13:815–826, 1983.

A. Denjoy. Sur les courbes définies par les équations différentielles à la surface du tore. *J. Math.*, 17(4):333–375, 1932.

Robert L. Devaney. *An Introduction to Chaotic Dynamical Systems*. The Benjamin/Cummings Publishing Co., Inc., 1989.

M. di Marco, A. Tesi, and M. Forti. Bifurcations and oscillatory behaviour in a class of competitive cellular neural networks. *International Journal of Bifurcation and Chaos*, 10(6):1267–1293, 2000.

Mauro di Marco, Mauro Forti, and Alberto Tesi. Existence and characterization of limit cycles in nearly symmetric neural networks. *IEEE Trans. Circuits and Systems - Part I: Fundamental Theory and Applications*, 49(7):979–992, 7 2002.

B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. Control of the transition to chaos in neural networks with random connectivity. *Int. Journal of Bifurcation and Chaos*, 3(2):279–291, 1993.

T. Elbert, W. J. Ray, Z. J. Kowalik, J. E. Skinner, K. E. Graf, and N. Birbaumer. Chaos and physiology: Deterministic chaos in excitable cell assemblies. *Physiological Reviews*, 74:1–47, 1994.

J. Feng and K. P. Hadeler. Qualitative behaviour of some simple networks. *Journal of Physics A*, 29:5019–5033, 1996.

Jianfeng Feng. Lyapunov functions for neural nets with nondifferentiable input-output characteristics and symmetric weight matrix. *Neural Computation*, 9(1):43–49, 1997.

R. FitzHugh. Mathematical models of excitation and propagation in nerve. In H. P. Schwan, editor, *Biological Engineering*, pages 1–85. McGraw-Hill, 1969.

Mauro Forti. A note on neural networks with multiple equilibrium points. *IEEE Trans. Circuits and Systems - Part I: Fundamental Theory and Applications*, 43(6):487–491, 1996.

Rainer W. Friedrich and Gilles Laurent. Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity. *Science*, 291:889–893, 2001.

R. M. Golden. Stability and optimization analysis of the generalized brain-state-in-a-box neural network model. *Journal of Mathematical Psychology*, 37:282–298, 1993.

M. Golubitsky and D. G. Schaefer. *Singularities and Groups in Bifurcation Theory*, volume 1. Springer, New York, 1985.

C. M. Gray, P. König, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature*, 338: 334–337, 1989.

Grossberg. Nonlinear neural networks: Priniciples, mechanisms, and architectures. *Neural Networks*, 1:17–61, 1988.

J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer, New York, 1993.

J. Hadamard. Sur l'itération et let solutions asymptotiques des éuations différentialles. *Proc. Soc. Math. France*, 29:224–228, 1901.

Wolfgang Hahn. *Stability of Motion*, volume 138 of *Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen*. Springer, New York, 1967.

J. K. Hale and H. Koçak. *Dynamics and Bifurcations*. Springer, New York, 1991.

Barbara Hammer and Jochen J. Steil. Tutorial: Perspectives on learning with rnns. In M. Verleysen, editor, *Proc. European Symposium on Artificial Neural Networks (ESANN)*, pages 357–368, 2002.

Robert Haschke, Jochen J. Steil, and Helge Ritter. Controlling oscillatory behaviour of a two neuron recurrent neural network using inputs. In Georg Dorffner, Horst Bischof, and Kurt Hornik, editors, *Proc. of the Int. Conf. on Artificial Neural Networks (ICANN)*, pages 1109–1114, Wien, Austria, 2001.

Harro Heuser. *Lehrbuch der Analysis*. Teubner Verlag, Stuttgart, 11. edition, 1994.

Morris W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2:331–349, 1989.

Morris W. Hirsch. Saturation at high gain in discrete time recurrent networks. *Neural Networks*, 7(3):449–453, 1994.

Morris W. Hirsch and Stephen Smale. *Differential equations, dynamical systems, linear algebra*. Academic Press College Division, 1974.

A. L. Hodgkin. The local electric changes associated with repetitive action in a non-medulated axon. *Journal of Physiology*, 1948.

A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544, 1954.

J. J. Hopfield. Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science USA*, 81:3088–3092, 1984.

F. C. Hoppensteadt. *Analysis and Simulation of Chaotic Systems*. Springer, New York, 1993.

F. C. Hoppensteadt and E. M. Izhikevich. Pattern recognition via synchronization in phase-locked loop neural networks. *IEEE Trans. Neural Networks*, 11(3):734–738, 2000.

Frank C. Hoppensteadt and Eugene M. Izhikevich. *Weakly Connected Neural Networks*. Springer, New York, applied mathematical sciences edition, 1997.

Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge Univ. Press, 1991.

Sanqing Hu and Jun Wang. Global stability of a class of discrete-time recurrent neural networks. *IEEE Trans. Circuits and Systems - Part I: Fundamental Theory and Applications*, 49(8):1104–1117, 2002.

Tingshu Hu and Zongli Lin. A complete stability analysis of planar linear systems under saturation. *IEEE Trans. Circuits and Systems*, 48(6):710–725, 2001.

Philip Husbands, Inman Harvey, Dave Cliff, and Geoffrey F. Miller. Artificial evolution: A new path for artificial intelligence? *Brain and Cognition*, 34:130–159, 1997a.

Philip Husbands, Inman Harvey, Nicholas Jakobi, Adrian Thompson, and Dave Cliff. Evolutionary robotics. In Thomas B"ack, David B. Fogel, and Zbigniew Michalewicz, editors, *Handbook of Evolutionary Computation*. Institute of Physics Publishing and Oxford University Press, 1997b.

Auke Jan Ijspeert. A connectionist central pattern generator for the aquatic and terrestrial gaits of a simulated salamander. *Biological Cybernetics*, 84:331–348, 2001.

E. R. Kandel, J. H. Schwartz, and T. M. Jessel. *Principles of Neural Science*. Prentice-Hall International Inc., 3 edition, 1991.

Eugenius Kaszkurewicz and Amit Bhaya. Robust stability and diagonal liapunov functions. *SIAM Journal on Matrix Analysis and Applications*, 14(2):508–520, 1993.

Eugenius Kaszkurewicz and Amit Bhaya. *Matrix Diagonal Stability in Systems and Computation*. Birkhäuser, Boston, 2000.

Tosio Kato. *A short introduction to perturbation theory for linear operators*. Springer-Verlag, New York, 1982.

Yakov Kazanovich and Roman Borisyuk. Object selection by an oscillatory neural network. *BioSystems*, 67:103–111, 2002.

John F. Kolen and Stefan C. Kremer, editors. *A Field Guide to Dynamical Recurrent Networks*. John Wiley & Sons, 2001.

S. C. Kremer. Spatio-temporal connectionist networks: A taxonomy and review. *Neural Computation*, 13(2):249–306, 2001.

A. Kuhn, A. Aertsen, and S. Rotter. Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, 15(1):67–101, 2003.

Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory*, volume 112 of *Applied Mathematical Sciences*. Springer-Verlag, New York, Berlin, Heidelberg, 1995.

Yuri A. Kuznetsov and V. V. Levitin. *CONTENT*. URL `ftp.cwi.nl://pub/CONTENT`.

Peter Lancaster and Miron Tismenetsky. *The Theory of Matrices*. Computer science and applied mathematics. Acad. Press, New York, 1985.

S. Lang. *Introduction to differentiable manifolds*. John Wiley & Sons, New York, 3 edition, 1967.

A. I. Lur'e and V. M. Postnikov. On the theory of stability of controlled systems. *Prikl. Mat.i Mekh*, 8:283–286, 1944.

C. M. Markus and R. M. Westervelt. Dynamics of iterated-map neural networks. *Physical Review A*, 40(1):501–504, 1989.

Patrick C. McGuire, Henrik Bohr, John W. Clark, Rober t Haschke, Chris L. Pershing, and Johann Rafelski. Threshold disorder as a source of diverse and complex behavior in random nets. *Neural Networks*, 15(10):1243–1258, 2002.

L. Meeden and D. Kumar. Trends in evolutionary robotics. In L. C. Jain and T. Fukuda, editors, *Soft Computing for Intelligent Robotic Systems*, pages 215–233. Physica, 1998.

M. Forti and A. Tesi. New conditions for global stability of neural networks with application to linear and quadratic programming problems. *IEEE Trans. Circuits and Systems - Part I: Fundamental Theory and Applications*, 42(7):354–366, 1995.

Zbigniew Michaelewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, Berlin, 3 edition, 1996.

Anthony N. Michel, Jie Si, and Gune Yen. Analysis and synthesis of a class of discrete-time neural networks described on hypercubes. *IEEE Trans. Neural Networks*, 2(1):32–46, 1991.

P. Milner. A model for visual shape recognition. *Psychological Review*, 81:521–535, 1974.

Alexander P. Molchanov and Derong Liu. Robust absolute stability of time-varying nonlinear discrete-time systems. *IEEE Trans. Circuits and Systems – I*, 49(8):1129–1137, 2002.

C. Morris and H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, 35:193–213, 1981.

Hiroyuki Nakahara and Kenji Doya. Near-saddle-node bifurcation behaviour as dynamics in working memory for goal-directed behaviour. *Neural Computation*, 10:113–132, 1998.

K. S. Narendra and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Trans. Neural Networks*, 1(1):4–27, 1990.

Eckart Nelle and Frank Pasemann. *Dynamical Systems - Theory and Applications*, chapter Elements of Nonconvergent Neurodynamics, pages 167–201. World Scientific, Singapore, 1993.

A. Nemirovskii. Several np-hard problems arising in robust stability analysis. *Mathematics of Control, Signals, and Systems*, 6:99–105, 1993.

Randall C. O'Reilly. iologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation*, 1996.

James M. Ortega. *Matrix Theory*. Plenum Press, 1987.

J. Park, H.-Y. Kim, Y. Park, and S.-W. Lee. A synsthesis procedure for associative memories based on space-varying cellular neural networks. *Neural Networks*, 14:107–113, 2001.

Frank Pasemann. Discrete dynamics of two neuron networks. *Open Systems & Information Dynamics*, 2(1):49–66, 1993a.

Frank Pasemann. Dynamics of a single model neuron. *Int. Journal of Bifurcation and Chaos*, 3 (2):271–278, 1993b.

Frank Pasemann. Characterization of periodic attractors in neural ring networks. *Neural Networks*, 8(3):421 – 429, 1995.

Frank Pasemann. A simple chaotic neuron. *Physica D*, 104:205–211, 1997.

Frank Pasemann. Synchronous and asynchronous chaos in coupled neuromodules. *International Journal of Bifurcation and Chaos*, 9(10):1957–1968, 1999.

Frank Pasemann. Complex dynamics and the structure of small neural networks. *Network: Computation in Neural Systems*, 13:195–216, 2002.

Barak A. Pearlmutter. Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Trans. Neural Networks*, 6(5):1212–1228, 1995.

O. Perron. Die stabilitätsfrage bei differentialgleichungen. *Math. Zeitschrift*, 32:703–728, 1930.

J. W. Polderman and J. C. Willems. *Introduction to Mathematical Systems Theory: A Behavioral Approach*. Springer, New York, 1998.

R. F. Port and T. van Gelder, editors. *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, 1995.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes*. Cambridge University Press, 1992.

Govindan Rangarajan, Salman Habib, and Robert D. Ryne. Lyapunov exponents without rescaling and reorthogonalization. *Physical Review Letters*, 80(17):3747–3750, 1998.

Volker Reitmann. *Reguläre und chaotische Dynamik*. Mathematik für Ingenieure und Naturwissenschaftler. Teubner, 1996.

Renals and Rhower. A study of network dynamics. *Journal of Statistical Physics*, 58:825–845, 1990.

J. Rinzel and G. B. Ermentrout. Analysis of neural excitability and oscillations. In C. Koch and I. Segev, editors, *Methods in Neuronal Modelling*. MIT Press, 1989.

Jiri Rohn. Checking positive definiteness or stability of symmetric interval matrices is np-hard. *Commentationes Mathematicae Universitatis Carolinae*, 35:795–797, 1994a.

Jiri Rohn. Positive definiteness and stability of interval matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(1):175–184, 1994b.

Ulf D. Schiller. Analysis and comparison of algorithms for training recurrent neural networks. diploma thesis, Faculty of Technology, University of Bielefeld, 2003.

Ulf D. Schiller and Jochen J. Steil. On the weight dynamcis of recurrent learning. In *Proc. ESANN*, 2003.

George R. Sell and Yuncheng You. *Dynamics of Evolutionary Equations*. Applied Mathematical Sciences. Springer, 2002.

Christine A. Skarda and Walter J. Freeman. How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10:161–195, 1987.

Jochen J. Steil. *Input-Output Stability of Recurrent Neural Networks*. PhD thesis, Technical Faculty, University of Bielefeld, 1999.

Jochen J. Steil. Local structural stability of recurrent networks with time-varying weights. *Neuro-computing*, 48(1-4):39–51, 2002.

C. Stéphanos. Sur une extension du calcul des substitutions linéares. *J. Math. Pures Appl.*, 6: 73–128, 1900.

N. Stollenwerk and F. Pasemann. Control strategies for chaotic neuromodules. *Int. Journal of Bifurcation and Chaos*, 6:693–703, 1996.

A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 1996.

J. M. T. Thompson and H. B. Stewart. *Nonlinear Dynamics and Chaos*. Wiley, 1986.

Peter Tiňo, Bill G. Horne, and C. Lee Giles. Attractive periodic sets in discrete-time recurrent neural networks (with emphasis on fixed-point stability and bifurcations in two-neuron networks). *Neural Computation*, 13:1379–1414, 2001.

A. Tonnelier, S. Meignen, H. Bosch, and J. Demongeot. Synchronization and desynchronization of neural oscillators. *Neural Networks*, 12(9):1213–1228, 1999.

Y. Tsypkin. "frequency criteria for the absolute stability of nonlinear sampled data systems". *Automation and Remote Control*, 25:261–267, 1964.

W. R. Utz. The embedding of homeomorphisms in continous flows. *Top. Proceedings*, 6, 1981.

C. von der Malsburg. The correlation theory of brain function. Technical Report 81-2, MPI Göttingen, 1981.

Peter Wallén, Örjan Ekeberg, Anders Lansner, Lennart Brodin, Hand Tråvén, and Sten Grillner. A computer-based model for realistic simulations of neural networks. ii. the segmental network generating locomotor rhythmicity in the lamprey. *Journal of Neurophysiology*, 68(6):1939–1949, 1992.

D. Wang and D. Terman. Image segmentation based on oscillatory correlation. *Neural Computation*, 9(4):805–836, 1997.

Qing-Guo Wang. Necessary and sufficient conditions for stability of a matrix polytope with normal vertex matrices. *Automatica*, 27(5):887–888, 1991a.

Xin Wang. *Discrete neural networks as dynamical systems*. PhD thesis, Universtity of Southern California, Los Angeles, 1991b.

Xin Wang. Period-doublings to chaos in a simple neural network: An analytic proof. *Complex Systems*, 5:425 – 441, 1991c.

Heiko Wersing, Wolf-Jürgen Beyn, and Helge Ritter. Dynamical stability conditions for recurrent neural networks with unsaturating piecewise linear transfer functions. *Neural Computation*, 13 (8):1811–1825, 2001a.

Heiko Wersing, Jochen J. Steil, and Helge Ritter. A competitive layer model for feature binding and sensory segmentation. *Neural Computation*, 13(2):357–387, 2001b.

H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.

Yang Xiao and Rolf Unbehauen. Robust Hurwitz and Schur stability test for interval matrices. In *39th IEEE Conf. on Decision and Control*, pages 4209–4214, 2000.

V. A. Yakubovich. A frequency theorem in control theory. *Siberian Math. J.*, 14(2):265–289, 1973.

G. Yune and A. N. Michel. A learning and forgetting algorithm in associative memories: The eigenstructure method. *IEEE Trans. Circuits and Systems – II*, 39(4):212–225, 1992.

Fuzhen Zhang. *Matrix Theory: Basic Results and Techniques*. Springer, 1999.

# Index