Stabilitätsuntersuchungen von Neuronalen Netzen unter Verwendung der Theorien dynamischer Systeme

Helko Lehmann

Inhaltsverzeichnis

Einleitung **3**

1 Grundlagen Dynamischer Systeme **5**

1.1 Zeitdiskrete und Zeitkontinuierliche Systeme 5

1.2 Stabilität von dynamischen Systemen 7

1.3 Attraktoren und Dissipation 10

1.4 Lineare Stabilitätsanalyse 11

2 Neuronale Netze **14**

2.1 Vorwärtsgerichtete Netze 15

2.2 Rekurrente Netze 16

3 Neuronale Netze als Dynamische Systeme **18**

3.1 Ein einelementiges „Netz“ 18

3.2 Ein Zweielementiges Netz 22

3.2.1 Fixpunkte in zweielementigen Netzen 22

3.2.2 Chaos in zweielementigen Netzen 23

3.3 Globale Stabilität 26

3.3.1 Das Cohen-Grossberg-Theorem 26

3.3.2 Neuronale Netze als Iterierte Funktionensysteme 28

Zusammenfassung **34**

Literaturverzeichnis **35**

Einleitung

Künstliche Neuronale Netze werden seit ihrer Wiederentdeckung zu Beginn der achziger Jahre sehr vielseitig angewendet. Neben traditionellen Mustererkennungs– und Klassifizierungsaufgaben können Neuronale Netze auch zur Steuerung und Regelung von Prozessen eingesetzt werden, das heißt von zeitlichen Abläufen von mitunter auch unbegrenzter Dauer.

Da *rekurrente* Neuronale Netze prinzipiell in der Lage sind, derartiges zeitabhängiges Verhalten zu erzeugen, ist insbesondere die Untersuchung der Dynamik dieser Netze aktueller Forschungsgegenstand.

Die vorliegende Arbeit beschreibt rekurrente Netze im Rahmen physikalischer Theorien als dynamische Systeme. Damit verbunden ist die Hoffnung, auf vorhandene Ergebnisse bei der Untersuchung anderer dynamischer Systeme zurückgreifen zu können. Es werden verschiedene Methoden der Stabilitätsanalyse auf rekurrente Neuronale Netze angewendet. Dabei zeigt sich, daß die *lineare Stabilitätsanalyse*, eine für viele physikalische Systeme einsetzbare Untersuchungsmethode, bereits für kleine Netze sehr aufwendig und für mittlere oder große Netze nicht praktikabel ist. Eine der Ursachen für diese Schwierigkeit ist die Tatsache, daß kleine Netze bereits ein komplexes Verhalten hervorrufen können, das mittels linearer Stabilitätsanalyse nicht untersucht werden kann. So wird hier beispielhaft ein zweielementiges Netz angegeben, das chaotisches Verhalten zeigt.

Ein anderes Untersuchungsverfahren dynamischer Systeme erlaubt Aussagen über dynamische Eigenschaften des Gesamtsystems. Das Verfahren besteht in der Ermittlung von sogenannten *Ljapunov–Funktionen* oder *Energie-Funktionen*, die einem System zugeordnet werden. Derartige Funktionen können jedoch nur für eingeschränkte Klassen von rekurrenten Netzen gefunden werden.

Bei einigen Netzen liefert auch die Betrachtung als Iteriertes Funktionensystem Aussagen über das Zeitverhalten des Systems. Die Theorie Iterierter Funktionensysteme wurde bisher vor allem zur Untersuchung chaotischer Attraktoren (bzw. von Fraktalen) verwendet. Im Rahmen dieser Arbeit wird ein Formalismus angegeben, in dem rekurrente Netze und Iterierte Funktionensysteme dargestellt werden können. Es wird gezeigt, daß Iterierte Funktionensysteme auch durch rekurrente Netze berechnet werden können und umgekehrt rekurrente Netze als Iterierte Funktionensysteme darstellbar sind. Dabei wird jedoch auch deutlich, daß die Analyse von bestehenden, für eine große Klasse von Anwendungen konzipierten Netzstrukturen, wie z.B. *Jordan–Netze*, dadurch nicht vereinfacht wird.

Die vorliegende Arbeit liefert zunächst im Kapitel 1.1 einen Überblick über die Theorie dynamischer Systeme und über Möglichkeiten zur Gewinnung von Aussagen über das zeitliche Systemverhalten. Das Kapitel ist in vier Abschnitte, die sich mit der formalen Beschreibung dynamischer Systeme, mit den Stabilitätsbegriffen nach Ljapunov und der linearen Stabilitätsanalyse beschäftigen, unterteilt. Eine kurze Beschreibung Neuronaler Netze erfolgt in Kapitel 2. Das umfangreichste Kapitel 3 enthält Abschnitte über die Anwendung der linearen Stabilitätsanalyse auf ein– bzw. zweielementige Neuronale Netze. Im Abschnitt über zweielementige Netze wird auch ein zweielementiges chaotisches Netz angegeben. Texte über das in der Analyse rekurrenter Netze fundamentale *Cohen-Grossberg-Theorem* und über die Interpretation von Netzen als Iterierte Funktionensysteme bilden den Abschnitt zum Thema "Globale Stabilität".

Kapitel 1

Grundlagen Dynamischer Systeme

1.1 Zeitdiskrete und Zeitkontinuierliche Systeme

Im folgenden werden dynamische Systeme betrachtet, die entweder als Systeme gewöhnlicher Differentialgleichungen in Normalform (zeitkontinuierlich)

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} \equiv \dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}) \tag{1.1}$$

oder als Systeme gewöhnlicher Differenzengleichungen (zeitdiskret)

$$\boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t)) \tag{1.2}$$

darstellbar sind[1]. Dabei entspreche $\boldsymbol{x}(t) \equiv (x_1(t), \ldots, x_n(t))$ dem Zustand des Systems zum Zeitpunkt t und $\boldsymbol{F}$ bzw. $\boldsymbol{f}$ seien Abbildungen der Form $\mathbb{R}^n \to \mathbb{R}^n$. Außerdem sei t im zeitkontinuierlichen Fall reellwertig ($t \in \mathbb{R}$) und im zeitdiskreten Fall ganzzahlig ($t \in \mathbb{Z}$). Dynamische Systeme werden hier und im weiteren durch die Gleichungen (1.1) und (1.2) identifiziert, der Begriff *System* bezeichne daher sowohl das physikalische System als auch dessen Beschreibung.

Sind die Funktionen F_i für $i = 1, \ldots, n$ stetig differenzierbar, dann existieren eindeutige Lösungen $\boldsymbol{x}(\boldsymbol{x}_0, t)$ des Differentialgleichungssystems (1.1) für einen Anfangszustand $\boldsymbol{x}_0$. Ein solches System werde *determiniert* genannt, da alle einem Anfangszustand $\boldsymbol{x}_0$ vorhergehenden und nachfolgenden Zustände eindeutig bestimmt sind.

Hat die Funktion $\boldsymbol{f}$ in (1.2) keine eindeutige Inverse, dann sind die vorhergegangenen Zustände nicht eindeutig durch einen Anfangszustand bestimmt; ein solches System heiße *halbdeterminiert*.

Wenn das in Rekursionsform (1.2) dargestellte zeitdiskrete System in eine explizite Form gebracht werden kann, bezeichne man es als *lösbar*.

[1] Definitionen der in diesem Kapitel verwendeten Begriffe findet man z.B. in [11] und [9].

Hängt ein durch $\boldsymbol{F}$ bzw. $\boldsymbol{f}$ modelliertes dynamisches System nicht explizit von der Zeit ab, dann wird es *autonom*, anderenfalls *nichtautonom* genannt. Nichtautonome Systeme können durch Hinzunahme eines weiteren Parameters $x_{n+1}(t) \equiv t$ in autonome Systeme überführt werden.

Eine spezielle Lösung von (1.1) oder (1.2) ordnet einem Zeitpunkt t und einem Anfangszustand $\boldsymbol{x}_0$ eindeutig einen Zustand $\boldsymbol{x}(\boldsymbol{x}_0, t)$ zu. Die Zuordnung der Menge aller möglichen Anfangszustände zu der Menge der Zustände nach der Zeit t heiße *Fluß* oder auch *Phasenfluß* (siehe Abbildung 1.1):

$$\boldsymbol{\varphi}_t : U \rightarrow \mathbb{R}^n \quad \text{mit} \quad U \subset \mathbb{R}^n, \quad \text{so daß} \quad \boldsymbol{\varphi}_t(\boldsymbol{x}_0) = \boldsymbol{x}(\boldsymbol{x}_0, t). \tag{1.3}$$

In (1.3) werde $\mathbb{R}^n$ *Phasenraum* und ein Element aus $\mathbb{R}^n$ *Phasenpunkt* genannt.

Es ist im allgemeinen nicht möglich, die explizite Darstellung des Phasenflusses aus (1.1) und (1.2) herzuleiten. Zum Beispiel entspräche im zeitkontinuierlichen Fall (1.1) die explizite Darstellung des zugehörigen Flusses der allgemeinen Lösung der Differentialgleichung.

$\boldsymbol{\varphi}_t$ definiert bei Nichtinvertierbarkeit von $\boldsymbol{f}$ eine Halbgruppe, ansonsten eine kommutative Gruppe mit der identischen Abbildung $\boldsymbol{\varphi}_0$ als Einselement.

Die Kurve, die man bei festem Anfangszustand $\boldsymbol{x}_0$ und Variation von t erhält, heiße auch *Bahnkurve*, *Orbit*, *Phasenbahn* oder *Trajektorie* des Flusses $\boldsymbol{\varphi}$ zu $\boldsymbol{x}_0$.

In (1.1) kann $\boldsymbol{F}(\boldsymbol{x}(t))$ auch als der Geschwindigkeitsvektor an die Bahnkurve $\boldsymbol{x}$ zum Zeitpunkt t interpretiert werden. Hingegen wird in (1.2) die zeitliche Veränderung des Zustandes direkt durch $\boldsymbol{f}(\boldsymbol{x}(t))$ ausgedrückt.

Da in die Funktionen $\boldsymbol{F}$ und $\boldsymbol{f}$ Parameter $\boldsymbol{r} = (r_1, \ldots, r_k)$ mit $r_i \in \mathbb{R}$ für $i = 1, \ldots, k$ eingehen können, werden $\{\boldsymbol{F}^r\}$ und $\{\boldsymbol{f}^r\}$ auch als Familien dynamischer Systeme bezeichnet. Eine qualitative Änderung des im allgemeinen von $\boldsymbol{r}$ abhängenden Phasenflusses $\boldsymbol{\varphi}_t^r$ sei eine *Bifurkation*. In zeitdiskreten Systemen

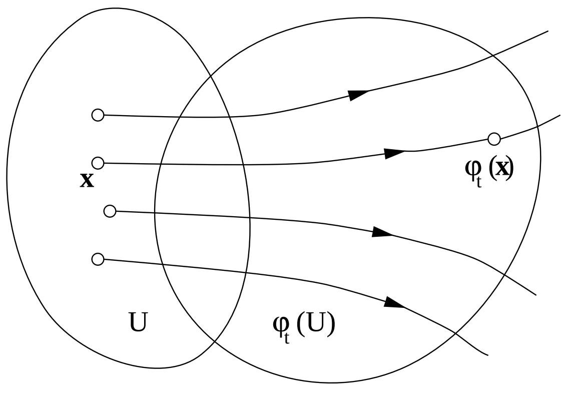


Abbildung 1.1: Phasenfluß $\boldsymbol{\varphi}_t(\boldsymbol{x})$

kann der Parameter t des Flusses $\boldsymbol{\varphi}$ nur bestimmte diskrete Werte annehmen. Ist das diskrete Zeitintervall auf die Größe 1 normiert, so gilt $\boldsymbol{x}_{t+1} = \boldsymbol{\varphi}(\boldsymbol{x}_t)$.

Zeitkontinuierliche Systeme können mitunter in zeitdiskrete Systeme überführt werden. Bei der *stroboskopischen Darstellung* wird der Zustand des Systems nur zu bestimmten Zeitpunkten untersucht. Diese Zeitpunkte sind durch feste Intervalle getrennt. Stroboskopische Darstellungen haben insbesondere dann Sinn, wenn die Zeitabhängigkeit in der Gleichung explizit periodisch ist (z.B. wenn $\boldsymbol{f}$ einen Term $\cos(t)$ enthält). Die gewonnene, um eine Dimension gegenüber der Differentialgleichung reduzierte Differenzengleichung kann nun trotz Vereinfachung qualitative Aufschlüsse über das Systemverhalten geben.

Ähnlich kann eine *Poincaré–Abbildung* angewendet werden. Die zu untersuchenden Zeitpunkte werden hier durch den transversalen Schnitt der Trajektorie mit einer $(n-1)$–dimensionalen Hyperebene im Phasenraum ermittelt. Eine periodische Oszillation eines zeitkontinuierlichen Systems beispielsweise wird als eine endliche Punktmenge in der Hyperebene repräsentiert. Der Poincaré–Schnitt sei das durch die Poincaré–Abbildung auf der Hyperebene generierte Bild (siehe Abbildung 1.2). Periodische Orbits zeitkontinuierlicher Systeme können so auf periodische Orbits zeitdiskreter Systeme abgebildet werden.

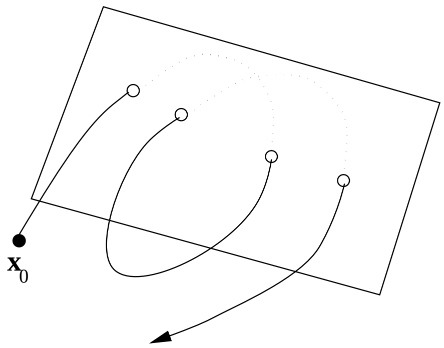


Abbildung 1.2: Poincaré–Schnitt zu einer zeitkontinuierlichen Bahn

1.2 Stabilität von dynamischen Systemen

Der Begriff *Stabilität* kann im allgemeinen zwei unterschiedliche Eigenschaften eines dynamischen Systems bezeichnen. Die eine dieser Eigenschaften beschreibt das qualitative Verhalten des Systems unter veränderten Anfangsbedingungen. Sie wird üblicherweise *Stabilität im Ljapunovschen Sinne* genannt (siehe Abbildung 1.3).

Definition 1 (Ljapunov-Stabilität)
Das System $\boldsymbol{\varphi}(\boldsymbol{u}_0, t_0)$ *ist genau dann stabil, wenn für jede Anfangszeit* t_0 *und jedes feste* $\varepsilon > 0$ *eine Zahl* $\delta > 0$ *existiert, so daß für alle* $t > t_0$ *mit*

$$\varrho_{t_0}(\boldsymbol{\varphi}_{t_0}(\boldsymbol{u}_0, t_0), \boldsymbol{\varphi}_{t_0}(\boldsymbol{v}_0, t_0)) = \varrho_{t_0}(\boldsymbol{u}_0, \boldsymbol{v}_0) < \delta$$

auch gilt:

$$\varrho(\boldsymbol{\varphi}_t(\boldsymbol{u}_0, t_0), \boldsymbol{\varphi}_t(\boldsymbol{v}_0, t_0)) < \varepsilon.$$

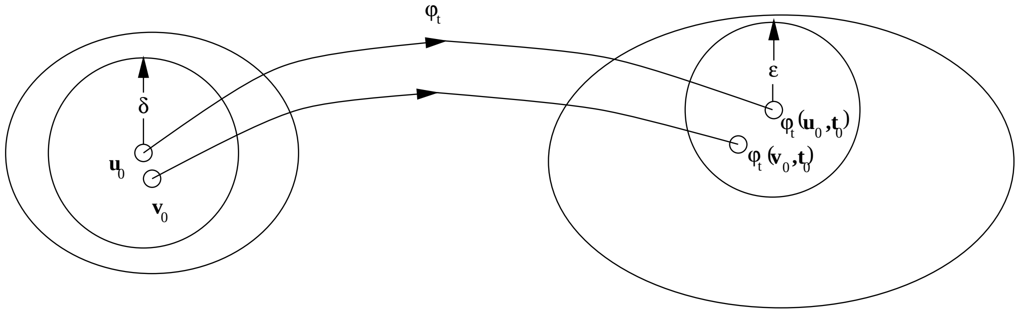


Abbildung 1.3: Ljapunov-Stabilität

Dabei seien ϱ_t und ϱ zwei nicht notwendigerweise identische Normen. Für endlichdimensionale Systeme kann die Unabhängigkeit von der Wahl der Norm mit dem folgenden Theorem gezeigt werden:

Theorem 1 (Prinzip der Normäquivalenz, nach [11])
Eine beliebige Norm ϱ in einem endlichdimensionalen System ist von unten und von oben durch die Euklidische Norm begrenzt:

$$\alpha\|\boldsymbol{x}\| \leq \varrho(\boldsymbol{x}) \leq \beta\|\boldsymbol{x}\| \quad mit \quad \alpha > 0, \beta > 0.$$

Die *Euklidische Norm* sei durch $\|\boldsymbol{x}\| = \sqrt{{x_1}^2 + {x_2}^2 + \ldots + {x_n}^2}$ gegeben.

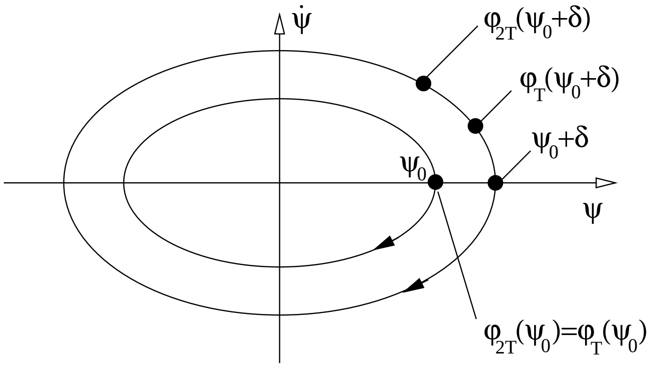
Desweiteren werde als Spezialfall der Ljapunovschen Stabilität noch die *asymptotische Stabilität* unterschieden:

Definition 2 (asymptotische Stabilität)
Wenn $\boldsymbol{\varphi}_t(u_0, t_0)$ im Ljapunovschen Sinne stabil ist und zusätzlich $\varrho(\boldsymbol{\varphi}_t(\boldsymbol{u}_0, t_0), \boldsymbol{\varphi}_t(\boldsymbol{v}_0, t_0)) \to 0$ für $t \to \infty$ gilt, so heißt $\boldsymbol{\varphi}_t(\boldsymbol{u}_0, t_0)$ asymptotisch stabil.

Die Definition der Ljapunovschen Stabilität kann jedoch zu restriktiv sein, um Stabilität allgemein zu beschreiben:

Beispiel 1 (Orbitalstabilität)

Beschreibt man ein ungedämpftes Pendel durch den Ausschlagwinkel $\psi(t)$ und untersucht das Pendel für die Anfangsbedingungen $\psi_1(0) = \psi_0$ und $\psi_2(0) = \psi_0 + \delta$ mit $\dot{\psi}_1(0) = \dot{\psi}_2(0) = 0$, dann stellt man fest, daß der Winkelabstand zwischen $\psi_1(t)$ und $\psi_2(t)$ jedes feste ε, unabhängig von der Wahl des δ, überschreitet. Der Bahnabstand, den man z.B. durch einen Poincaré-Schnitt im durch ψ und $\dot{\psi}$ definierten Phasenraum erhalten kann, bleibt jedoch in Abhängigkeit von t konstant, das System ist also in diesem Sinne durchaus stabil:



□

Auf die zu Beginn des Abschnitts genannte zweite Stabilitätseigenschaft eines Systems zielt die Definition der *Strukturellen Stabilität* bzw. *Robustheit*. Diese Eigenschaft beschreibt, inwieweit das Systemverhalten durch Strukturveränderungen des Systems selbst verändert wird. In der folgenden Definition ist M ein gegebener Phasenraum und $\boldsymbol{v}(\boldsymbol{x})$ ein Vektorfeld.

Definition 3 (strukturelle Stabilität)
Ein dynamisches System $(\boldsymbol{v}, M)$ *wird strukturell stabil genannt, wenn bei einer hinreichend kleinen Störung* $\delta(\boldsymbol{x})$ *des Vektorfeldes* $\boldsymbol{v}$ *das gestörte System* $(\boldsymbol{v} + \boldsymbol{\delta}, M)$ *dem ungestörten System topologisch äquivalent ist.*

Definition 4
Zwei Systeme, die durch die Vektorfelder $\boldsymbol{v}$ *und* $\boldsymbol{v}'$ *beschrieben seien, sind genau dann* topologisch äquivalent*, wenn eine stetige Funktion existiert, die den durch* $\boldsymbol{v}$ *definierten Fluß auf den durch* $\boldsymbol{v}'$ *definierten abbildet und deren Inverse ebenfalls stetig ist.*

Wenn die strukturell stabilen Systeme im Raum der durch die Parameter des Vektorfeldes bestimmten dynamischen Systeme eine offene und dichte Menge bilden, dann ist es möglich, jedes durch Differentialgleichungen beschriebene System durch ein strukturell stabiles System zu approximieren.

Theorem 2 (strukturelle Stabilität, nach [11])
Strukturell stabile Systeme bilden nur für $\dim(M) \leq 2$ *eine offene und dichte Menge im Vektorfeldraum. Wobei* $\dim(M)$ *die Dimension von von* M *angibt.*

Die instabilen Systeme teilen den Parameterraum M in Systemklassen mit robustem Verhalten. Ziel der *Bifurkationstheorie* ist es, den Raum auf höchstens zwei Dimensionen zu reduzieren und damit auch Aussagen über höherdimensionale Systeme zu gewinnen.

Andererseits gibt es auch Klassen höherdimensionaler Systeme, die offenene und dichte Mengen bilden wie z.B. *Gradientensysteme*, mit denen sich die *Katastrophentheorie* beschäftigt.

1.3 Attraktoren und Dissipation

Mit der Wirkung des Phasenflusses $\boldsymbol{\varphi}_t$ auf Teilmengen U des Phasenraumes $\mathbb{R}^n$ kann man einen globalen Überblick über das Systemverhalten gewinnen. Dabei können mitunter auch Aussagen über das System gewonnen werden, ohne daß der Phasenfluß explizit bestimmt werden muß. Die Divergenz des in (1.1) gegebenen Vektorfeldes $\boldsymbol{F}$ des Phasenflusses,

$$\mathrm{div}\boldsymbol{F}(\boldsymbol{x}) \equiv \sum_{i=1}^{n} \frac{\partial F_i(\boldsymbol{x})}{\partial x_i}, \tag{1.4}$$

gibt die Geschwindigkeit der Größenänderung eines infinitesimalen Volumenelements an der Stelle $\boldsymbol{x} \in \mathbb{R}^n$ unter Einwirkung des Flusses $\boldsymbol{\varphi}_t$ an. Ein Gebiet U aus $\mathbb{R}^n$ wird durch den Fluß $\boldsymbol{\varphi}_t$ deformiert.

Theorem 3 (Liouvillescher Satz, nach [9])
Sei $V(t)$ das Volumen von $\boldsymbol{\varphi}_t U$, dann gibt

$$\frac{dV(t)}{dt} = \int_{\boldsymbol{\varphi}_t U} div\boldsymbol{F}(\boldsymbol{x})d\boldsymbol{x}$$

die Volumenänderung zum Zeitpunkt t an.

Für eine konstante Divergenz $\mathrm{div}\boldsymbol{F}$ in $\cup_{t\geq 0}\boldsymbol{\varphi}_t U$ gilt $dV(t)/dt = V(t)\mathrm{div}\boldsymbol{F}$ und

$$V(t) = V(0)\exp(t\,\mathrm{div}\boldsymbol{F}). \tag{1.5}$$

Gilt $\mathrm{div}\boldsymbol{F} = 0$ in $\mathbb{R}^n$, so bleibt das Volumen konstant und das dynamische System werde als *konservativ* bezeichnet. Wird ein beliebiges Gebiet $U \subset \mathbb{R}^n$ durch $\boldsymbol{\varphi}_t$ im zeitlichen Mittel kontrahiert, so heiße das System *dissipativ*.

Beispielsweise sind Systeme dissipativ, wenn im zeitkontinuierlichen Fall

$$\mathrm{div}\boldsymbol{F}(\boldsymbol{x}) < 0 \qquad \forall \boldsymbol{x} \in \mathbb{R}^n \tag{1.6}$$

oder, im zeitdiskreten Fall, für die Funktionalmatrix des Phasenflusses

$$\mathrm{D}\boldsymbol{f} \equiv \begin{pmatrix} \partial f_1/\partial x_1 & \dots & \partial f_1/\partial x_n \\ \vdots & & \vdots \\ \partial f_n/\partial x_1 & \dots & \partial f_n/\partial x_n \end{pmatrix} \tag{1.7}$$

$$|\mathrm{det}\mathrm{D}\boldsymbol{f}(\boldsymbol{x})| < 1 \qquad \forall \boldsymbol{x} \in \mathbb{R}^n \tag{1.8}$$

gilt.

Im folgenden werde unter einem *Attraktor* eines dissipativen dynamischen Systems eine Teilmenge des Phasenraumes verstanden, in der die Bewegung des

Systems nach einer Einschwingphase, auch *transiente* oder *instationäre* Phase genannt, stattfindet. Die Bewegung auf dem Attraktor selbst heißt *stationär* oder *posttransient*. Die Menge der Anfangszustände $\boldsymbol{x}_0$, die durch Wirkung des Phasenflusses den Attraktor erreichen, sei das *Einzugsgebiet* des Attraktors. Trajektorien von außerhalb des Attraktors liegenden Anfangszuständen führen mitunter erst nach unendlicher Einschwingzeit auf den Attraktor. Bei stetigem Phasenfluß $\boldsymbol{\varphi}$ gleicht jedoch die Bewegung in der Nähe des Attraktors der Bewegung auf dem Attraktor.

Fixpunkte

Ein wichtiger Spezialfall eines Attraktors ist ein Fixpunkt. Für den Phasenfluß in einem Fixpunkt $\boldsymbol{x}_s$ gilt:

$$\boldsymbol{\varphi}_t(\boldsymbol{x}_s) = \boldsymbol{x}_s \qquad \text{für alle} \quad t. \tag{1.9}$$

Definition 5
Ein Fixpunkt $\boldsymbol{x}_s$ heißt stabil*, wenn zu einer beliebigen Umgebung U^* von $\boldsymbol{x}_s$ eine Umgebung $U \subset U^*$ von $\boldsymbol{x}_s$ existiert, so daß $\boldsymbol{\varphi}_t U \subset U^*$ für alle $t \geq 0$ gilt.*

Wenn außerdem $\lim_{t\to\infty} \boldsymbol{\varphi}_t \boldsymbol{x}_0 = \boldsymbol{x}_s$ für alle $\boldsymbol{x}_0 \in U$ gilt, so heißt $\boldsymbol{x}_s$ asymptotisch stabil.

Da die Umgebung U nach Definition 5 beliebig klein gewählt werden kann, ist es hinreichend die durch (1.1) und (1.2) beschriebenen Systeme linearisiert zu betrachten. Die linearen Systeme, die aus der Taylorentwicklung gewonnen werden können, lauten dann im zeitkontinuierlichen Fall

$$\dot{\boldsymbol{z}} = \mathrm{D}\boldsymbol{F}(\boldsymbol{x}_s)\boldsymbol{z} \tag{1.10}$$

und wenn das System zeitdiskret beschrieben wird

$$\boldsymbol{z}(t+1) = \mathrm{D}\boldsymbol{f}(\boldsymbol{x}_s)\boldsymbol{z}(t). \tag{1.11}$$

1.4 Lineare Stabilitätsanalyse

Endlichdimensionale zeitkontinuierliche Systeme

Definition 6
Ein Fixpunkt $\boldsymbol{x}_s$ eines zeitkontinuierlichen Systems heißt hyperbolisch*, wenn kein Eigenwert von $D\boldsymbol{F}(\boldsymbol{x}_s)$ rein imaginär oder null ist.*

Ist ein Fixpunkt hyperbolisch, dann kann er mittels der Methode der *Zentrumsmannigfaltigkeit* untersucht werden. Die Ermittlung der Zentrumsmannigfaltigkeit und Verfahren der Eigenwertberechnung können [11] entnommen werden.

Oft genügt dann auch die Untersuchung auf die ersten beiden folgenden Kriterien nach Ljapunov:

Theorem 4 (nach [11])

1. *Ein Fixpunkt* $\boldsymbol{x}_s$ *ist stabil, wenn alle Eigenwerte von* $\boldsymbol{DF}(\boldsymbol{x}_s)$ *negative Realteile haben.*
2. $\boldsymbol{x}_s$ *ist instabil, wenn mindestens ein Eigenwert einen positiven Realteil hat.*
3. *Ein kritischer Fall tritt auf, wenn der Realteil mindestens eines Eigenwertes null ist und alle anderen Eigenwerte negative Realteile besitzen.*

Ist ein Fixpunkt nicht hyperbolisch, dann kann mitunter eine geschlossene Lösung des durch die Linearisierung entstandenen Differentialgleichungssystems gefunden werden und den Nachweis für Stabilität oder Instabilität liefern.

Außerdem kann die globale Stabilität eines Systems auch Aufschluß über lokale Stabilitätseigenschaften geben. Globale Stabilitätseigenschaften gewinnt man, wenn man eine *Ljapunovfunktion* für das System finden kann.

Definition 7
Eine positiv definite Funktion $L(\boldsymbol{x})$ *heißt* Ljapunovfunktion *von* $F(\boldsymbol{x})$ *auf der Menge* U*, falls*

$$\sum_{i=1}^{n} \frac{\partial L(\boldsymbol{x})}{\partial x_i} F_i(\boldsymbol{x}) \leq 0$$

für alle $\boldsymbol{x} \in U$ *gilt.* L *heißt* regulär *in* $\boldsymbol{x}_s$*, wenn* L *in einer Umgebung von* $\boldsymbol{x}_s$ *streng negativ ist.*

Für eine Trajektorie $\boldsymbol{x} = \boldsymbol{x}(t)$ folgt daraus die Beziehung

$$\frac{\mathrm{d}}{\mathrm{d}t} L(x_1(t), \ldots, x_n(t)) \leq 0. \tag{1.12}$$

L wächst also entlang der Trajektorie nicht.

Theorem 5 (nach [11])
Sei $\boldsymbol{x}_s$ *ein Fixpunkt von* $\boldsymbol{F}(\boldsymbol{x})$ *und* L *eine Ljapunovfunktion, die in der Umgebung von* $\boldsymbol{x}_s$ *im* $\mathbb{R}^n$ *stetige erste partielle Ableitungen besitzt.*

1. *Hat* L *in* $\boldsymbol{x}_s$ *ein strenges Minimum, dann ist der Fixpunkt* $\boldsymbol{x}_s$ *stabil.*
2. *Ist* L *zusätzlich in* $\boldsymbol{x}_s$ *regulär, dann ist der Fixpunkt* $\boldsymbol{x}_s$ *asymptotisch stabil.*

Endlichdimensionale zeitdiskrete Systeme

Definition 8
Ein Fixpunkt $\boldsymbol{x}_s$ *eines invertierbaren zeitdiskreten Systems heißt* hyperbolisch, *wenn kein Eigenwert von* $D\boldsymbol{f}(\boldsymbol{x}_s)$ *eins ist.*

Analog zum zeitkontinuierlichen Fall kommt man bei den ersten beiden Stabilitätskriterien ohne Untersuchung der zugehörigen Zentrumsmannigfaltigkeit aus:

Theorem 6 *[nach [11]]*

1. *Ein Fixpunkt $\boldsymbol{x}_s$ ist stabil, wenn die Beträge aller Eigenwerte von $D\boldsymbol{f}(\boldsymbol{x}_s)$ kleiner als eins sind.*
2. *$\boldsymbol{x}_s$ ist instabil, wenn mindestens ein Eigenwert einen Betrag größer eins hat.*
3. *Ein kritischer Fall tritt auf, wenn der Betrag mindestens eines Eigenwertes eins ist und alle anderen Eigenwerte Beträge kleiner eins besitzen.*

Ebenso kann auch bei zeitdiskreten Systemen globale Stabilität mittels einer hier diskreten Ljapunovfunktion $l(t)$ untersucht werden. Dabei gilt dann

$$\Delta l_t = l(\boldsymbol{x}(t)) - l(\boldsymbol{x}(t-1)) \leq 0 \tag{1.13}$$

entsprechend $\frac{\mathrm{d}L}{\mathrm{d}t}$ aus Gleichung (1.12).

Periodische Orbits in zeitkontinuierlichen und zeitdiskreten Systemen

Für periodische Orbits dynamischer Systeme, beschrieben durch die Phasenpunktmenge γ, gilt:

$$\varphi_{t+T}(\boldsymbol{x}) = \varphi_t(\boldsymbol{x}) \quad \text{für alle} \quad t \quad \text{und alle} \quad \boldsymbol{x} \in \gamma. \tag{1.14}$$

Dabei sei T die Periodenlänge, d.h. für alle $\boldsymbol{x} \in \gamma$ und $0 < t < T$ gilt $\varphi_t(\boldsymbol{x}) \neq \boldsymbol{x}$.

In einer geeigneten Poincaré–Abbildung $\boldsymbol{P}$ erscheint der periodische Orbit eines zeitkontinuierlichen Systems als periodischer Orbit oder Fixpunkt des zeitdiskreten Systems $\boldsymbol{P}(\boldsymbol{x})$, so daß $\boldsymbol{x} = \boldsymbol{P}^m(\boldsymbol{x})$ gilt, wobei m die Periodenlänge dieses zeitdiskreten Systems sei. Bezüglich der Funktion $\boldsymbol{P}^m$ ist $\boldsymbol{x}$ also ein Fixpunkt.

Die Stabilitätsuntersuchung eines periodischen Orbits eines zeitkontinuierlichen oder zeitdiskreten Systems kann also auf die Untersuchung eines Fixpunktes eines zeitdiskreten Systems zurückgeführt werden.

Kapitel 2

Neuronale Netze

Im folgenden werde der Begriff *Neuronales Netz* für dynamische Systeme verwendet, die die folgenden Eigenschaften besitzen:

1. Ein Neuronales Netz ist eine (parallele) Architektur, die aus einzelnen Berechnungselementen, *Neuronen* genannt, besteht, die miteinander verbunden sind.

2. Die Berechnungsfunktion der Neuronen kann durch eine nichtlineare algebraische Gleichung oder eine Differentialgleichung beschrieben werden.

3. Jeder Verbindung zwischen zwei Neuronen ist ein veränderbarer Parameter zugeordnet.

Üblicherweise wird ein Neuronales Netz durch seine Struktur und einen zugeordneten Lernalgorithmus definiert. Dabei beschreibt der Lernalgorithmus Veränderungen der Parameter, die den Verbindungen zugeordnet sind. In der vorliegenden Arbeit wird jedoch nur das durch feste Verbindungsparameter determinierte Systemverhalten untersucht. Das Verhalten eines Neurons eines Netzes sei im folgenden durch die Differentialgleichung

$$\frac{\mathrm{d}}{\mathrm{d}t} x_i = a_i(x_i) \left[b_i(x_i) - \sum_{j=1}^{n} c_{ij} d_j(x_j) \right] \tag{2.1}$$

hinreichend beschrieben.

Dabei seien a_i, b_i Funktionen der Aktivierung x_i des i–ten Neurons und d_j die dem j-ten Neuron zugeordnete *Schwellwertfunktion* nach Definition 9. c_{ij} ist der konstante Parameter, der der Verbindung zwischen dem i-ten und dem j-ten Neuron zugeordnet ist. Im allgemeinen seien a_i, b_i, c_{ij} und x_i reelle Zahlen und a_i, b_i, d_j sind Funktionen der Form $\mathbb{R} \to \mathbb{R}$.

Definition 9
Gelten für eine nichtkonstante Funktion $f : \mathbb{R} \to \mathbb{R}$ *die folgenden Bedingungen:*

1. *Monotonie:* $f'(x) \geq 0$,
2. *Beschränktheit:* $lim_{x \to \infty} = s_u$ *und* $lim_{x \to -\infty} = s_l$, *wobei* s_u *obere und* s_l *untere Schranken der Funktion* f *sind,*

dann heißt f Schwellwertfunktion.

Bei der Untersuchung spezieller Architekturen werden stark vereinfachte Formen der Gleichung (2.1) benötigt. Da eine wichtige Aussage über globale Stabilität, das *Cohen–Grossberg–Theorem*, in dieser allgemeinen Form beweisbar ist, wird jedoch auch an dieser Stelle die angegebene Notation verwendet.

Viele Architekturen Neuronaler Netze werden auch zeitdiskret beschrieben. In diesem Fall werde eine Version der Differenzengleichung

$$x_i^{t+1} = a_i(x_i^t)\left[b_i(x_i^t) - \sum_{j=1}^{n} c_{ij} d_j(x_j^t)\right] \tag{2.2}$$

verwendet, die sich für Stabilitätsuntersuchungen mit dem Cohen-Grossberg-Theorem im allgemeinen auf Gleichung (2.1) zurückführen läßt.

2.1 Vorwärtsgerichtete Netze

Seien $\mathrm{from}(c_{ij}) = j$ und $\mathrm{to}(c_{ij}) = i$ Funktionen, die den Index des Anfangs- bzw. Endneurons einer Verbindung angeben.

Definition 10
Sind in den Gleichungen (2.1) bzw. (2.2) die Parameter c_{ij} *derart gewählt, so daß jede Sequenz von Verbindungen* $c_1, \ldots, c_m$ *mit* $\mathrm{from}(c_1) = \mathrm{to}(c_m)$ *und* $\mathrm{to}(c_k) = \mathrm{from}(c_{k+1})$ *(für* $k = 1, \ldots, m-1$*) mindestens eine Verbindung* $c_l = 0$ *(*$1 \leq l \leq m$*) enthält und sowohl alle* a *als auch alle* b *konstant sind, dann heißt das durch die Verbindungsparameter* c_{ij} *beschriebene Netz* vorwärtsgerichtet.

Vorwärtsgerichtete Netze enthalten also weder direkte noch indirekte Rückkopplungen. Damit erreicht das Netz immer nach endlicher, durch die Anzahl der Neuronen beschränkter Zeit einen Fixpunkt. In den folgenden Kapiteln wird daher nur auf rückgekoppelte (rekurrente) Neuronale Netze eingegangen.

Häufig werden vorwärtsgerichtete Netze betrachtet, die eine Schichtenstruktur besitzen. Die Berechnungselemente werden dabei so gruppiert, daß jede Gruppe nur Verbindungen zu genau einer vorhergehenden Gruppe und zu genau einer nachfolgenden Gruppe besitzt (siehe Abbildung 2.1).

Nach [5] sind dreischichtige vorwärtsgerichtete Neuronale Netze in der Lage jede *Borel–meßbare* Funktion beliebig genau zu approximieren, wenn in der mittleren Schicht ausreichend viele Neuronen existieren und die den Neuronen zugeordneten Aktivierungsfunktionen ($d_j(x_j)$ in den Gleichungen (2.1) und (2.2)) Schwellwertfunktionen sind. Zu den Borel–meßbaren Funktionen gehören z.B. auch alle stetigen Funktionen.

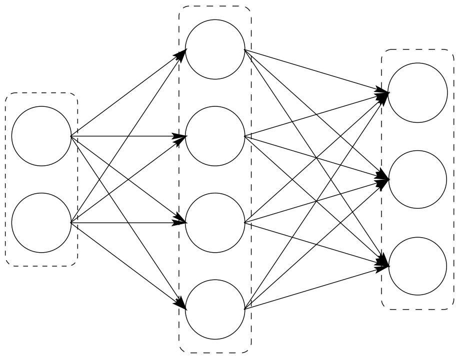


Abbildung 2.1: Mehrschichtiges vorwärtsgerichtetes Netz.

2.2 Rekurrente Netze

Ist in einem Neuronalen Netz eine der in Definition 10 angegebenen Bedingungen verletzt, dann heißt das Netz *rekurrent*. In einem rekurrenten Netz existiert daher mindestens ein Neuron, dessen aktueller Zustand den Zustand dieses Neurons zu einem späteren Zeitpunkt beeinflussen kann.

Wie in Beispielen noch zu zeigen sein wird, sind rekurrente Netze im Gegensatz zu vorwärtsgerichteten zu sehr komplexem dynamischen Verhalten in der Lage. So kann ein solches Netz neben Fixpunkten auch periodische, quasiperiodische oder auch chaotische Attraktoren besitzen.

Besitzen in einem rekurrenten Netz alle Elemente c_{ij} der Gewichtsmatrix Werte verschieden von Null, dann heißt das Netz auch *vollständig verbunden*.

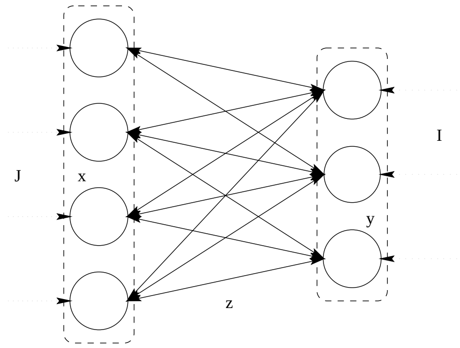


Abbildung 2.2: Gewichtsmatrix in einem Bidirektionalen Assoziativen Speicher.

Beispiel 2 (Bidirektionaler Assoziativer Speicher)

Typische Vertreter rekurrenter Netze, die als Assoziativspeicher verwendet werden, sind Bidirektionale Assoziative Speicher, *die von Kosko in [8] vorgestellt* *wurden. Diese Netze bestehen aus zwei Schichten von Neuronen, die ausschließlich über Verbindungen zur jeweils anderen Schicht verfügen. Das Gleichungssystem (2.1) kann daher in zwei Gleichungssysteme zerlegt werden, so daß für den Informationsfluß von der ersten zur zweiten Schicht*

$$\frac{d}{dt} y_j = -A_j y_j + \sum_{k=1}^{n} f_k(y_k) z_{kj} + I_j \tag{2.3}$$

und von der zweiten zur ersten Schicht

$$\frac{d}{dt} x_i = -B_i x_i + \sum_{l=1}^{n} g_l(y_l) z_{il} + J_i \tag{2.4}$$

gilt (siehe Abbildung 2.2), wobei die Konstanten J *und* I *Eingaben an das System und die Funktionen* f *bzw.* g *Schwellwertfunktionen wie beispielsweise die Sigmoide repräsentieren. Da jedes Neuron* x_i*, daß mit dem Gewicht* z_{ij} *durch das Neuron* y_j *beeinflußt wird mit dem gleichen Gewicht* z_{ij} *das Neuron* y_j *beeinflußt, ist die aus den Gewichten zwischen den Neuronen* $(x_1, \ldots, x_m, y_1, \ldots, y_n)$ *gebildete Matrix der Dimension* $(m+n) \times (m+n)$ *symmetrisch.*

$\square$

Kapitel 3

Neuronale Netze als Dynamische Systeme

3.1 Ein einelementiges „Netz“

Das einfachste rückgekoppelte neuronale Netz besteht aus einem einzelnen Neuron mit einer Rückführung seines Ausgangs auf seinen Eingang. Die Berechnung der Ausgabe erfolge zeitdiskret, d.h. das Neuron iteriert die Aktivierungsfunktion. Das Verhaltenspotential eines derart einfachen Systems werde hier exemplarisch unter Verwendung der Sigmoide

$$\sigma(x) = (1 + e^{-ax+b})^{-1} \tag{3.1}$$

als Aktivierungsfunktion untersucht.

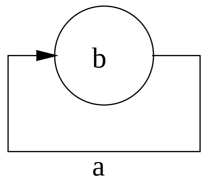


Abbildung 3.1: Neuron mit Schwellwert b und Rückkopplungsgewicht a

Die erste Ableitung von $\sigma(x)$ lautet

$$\sigma'(x) = a(1 + e^{-ax+b})^{-2} e^{-ax+b} = a\sigma(x)(1 - \sigma(x)). \tag{3.2}$$

Das Verhalten des Neurons spiegelt sich u.a. in der Existenz unterschiedlicher Attraktoren im Phasenraum wieder (hier sei der Phasenraum durch die möglichen Ausgabewerte, d.h. $0 < \sigma(x) < 1$, definiert), denen sich das System während des Iterationsprozesses je nach gewähltem Anfangszustand x_0 nähert.

Im Graphen der Sigmoide entsprechen die Schnittpunkte mit der Geraden $y = x$ den Fixpunkten. Drei Regionen qualitativ unterschiedlichen Systemverhaltens

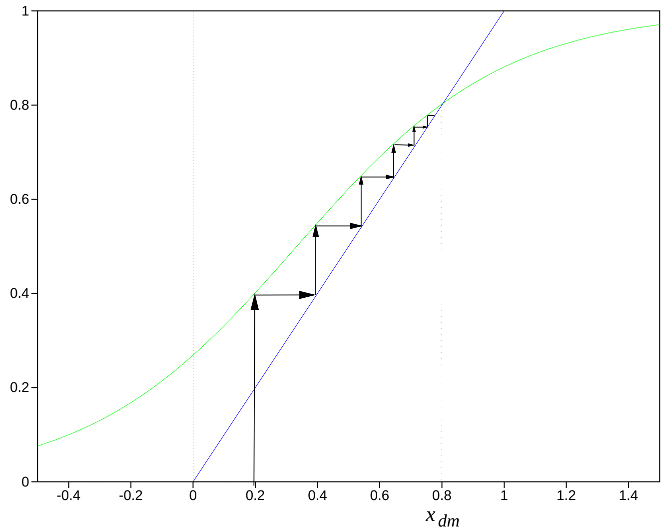


Abbildung 3.2: Iterierte Sigmoide mit einem Fixpunkt, Parameter $a = 3$ und $b = 1$, Startpunkt $x_0 = 0.2$ (Region 1)

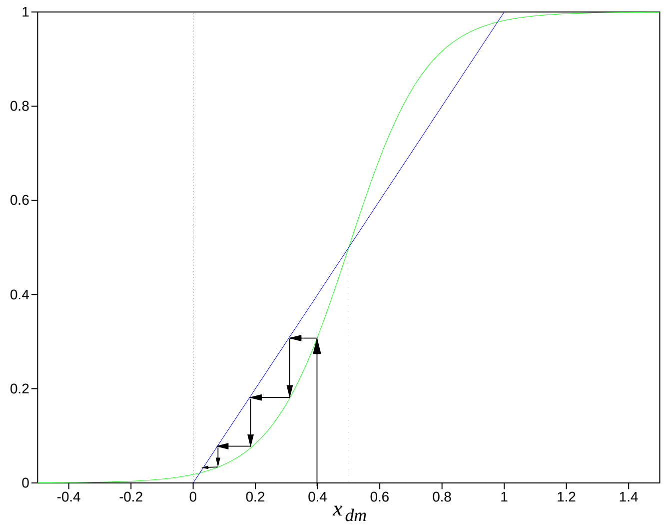


Abbildung 3.3: Iterierte Sigmoide mit drei Fixpunkten, Parameter $a = 8$ und $b = 4$, Startpunkt $x_0 = 0.4$ (Region 2)

im durch die Parameter a und b aufgespannten Raum sind leicht durch die Klassifizierung des Schnittpunktes, in dem der Betrag des Anstiegs der Sigmoiden maximal ist (x_{dm}), voneinander abgrenzbar (siehe Abbildung 3.4 und Abbildung 3.6).

Die Bifurkationspunkte, an denen eine qualitative Verhaltensänderung des Neurons stattfindet, bilden Kurven im Parameterraum und grenzen die verschiede-

Fall	Anstieg	Stabilität von x_{dm}	weitere Attraktoren
1	$-1 < \sigma'(x_{dm}) < 1$	stabil	keine
2	$\sigma'(x_{dm}) > 1$	instabil	zwei stabile Fixpunkte
3	$\sigma'(x_{dm}) < -1$	instabil	ein stabiler 2-periodischer Attraktor

Abbildung 3.4: Charakterisierung von Regionen unterschiedlichen Systemverhaltens entsprechend des Fixpunktes x_{dm}.

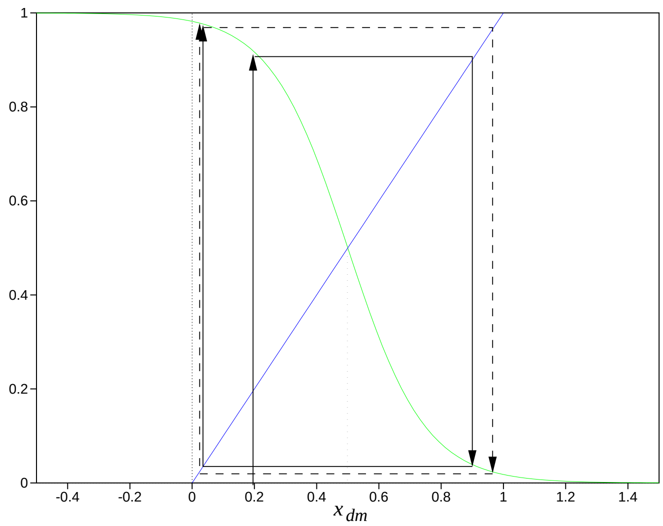


Abbildung 3.5: Iterierte Sigmoide mit einem Fixpunkt und einem periodischen Attraktor, Parameter $a = -8$ und $b = -4$, Startpunkt $x_0 = 0.2$ (Region 3)

nen Regionen voneinander ab. Die Grenzkurven können aus der Fixpunktbedingung

$$\sigma(x_f) = x_f \tag{3.3}$$

und den Aussagen über die erste Ableitung entsprechend Abbildung 3.4 gewonnen werden. Für die Grenzkurven, die die Regionen trennt, gilt daher neben (3.3) auch:

Grenzkurve von Region 1 und 2	Grenzkurve von Region 2 und 3
$\sigma'(x_f) = a\sigma(x_f)(1-\sigma(x_f)) = 1$	$\sigma'(x_p) = a\sigma(x_p)(1-\sigma(x_p)) = -1.$

Daraus folgt

$$\sigma_{f_{1,2}} = \frac{1 \pm \sqrt{1 - \frac{4}{a}}}{2} \qquad \sigma_{p_{1,2}} = \frac{1 \pm \sqrt{1 + \frac{4}{a}}}{2}$$

und mit 3.1 erhält man die Grenzkurven

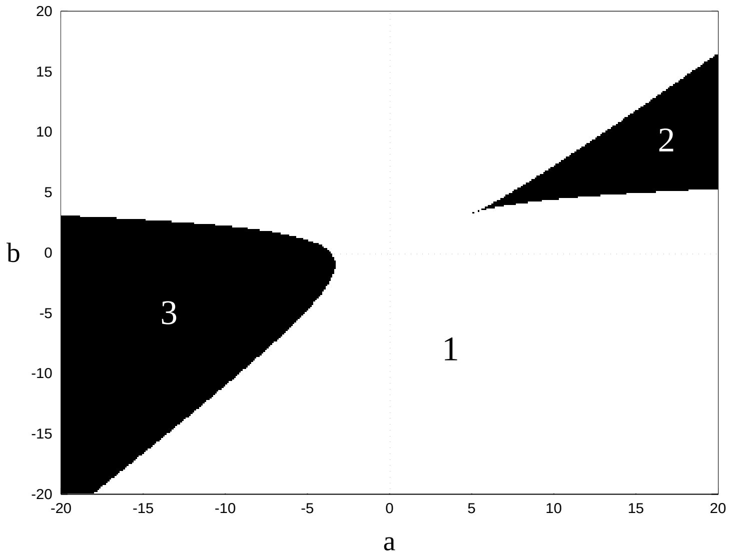


Abbildung 3.6: Regionen im Parameterraum.

$$b = a\sigma_{f_{1,2}} + ln(\frac{1}{\sigma_{f_{1,2}}} - 1) \qquad\qquad b = a\sigma_{p_{1,2}} + ln(\frac{1}{\sigma_{p_{1,2}}} - 1).$$

Für die gesamten Regionen gilt dann

$$a\sigma_{f_2} + ln(\frac{1}{\sigma_{f_2}} - 1) < b < a\sigma_{f_1} + ln(\frac{1}{\sigma_{f_1}} - 1) \tag{3.4}$$

bzw.

$$a\sigma_{p_2} + ln(\frac{1}{\sigma_{p_2}} - 1) < b < a\sigma_{p_1} + ln(\frac{1}{\sigma_{p_1}} - 1). \tag{3.5}$$

Der Parameterraum des einelementigen "Netzes" ist nun durch die Grenzkurven (3.4) und (3.5) vollständig klassifiziert. Da die der Region 2 zugeordneten Systeme zwei unterschiedliche stabile Attraktoren besitzen, hängt deren Verhalten von einem Anfangswert x_0 ab. Gilt $x_0 < x_{dm}$, dann konvergiert x gegen den kleineren der beiden stabilen Fixpunkte (siehe auch Abbildung 3.3). Gilt hingegen $x_0 > x_{dm}$, dann konvergiert x gegen den größeren (x_{dm} ist wie in Region 3 instabil). Der Anfangswert kann z.B. auch durch eine einmalige Eingabe I an das System zum Zeitpunkt t_0 bestimmt werden. Die Aktivierung des Neurons zum Zeitpunkt t_0 entspricht dann $x_0 = b + I$.

Wird I dem System permanent zugeführt, dann kann I auch als Veränderung des Schwellwertes b interpretiert werden und das Systemverhalten entspricht dem des Systems zum Punkt $(a, b + I)$ im Parameterraum.

3.2 Ein Zweielementiges Netz

3.2.1 Fixpunkte in zweielementigen Netzen

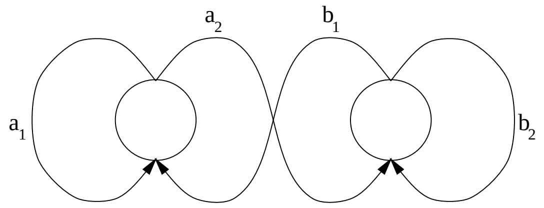


Abbildung 3.7: Rekurrentes Netz mit zwei Neuronen.

Zunächst werde entsprechend der linearen Stabilitätsanalyse aus Abschnitt 1.4 das aus zwei vollständig vernetzten Neuronen bestehende dynamische System

$$x_i^{t+1} = \sigma_i(\boldsymbol{x^t}) = (1 + e^{-a_i x_1^t + b_i x_2^t})^{-1} \qquad i = 1,2 \tag{3.6}$$

durch Taylorentwicklung an der Stelle $\boldsymbol{x}_s$ linearisiert:

$$\boldsymbol{z}(t+1) = D\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{z}(t) = \begin{pmatrix} \frac{\delta\sigma_1(\boldsymbol{x})}{\delta x_1} & \frac{\delta\sigma_1(\boldsymbol{x})}{\delta x_2} \\ \frac{\delta\sigma_2(\boldsymbol{x})}{\delta x_1} & \frac{\delta\sigma_2(\boldsymbol{x})}{\delta x_2} \end{pmatrix}_{\boldsymbol{x}=\boldsymbol{x}_s} \boldsymbol{z}(t) \tag{3.7}$$

$$\boldsymbol{z}(t+1) = \begin{pmatrix} a_1\sigma_1(1-\sigma_1) & b_1\sigma_1(1-\sigma_1) \\ a_2\sigma_2(1-\sigma_2) & b_2\sigma_2(1-\sigma_2) \end{pmatrix}_{\boldsymbol{x}=\boldsymbol{x}_s} \boldsymbol{z}(t) \tag{3.8}$$

Durch Bestimmung der Eigenwerte der charakteristischen Gleichung

$$\begin{vmatrix} a_1\sigma_1(1-\sigma_1)-\lambda & b_1\sigma_1(1-\sigma_1) \\ a_2\sigma_2(1-\sigma_2) & b_2\sigma_2(1-\sigma_2)-\lambda \end{vmatrix}_{\boldsymbol{x}=\boldsymbol{x}_s} = 0 \tag{3.9}$$

wird die Quadratische Gleichung gewonnen:

$$\lambda_{1,2} = \frac{a_1\sigma_2^* + b_2\sigma_1^* \pm \sqrt{(a_1\sigma_2^* - b_2\sigma_1^*)^2 + 4a_2b_1\sigma_1^*\sigma_2^*}}{2}, \tag{3.10}$$

mit $\sigma^* = \sigma(\boldsymbol{x}_s)(1-\sigma(\boldsymbol{x}_s))$.

Nach den Bedingungen für stabile Fixpunkte in Theorem 6 muß gelten:

$$|\lambda_{1,2}| < 1 \tag{3.11}$$

Beispielhaft sei in der Gleichung 3.10 der Fall $a_1, b_2 > 0$ näher untersucht. Durch Umformungen erhält man die Ungleichung

$$k(\sigma_1*, \sigma_2*) = (a_1 b_2 - a_2 b_1)\sigma_1^* \sigma_2^* - a_1 \sigma_1^* - b_2 \sigma_2^* + 1 > 0 \tag{3.12}$$

Analog zur Vorgehensweise in Abschnitt 3.1 kann daraus eine Grenzkurve für $k(\sigma_1*, \sigma_2*) = 0$ ermittelt werden, die Regionen, in denen mindestens ein Eigenwert λ_1 oder λ_2 größer eins ist und in denen beide Eigenwerte kleiner eins sind, voneinander trennt.

$$\sigma_2^* = \frac{a_1 \sigma_1^*}{(a_1 b_2 - a_2 b_1)\sigma_1^* - b_2} - \frac{1}{(a_1 b_2 - a_2 b_1)\sigma_1^* - b_2} \tag{3.13}$$

Neben dieser Grenzkurve können auf ähnliche Weise weitere Grenzkurven z.B. für $a_1, b_2 < 0$ gewonnen werden. Ausführliche Untersuchungen diskreter aus zwei Neuronen bestehender Systeme wurden von Tino in [10] durchgeführt.

Am Beispiel dieses kleinen Netzes wird bereits deutlich, wie aufwendig die lineare Stabilitätsanalyse für sehr niedrigdimensionale Systeme ist. Wie im nächsten Abschnitt gezeigt wird, kann ein zweielementiges Netz außerhalb der Regionen im Parameterraum, in denen stabile Fixpunkte existieren außerordentlich komplexes Verhalten zeigen.

Stabilitätsuntersuchungen von periodischen Attraktoren und Ermittlung von Grenzkurven für zweielementige Netze gibt es bisher nicht.

3.2.2 Chaos in zweielementigen Netzen

Für zweielementige rückgekoppelte Netze gelte wie in Gleichung 3.6 (hier in Matrixnotation):

$$\boldsymbol{x}^{t+1} = \boldsymbol{\sigma}(W\boldsymbol{x}) = \boldsymbol{\sigma}\left(\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \boldsymbol{x}^t\right)$$

Wang untersuchte in [12] einen Spezialfall dieser Gewichtsmatrix

$$W_g = a \begin{pmatrix} b & lb \\ c & lc \end{pmatrix}. \tag{3.14}$$

Er transformierte die durch W_g beschriebene Funktionsklasse in eine topologisch äquivalente Familie von eindimensionalen Funktionen, in der Systeme mit chaotischem Verhalten enthalten sind. Dieses Vorgehen ist möglich, da die Matrix W_g zwei linear abhängige Spaltenvektoren besitzt. Die Matrix bildet also eine Fläche auf ein Geradenstück ab. Gilt für den Parameter $l \neq 1$, dann ist die Abbildung nicht symmetrisch:

$$\begin{pmatrix} b & lb \\ c & lc \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} bx_1 + lbx_2 \\ cx_1 + lcx_2 \end{pmatrix} \neq \begin{pmatrix} b & lb \\ c & lc \end{pmatrix} \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} bx_2 + lbx_1 \\ cx_2 + lcx_1 \end{pmatrix} \tag{3.15}$$

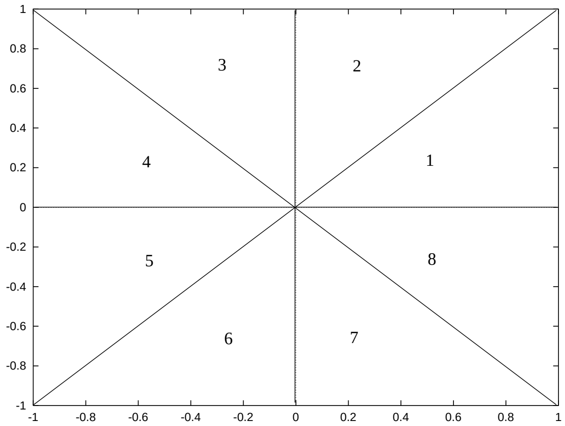


Abbildung 3.8: Regionen, in denen die Spaltenvektoren und damit auch die Geraden, auf die W_g bzw. W_s abbilden, liegen können.

Ist W_g eine in diesem Sinn symmetrische Abbildung, dann kann W_g auch wie folgt dargestellt werden:

$$W_s = a \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21}k & \lambda_{22}k \end{pmatrix}; \quad \lambda_i \in \{-1, 1\}; \quad a, k > 0 \qquad (3.16)$$

Die experimentelle Untersuchung der Wirkung von W_s für unterschiedliche Wahl der Parameter λ_{ij} und k zeigt, daß in Abhängigkeit von der Lage der Spaltenvektoren in den verschiedenen Regionen des Koordinatensystems (siehe Abbildung 3.8) unterschiedliche Bifurkationsdiagramme bei Variation des Parameters a erzeugt werden. Gilt $\lambda_{11} = \lambda_{12}$ und $\lambda_{21} = \lambda_{22}$, dann befinden sich die Geradenstücke, auf die die Matrix W_s abbildet vollständig in einer Region des Koordinatensystems (die Spaltenvektoren beschreiben im allgemeinen das Parallelogramm, auf das das Einheitsquadrat bei der Lineartransformation abgebildet wird, im hier untersuchten Fall entartet das Parallelogramm zu einer Geraden).

λ_{11}	λ_{12}	λ_{21}	λ_{22}	k	Beschreibung	Regionen
1	-1	1	-1	$0 < k < 1$	Periodenverdopplung (halb)	1,5
1	-1	1	-1	$1 < k$	Periodenverdopplung (voll)	2,6
-1	1	1	-1	$1 < k$	Hopf-Bifurkation (voll)	3,7
-1	1	1	-1	$0 < k < 1$	Hopf-Bifurkation (voll)	4,8
-1	1	-1	1	$0 < k < 1$	Periodenverdopplung (voll)	5,1
-1	1	-1	1	$1 < k$	Periodenverdopplung (halb)	6,2
1	-1	-1	1	$1 < k$	Fixpunkt	7,3
1	-1	-1	1	$0 < k < 1$	Fixpunkt	8,4

Die Angaben *voll* und *halb* hinter der qualitativen Beschreibung des Bifurkationsdiagramms geben an, ob der gesamte Phasenraum oder nur der halbe

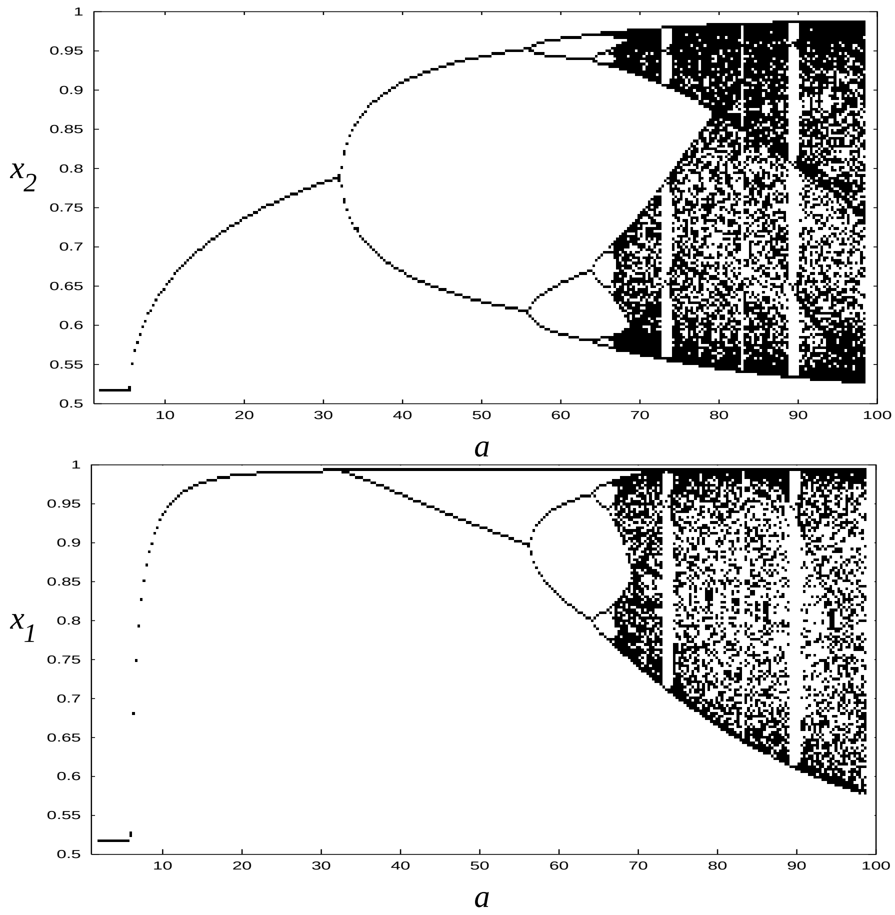


Abbildung 3.9: Bifurkationsdiagramme der beiden Elemente des 2-Neuronen-Systems mit $\lambda_{11} = \lambda_{21} = -1$, $\lambda_{12} = \lambda_{22} = 1$, $k = 5$ und Veränderung des Parameters a. Die Diagramme zeigen typische Periodenverdopplungen bis hin zu chaotischem Verhalten.

durchlaufen wird. Wird nur der halbe Phasenraum durchlaufen, so kann durch andere Wahl des Startwertes $\boldsymbol{x}_0$ ein zweiter Attraktor in der anderen Hälfte des Phasenraums erreicht werden.

In Abbildung 3.9 sind beispielhaft die Bifurkationsdiagramme für den von Wang beschriebenen chaotischen Fall des Systems angegeben. Mit der Wahl des Startwertes $\boldsymbol{x}_0 = (0.2, 0.7)$ wird der Attraktor in der oberen Hälfte des Phasenraums angelaufen. Wählt man hingegen $\boldsymbol{x}_0 = (0.7, 0.2)$, so strebt das System zu einem Attraktor in der unteren Hälfte des Phasenraums. Das entsprechende Bifurkationsdiagramm liegt spiegelbildlich bezüglich der Geraden $x = 0.5$ zu dem Diagramm in der oberen Hälfte. Die Spaltenvektoren der Matrix W_s liegen in den Regionen 6 bzw. 2.

3.3 Globale Stabilität

Wie in den vorangegangegen Abschnitten gezeigt wurde, sind Untersuchungen des dynamischen Verhaltens sehr kleiner Neuronaler Netze mittels linearer Stabilitätsanalyse möglich. Für zweielementige Netze ist die Ermittlung der Regionen im Parameterraum, in denen ein derartiges Netz genau einen stabilen Fixpunkt besitzt, bereits sehr aufwendig und in voller Allgemeinheit bisher nicht realisiert worden. Die Identifikation von Parameterregionen, in denen das zugeordnete System kompliziertere periodische oder auch chaotische Attraktoren besitzt, ist nur in Ausnahmefällen möglich.

Besonders wichtig bei der Anwendung Neuronaler Netze ist außerdem die Abhängigkeit der Dynamik von einer Menge von Parametern, die als Eingabesequenz oder Startwerte das Systemverhalten beeinflussen.

Für einige Klassen von Neuronalen Netzen läßt sich durch Angabe einer Ljapunov-Funktion entsprechend Definition 7 nachweisen, daß das Netz für beliebige Startwerte einen Fixpunkt erreicht. Cohen und Grossberg konnten in [2] eine Ljapunovfunktion finden, anhand derer globale Stabilitätsbedingungen für eine große Klasse von Netzen formulierbar sind.

Häufig werden die Berechnungsmöglichkeiten rekurrenter Netze auch mit denen von deterministischen Automaten wie z.B. Turingmaschinen verglichen. Teilmengen des durch die Neuronenzustände gebildeten Zustandsraums rekurrenter Netze werden mit Zuständen von deterministischen Automaten zu identifizieren versucht. Obwohl diese Sicht bei der Konstruktion rekurrenter Netze, die über die Funktionalität endlicher Automaten verfügen sollen, sinnvoll ist, sind Modelle endlicher deterministischer Automaten für die Analyse rekurrenter Netze, die potentiell über einen unendlichen Zustandsraum verfügen (z.B. bei reellwertiger Aktivierung), ungeeignet (siehe [7]).

Im Abschnitt 3.3.2 wird eine alternative Analysemethode mittels Iterierter Funktionensysteme nach [7] vorgestellt. Diese Analysemethode wird ebenso wie die Modellierung als Automaten nur zeitdiskrete rekurrente Netze behandeln. Eine Erweiterung auf zeitkontinuierliche Systeme ist auch hier denkbar, auf sie wird jedoch in der vorliegenden Arbeit nicht eingegangen.

3.3.1 Das Cohen-Grossberg-Theorem

Die von Cohen und Grossberg vorgeschlagene globale Ljapunov-Funktion für rekurrente Netze lautet:

$$L(\boldsymbol{x}) = -\sum_{i=1}^{n} \int_0^{x_i} b_i(\xi_i) d_i'(\xi_i) \mathrm{d}\xi_i + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk} d_j(x_j) d_k(x_k) \tag{3.17}$$

Dabei müssen unter anderem folgende Bedingungen in Gleichung 2.1 eingehalten werden (siehe [4]):

1. $c_{ij} = c_{ji}$;

2. $a_i(x) \geq 0$;
3. $d_i(x) \geq 0$ für $x \in \mathbb{R}$ ist differenzierbar und monoton, $d_i'(x) \geq 0$.

Sind die Bedingungen erfüllt, dann hat die Ljapunov-Funktion $L(\boldsymbol{x})$ die Ableitung $\dot{L} \leq 0$ der Form:

$$\dot{L}(\boldsymbol{x}) = -\sum_{i=1}^{n} a_i(x_i) d_i'(x_i)[b_i(x_i) - \sum_{j=1}^{n} c_{ij} d_j(x_j)]^2 \tag{3.18}$$

Beispiel 3 (McCulloch-Pitts Modell)

$$x_i(t+1) = sgn\left(\sum_{j=1}^{n} A_{ij} x_j(t) - B_i\right) \tag{3.19}$$

Mit $M(w) = sgn(w - B_i)$ kann Gleichung 3.19 als

$$x_i(t+1) = M\left(\sum_{j=1}^{n} A_{ij} x_j(t) - B_i\right) \tag{3.20}$$

umgeschrieben werden und mit der Ersetzung

$$y_i = \sum_{j=1}^{n} A_{ij} x_j(t) \tag{3.21}$$

und der Umformung des diskreten Modells in ein kontinuierliches Modell erhält man

$$\frac{d}{dt} y_i = y_i + \sum_{j=1}^{n} A_{ij} M(y_j) \tag{3.22}$$

und damit die Ljapunov-Funktion, die für $b_i(x) = 0$ der Form aus Gleichung 3.17 entspricht:

$$L = \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{jk} M(y_j) M(y_k). \tag{3.23}$$

Die Schwellwertfunktion M besitzt außer an der Stelle $y_j = 0$ die Ableitung $M'(y_j) = 0$.

□

Die Verwendung des Cohen-Grossberg-Theorems in der Analyse rekurrenter Netze ist dann sinnvoll, wenn die Funktionalität des zu untersuchenden Netzes der eines Assoziativspeichers entspricht oder entsprechen soll. Der Phasenraum eines Assoziativspeichers enthält dabei im allgemeinen mehrere Fixpunkt-Attraktoren, von denen einer nach Ablauf einer gewissen Zeit erreicht wird in Abhängigkeit vom Zustand des Systems zum Zeitpunkt $t = 0$ (dieser Zustand wird oft durch Eingabevektoren bestimmt, wie z.B. in Bidirektionalen Speichern (siehe Gleichungen 2.4 und 2.3)).

Es sei jedoch noch einmal darauf hingewiesen, daß es sich beim Cohen-Grossberg-Theorem nicht um eine notwendige, sondern nur um eine hinreichende Bedingung für Stabilität handelt. Es können also rekurrente Netze existieren, die die Cohen-Grossberg-Bedingungen nicht erfüllen, aber dennoch unabhängig von der Wahl der Parameter in einen Fixpunkt konvergieren, also stabil sind.

3.3.2 Neuronale Netze als Iterierte Funktionensysteme

Iterierte Funktionensysteme

Sei (X, d) ein vollständiger metrischer Raum. Hier sei dieser Raum eine beschränkte Teilmenge von $\mathbb{R}^d$, die nach [3] immer ein vollständiger metrischer Raum ist. Sei $\mathcal{H}(X)$ der Raum der kompakten Teilmengen von X ohne die leere Menge.

Nach [3] sind alle beschränkten und abgeschlossenen Teilmengen von $\mathbb{R}^d$ genau die kompakten Teilmengen von $\mathbb{R}^d$. So sind z.B. alle abgeschlossenen Intervalle aus $\mathbb{R}^d$ kompakt, aber auch die Vereinigung von endlich vielen solcher Intervalle.

Mit $d(x, B)$ sei der Abstand vom Punkt x zur Menge B gemeint, definiert wie folgt:

$$d(x, B) = min\{d(x, y) : y \in B\} \quad x \in X, B \in \mathcal{H}(X)$$

und $d_h(A, B)$ sei der Abstand der Menge A zur Menge B mit

$$d_h(A, B) = max\{d(x, B) : x \in A\} \quad A, B \in \mathcal{H}(X).$$

Dann heißt d_h die Hausdorff-Metrik und $(\mathcal{H}(X), d_h)$ ist ein vollständiger metrischer Raum (siehe [1]).

Definition 11
Ein Iteriertes Funktionensystem (IFS) auf X ist eine endliche Menge von stetigen Abbildungen $\{f_1, \ldots, f_n\}$ mit $f_i : X \to X$.

Ein IFS stellt damit auch eine Transformation F in $\mathcal{H}(X)$ dar:

$$F(B) = f_1(B) \cup \ldots \cup f_n(B) \quad \text{für alle} \quad B \in \mathcal{H}(X)$$

Theorem 7 (nach [1])
Wenn alle Funktionen f_i eines IFS auf X Kontraktionen sind (das IFS wird dann auch hyperbolisch *genannt), dann ist F eine Kontraktion auf $\mathcal{H}(X)$.*

Damit besitzt F einen eindeutigen Fixpunkt A_F, der Attraktor des IFS genannt wird (und einer kompakten Menge entspricht).

In der Metrik (X, d) kann ein solcher Fixpunkt–Attraktor aus der Hausdorff–Metrik jedoch einem periodischen oder chaotischen Attraktor entsprechen, da alle kompakten Teilmengen von X Punkte in der Hausdorff–Metrik sind.

Zur näherungsweisen Erzeugung des Attraktors eines IFS kann folgender Algorithmus durch einen determinierten Automaten ausgeführt werden:

Algorithmus 1 (Deterministische Erzeugung des Attraktors)
Sei $\{X; f_1, \ldots, f_m\}$ ein hyperbolisches IFS und $A_0 \subset X$ eine beliebige kompakte Menge. Dann berechne $A_n = F^n(A)$ mittels

$$A_n = \cup_{j=1}^m f_j(A_{n-1}) \quad \textit{für} \quad n = 1, 2, \ldots$$

Die so konstruierte Sequenz $\{A_n : n = 0, 1, 2, \ldots\} \subset \mathcal{H}(X)$ konvergiert wegen Theorem 7 zu dem Attraktor des IFS in der Hausdorff-Metrik.

Wird das IFS auf einem endlichen Automaten berechnet, dann muß mit einer endlichen Teilmenge $A_0 \subset X$ begonnen werden. Da ein IFS aus endlich vielen Funktionen f_i besteht, erzeugt eine endliche Menge A_i mittels F auch wieder eine endliche Menge $A_{i+1} = F(A_i)$, die im allgemeinen jedoch größer sein kann als A_i. Die angegebene Form des Algorithmus benötigt also unter Umständen in jedem Schritt zusätzliche Speicherkapazität und ist daher als Modell für die in einem rekurrenten Netz stattfindenden Berechnungen ungeeignet.

Probabilistische Iterierte Funktionensysteme

Eine ander Möglichkeit der Berechnung von Attraktoren Iterierter Funktionensysteme mit rekurrenten Netzen bietet der folgende Algorithmus.

Algorithmus 2 (Nichtdeterministische Erzeugung des Attraktors)
Sei $\{X; f_1, \ldots, f_m\}$ ein hyperbolisches IFS, wobei jeder Abbildung f_i eine Wahrscheinlichkeit $p_i > 0$ mit $\sum_{i=1}^m p_i = 1$ zugeordnet ist und $x_0 \in X$. Dann wähle

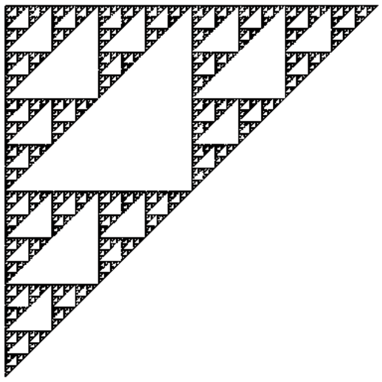
$$x_n \in \{w_1(x_{n-1}), w_2(x_{n-2}), \ldots, w_N(x_{n-1})\} \quad \textit{für} \quad n = 1, 2, \ldots,$$

wobei p_i die Wahrscheinlichkeit für das Ereignis $x_n = w_i(x_{n-1})$ ist. Werden die Abbildungen zur Generierung der Sequenz $\{x_n : n = 0, 1, 2, \ldots\} \subset X$ unabhängig ausgewählt, dann konvergiert $\{x_n\}_{n=0}^\infty$ zu dem Attraktor des probabilistischen IFS.

Beispiel 4 *Man betrachte das IFS:*

$$\begin{aligned}
\boldsymbol{f}_1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\boldsymbol{f}_2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \end{bmatrix} \\
\boldsymbol{f}_3 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}
\end{aligned}$$

Das Bild des Sierpinski-Dreiecks, generiert durch den nichtdeterministischen Algorithmus mit dem angegebenen IFS auf der Menge $\mathbb{R}^2$ *und den Wahrscheinlichkeiten* $p_i = \frac{1}{3}$ *für* $i = 1, 2, 3$*, sieht wie folgt aus:*



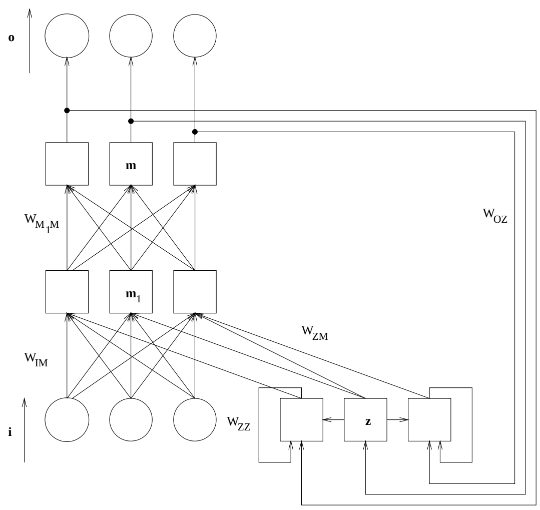
□

Ein gemeinsamer Formalismus für RNN und IFS

Neben den Anwendungen rekurrenter Neuronaler Netze als Assoziativspeicher wie im Abschnitt 3.3.1 beschrieben, existieren auch einige Architekturen, die zur Generierung von Sequenzen geeignet sind. Das heißt diese Netze können periodische oder chaotische Attraktoren besitzen.

Beispiel 5 (Rekurrentes Netz zur Generierung von Sequenzen)

Der Parametervektor $\boldsymbol{i}$*, auch Plan genannt, bleibt während der Dauer der Berechnung konstant. Sind die Gewichte im Netz so gewählt, daß im Phasenraum unterschiedliche periodische Attraktoren existieren, dann kann über diesen Plan einer der Attraktoren selektiert werden. Das Netz wird oft auch als* Jordan*–Netz (siehe [6]) bezeichnet:*



□

Ein durch die Eingabe gesteuertes dynamisches System bestehe aus einem Eingabevektor ($\boldsymbol{i}^t$) und einem Ausgabevektor ($\boldsymbol{o}^t$) sowie einem Zustandsvektor ($\boldsymbol{z}^t$). Die Funktion $\boldsymbol{F}_M(\boldsymbol{z}^{(t-1)}, \boldsymbol{i}^t)$ berechne aus der aktuellen Eingabe und dem alten Zustand einen Vektor $\boldsymbol{m}^t$, aus dem mit der Funktion $\boldsymbol{f}_O(\boldsymbol{m}^t)$ die Ausgabe des Systems und mit $\boldsymbol{f}_Z(\boldsymbol{m}^t)$ der Zustand zum Zeitpunkt t gebildet werden (siehe Abbildung 3.10). Dabei existieren nur endlich viele mögliche Eingabevektoren $\boldsymbol{i}$.

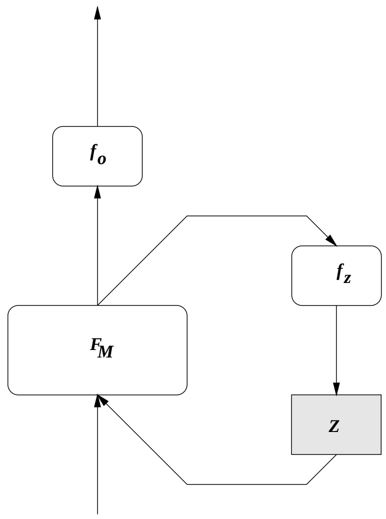


Abbildung 3.10: Gemeinsame Beschreibung von IFS und RNN

Im Sinne eines Iterierten Funktionensystems wird der Berechnungsvorgang innerhalb der Systemarchitektur wie folgt interpretiert:

1. Der Zustandsvektor $\boldsymbol{z}^t$ gibt den Zustand des Systems im Phasenraum zum Zeitpunkt t wieder.

2. Durch den Eingabevektor $\boldsymbol{i}_k$ zum Zeitpunkt t mit $1 \geq k \geq n$ wird die Funktion $\boldsymbol{f}_k$ der n Funktionen des iterierten Funktionensystems $\boldsymbol{F}_M$ selektiert und auf den Vektor $\boldsymbol{z}^t$ angewendet.

3. Die Funktionen $\boldsymbol{f}_O$ und $\boldsymbol{f}_Z$ transformieren $\boldsymbol{f}_k(\boldsymbol{z}^t)$ dann in neue Ausgabe- und Zustandsvektoren $\boldsymbol{o}^t$ und $\boldsymbol{z}^t$.

Im Gegensatz zur Darstellung endlicher Automaten kann der Zustandsvektor hier im allgemeinen unendlich viele Werte annehmen.

Beispiel 6 (Überführung des Jordan–Netzes in den obigen Formalismus)

$$\begin{aligned}
\boldsymbol{F}_M(\boldsymbol{z}, \boldsymbol{i}) &= \boldsymbol{S}_{\boldsymbol{m}}(W_{M_1M} \cdot \boldsymbol{S}_{\boldsymbol{m}_1}(W_{ZM} \cdot \boldsymbol{z} + W_{IM} \cdot \boldsymbol{i})) \\
\boldsymbol{f}_O(\boldsymbol{m}) &= \boldsymbol{F}_M(\boldsymbol{z}, \boldsymbol{i}) \\
\boldsymbol{f}_Z(\boldsymbol{m}) &= \boldsymbol{S}_{\boldsymbol{z}}(W_{OZ} \cdot \boldsymbol{F}_M(\boldsymbol{z}, \boldsymbol{i}) + W_{ZZ} \cdot \boldsymbol{z})
\end{aligned}$$

Sei $\boldsymbol{S} = \cup_{i=1}^{n} s_i(x_i)$*, wobei* s_i *eine Schwellwertfunktion ist.*

Da der Plan während der gesamten Berechnung unverändert bleibt, wird also in jedem Zeitschritt die gleiche Funktion angewendet. Wenn diese durch den Eingabevektor selektierte Funktion keine Kontraktion ist, d.h. das System besitzt z.B. einen periodischen Attraktor, dann ist die Bedingung des Kontraktionstheorems 7 bereits unerfüllt und die Existenz eines eindeutigen Attraktors nicht gesichert. □

Beispiel 7 (Das affine IFS zur Generierung des Sierpinski-Dreiecks)

$$\begin{aligned} \boldsymbol{F}_M(\boldsymbol{z}, \boldsymbol{i}) &= \boldsymbol{f}_{\boldsymbol{i}}(\boldsymbol{z}) \quad \mathit{mit} \quad \boldsymbol{i} = i \\ \boldsymbol{f}_O(\boldsymbol{m}) &= \boldsymbol{m} \\ \boldsymbol{f}_Z(\boldsymbol{m}) &= \boldsymbol{m} \end{aligned}$$

Die Abbildung dieses IFS auf den allgemeinen Formalismus ist im Gegensatz zum Jordan-Netz sehr einfach und da es sich bei den Funktionen um Kontraktionen handelt, führt auch eine zufällige Eingabesequenz zu dem einzigen existierenden Attraktor. □

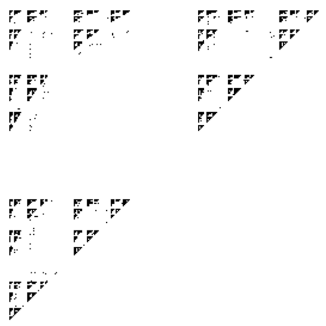
Beispiel 8

Ein dem Sierpinkski-Dreieck ähnlicher Attraktor in $\mathbb{R}^2$ *kann auch durch ein einfaches RNN mit der Sigmoide als Schwellwertfunktion berechnet werden, in das zu jedem Zeitpunkt* i_1, i_2 *oder* i_3 *zufällig und ergodisch eingegeben werden:*

$$i_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \quad i_2 = \begin{bmatrix} 0 \\ 0.5 \end{bmatrix}; \quad i_3 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}; \quad \boldsymbol{f}_O(\boldsymbol{m}) = \boldsymbol{m}; \quad \boldsymbol{f}_Z(\boldsymbol{m}) = \boldsymbol{m}$$

$$\boldsymbol{F}_M(\boldsymbol{z}, \boldsymbol{i}) = \boldsymbol{S}\left(\begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \cdot \boldsymbol{z} + \boldsymbol{i}\right) \quad \mathit{mit} \quad S = \begin{bmatrix} (1+e^{-4x_1+2})^{-1} \\ (1+e^{-4x_2+2})^{-1} \end{bmatrix}$$

Der resultierende Attraktor ist gegenüber dem linearen IFS durch die Sigmoide verzerrt:



□

Eine Analyse eines rekurrenten Neuronalen Netzes mittels der Theorie iterierter Funktionensysteme ist insbesondere dann sinnvoll, wenn sich das rekurrente Netz so in den allgemeinen Formalismus überführen läßt, daß mit den endlich vielen Eingabevektoren Funktionen eines IFS identifizierbar sind.

Sind diese Funktionen Kontraktionen, dann kann man einer Eingabesequenz einen stabilen Attraktor zuordnen. Werden die Eingabevektoren zufällig und ergodisch dem Netz zugeführt, dann existiert ein eindeutiger Attraktor.

Zusammenfassung

In der Arbeit wurde gezeigt, daß die etablierte Theorie dynamischer Systeme auf Neuronale Netze prinzipiell anwendbar ist. Die dabei verwendeten Methoden können in zwei Klassen unterteilt werden:

- lokale Stabilitätsanalyse
  - lineare Stabilitätsanalyse
- globale Stabilitätsanalyse
  - Suche nach einer Ljapunov-Funktion
  - Analyse als Iteriertes Funktionensystem

Lokale Verfahren liefern Eigenschaften eines Systems in der Umgebung eines bestimmten Phasenraumpunktes. Globale Verfahren hingegen geben Aufschluß über Eigenschaften von Trajektorien im Phasenraum.

Die Anwendung lokaler Verfahren ist bereits bei kleinen Netzen schwierig und eine vollständige Klassifizierung des Parameterraumes in Regionen, denen stabile Attraktoren des Systems zugeordnet werden können, nur im Einzelfall möglich (z.B. bei einelementigen Netzen).

Globale Verfahren wiederum liefern nur unter der Einhaltung strenger Bedingungen (wie Symmetrie der Gewichtsmatrix beim Cohen-Grossberg-Theorem) Aussagen über dynamische Eigenschaften des Systems.

Für rekurrente Netze, die durch Eingabevektoren gesteuert werden, d.h. Systeme, die z.B. zur Steuerung und Regelung verwendbar sind, ist neben den in dieser Arbeit nicht untersuchten Automatenmodellen nur die Darstellung als Iteriertes Funktionensystem geeignet.

Neben fortführenden Untersuchungen von kleinen Netzen, aus denen möglicherweise für größere Netze verallgemeinerbare Aussagen gewonnen werden können, verspricht die Theorie Iterierter Funktionensysteme weitere Einsichten in das Verhalten dynamischer Systeme und Neuronaler Netze. Basis könnte dabei die erfolgreiche Untersuchung eines existierenden Regelungsnetzes als IFS sein.

Hinreichende Bedingung für die Existenz eindeutiger Attraktoren eines IFS ist, daß alle Abbildungen des IFS Kontraktionen sind. Experimente und Analysen mit IFS, in denen diese Bedingung verletzt ist, könnten wiederum Rückschlüsse auf das Verhalten rekurrenter Netze zulassen.

Literaturverzeichnis

[1] M. Barnsley. *Fractals Everywhere.* San Diego, CA: Academic Press, 1988.

[2] M.A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions SMC-13*, pages 815–826, 1983.

[3] E.T. Copson. *Metric Spaces.* Cambridge University Press, 1968.

[4] S. Grossberg. Content-addressable memory storage by neural networks: A general model and global ljapunov method. *Computational Neurosciences*, 1990.

[5] K. Hornik. Approximation capabilities of multilayered feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[6] M.I. Jordan. Serial order: A parallel, distributed processing approach. Technical report, University of California, San Diego, Institute for Cognitive Science., 1986.

[7] J. F. Kolen. *Exploring the Computational Capabilities of Recurrent Neural Networks.* PhD thesis, Ohio State University, 1994.

[8] B. Kosko. *Neural Networks and Fuzzy Systems.*

[9] R.W. Leven, P. Koch, and Pompe B. *Chaos in dissipativen Systemen.* Berlin: Akademie-Verlag, 1994.

[10] P. Tino. Fixed-points in two-neuron discrete time recurrent neural networks. Technical report, University of Maryland, 1995.

[11] H. Troger and A. Steindl. *Nonlinear Stability and Bifurcation Theory.* Springer, 1991.

[12] X. Wang. Period-doublings to chaos in a simple neural network: An analytical proof. *Complex Systems*, 5, 1991.