
Kapitel 2: Netzwerke bei niedriger Speicherdichte

Der Zustandsraum Autoassoziativer Neuronaler Netzwerke bei niedriger Speicherdichte α ist sehr regelmäßig strukturiert. Unter niedriger Speicherdichte wollen wir dabei den Grenzfall verstehen, daß die Anzahl der eingespeicherten Muster, p , fest bleibt, während die Anzahl der Neuronen, N , gegen unendlich geht. Für die Speicherdichte $\alpha = \frac{p}{N}$ gilt also $\alpha \rightarrow 0$.

In diesem Limes stellt sich die Dynamik des autoassoziativen Netzwerks als trivial heraus: Eine einfache Signal/Rauschanalyse (vergl. Abschnitt 3.1.1) zeigt, daß die Muster ein volles Attraktionsgebiet besitzen. Dies gilt allerdings nur, falls im verrauschten Startmuster keine Beimischungen von anderen gelernten Mustern zu finden sind.

Die Situation wird komplizierter, wenn man im Systemzustand des Neuronalen Netzes Anteile mehrerer Muster zuläßt. In diesem Fall wird eine äußerst komplexe Struktur des Zustandsraumes sichtbar, die sich einer vollständigen Analyse bislang entzogen hat. Neben den gelernten Mustern treten dann die schon in der Einleitung erwähnten, metastabilen Niveaus als zusätzliche dynamische Attraktoren auf. Diese wirken als Fallen der Dynamik und schränken das Attraktionsgebiet der gelernten Muster ein.

2.1 Dynamische Kenngrößen

2.1.1 Überlapp und Untergitter-Magnetisierungen

Der Überlapp

Der Überlapp ist die natürliche Erweiterung des Begriffs der Magnetisierung aus der Spinglas-Physik. Er ist das normierte Skalarprodukt zwischen zwei Zustandsvektoren S_i^1 und S_i^2 des Neuronalen Netzwerkes:

$$m = N^{-1} \sum_i S_i^1 \cdot S_i^2 .$$

Klarerweise gilt $m \in [-1, +1]$. Prinzipiell können, für $N \rightarrow \infty$, alle Werte dieses Intervalls angenommen werden. Gilt $m = +1$, so sind beide Zustandvektoren identisch, d.h., es gilt $\bigwedge_i : S_i^1 = S_i^2$.

Der aktuelle Systemzustand eines Neuronalen Netzwerkes wird üblicherweise durch die Angabe der Musterüberlappungen

$$m_\mu(t) = N^{-1} \sum_i \xi_i^\mu \cdot S_i(t)$$

beschrieben. Dabei tritt allerdings eine Schwierigkeit auf: Die Überlappungen zu verschiedenen Mustern sind nicht unabhängig voneinander. Zwar kann jeder einzelne Musterüberlapp an sich alle erlaubten Werte zwischen $+1$ und -1 annehmen, doch ist bei weitem nicht jede mögliche Kombination der p Musterüberlappungen erlaubt. Ist beispielsweise der Überlapp zum ersten Muster, $m_1 = 1$, so kann keiner der anderen Musterüberlappungen gleich eins sein — sonst wäre ja dieses andere Muster identisch dem ersten Muster.

Die erlaubten Bereiche im Überlapp-Raum liegen also nicht innerhalb eines p -dimensionalen Hyperkubus, wie manche Autoren fälschlicherweise angenommen haben ([Co88]), sondern besitzen komplexere Grenzflächen. Um diese zu bestimmen, ist es günstig, dynamische Kenngrößen einzuführen, deren erlaubte Bereiche durch einfache Grenzflächen festgelegt sind.

Die Untergitter-Magnetisierungen

Ein dazu geeignetes Konzept stellen die Untergitter (UG) dar [vH86]. Sie sind im wesentlichen eine Einteilung der Neuronen in voneinander disjunkte Klassen. Für die einzelnen Klassen kann man Magnetisierungen definieren, die wegen der Disjunktheit der Klassen unabhängige dynamische Parameter werden.

Zur Definition der Untergitter nutzt man die verschiedenen Vorzeichenkombinationen der Muster an verschiedenen Neuronen aus. Alle Neuronen, die die gleichen Vorzeichenkombinationen besitzen, werden in eine Untergitter-Klasse eingeordnet.

Betrachten wir beispielsweise den Fall zweier Muster ξ_i^1 und ξ_i^2 . Die Vorzeichen von ξ_i^1 und ξ_i^2 können am Ort i übereinstimmen oder verschieden sein. Es findet also — bei zwei Mustern — eine Einteilung der Neuronen in zwei verschiedene Klassen statt. Wählen wir der Einfachheit halber $\xi_i^1 = +1$, was wegen einer Eichtransformation der Dynamik Neuronaler Netzwerke immer geht (siehe Anhang C), so hat das eine Untergitter die Vorzeichenkombinationen $(++)$, das andere die Kombination $(+-)$.

Die Erweiterung dieses Untergitter-Konzepts auf mehr als zwei Muster stellt kein Problem dar. Bei p Mustern sind $2^{(p-1)}$ verschiedene Vorzeichenkombinationen der Muster möglich, die in folgender Weise angeordnet werden können:

<i>Klasse</i>	\rightarrow	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	\dots	<i>u</i>	<i>v</i>	<i>w</i>	<i>x</i>	<i>y</i>	<i>z</i>
<i>Muster</i>	1	+	+	+	+	+	+		+	+	+	+	+	+
<i>Muster</i>	2	+	+	+	+	+	+		-	-	-	-	-	-
	\vdots													
<i>Muster</i>	$p-2$	+	+	+	+	-	-		+	+	-	-	-	-
<i>Muster</i>	$p-1$	+	+	-	-	+	+	\dots	-	-	+	+	-	-
<i>Muster</i>	p	+	-	+	-	+	-		+	-	+	-	+	-

Interpretiert man die Spalten dieser Tabelle als Binärzahlen, mit "0" für "+" und "1" für "-", so treten, von links nach rechts, alle Binärzahlen von 0 bis $2^{(p-1)} - 1$ auf.

Damit sind $2^{(p-1)}$ verschiedene Untergitter definiert, die wir im folgenden durch lateinische Indices kenntlich machen. Die Matrix der Vorzeichenkombinationen,

$$\xi_a^\mu = \begin{pmatrix} +1 & +1 & \dots & +1 \\ +1 & +1 & \dots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ +1 & -1 & \dots & -1 \end{pmatrix},$$

bezeichnen wir als Untergitter-Matrix. Die Zeilen der UG-Matrix sind mit Zeilen der Hadamard-Matrix der Dimension $2^{(p-1)}$ [Go86] identisch und damit orthogonal untereinander:

$$\sum_a \xi_a^\mu \xi_a^\nu = 2^{(p-1)} \delta_{\mu\nu}. \quad (2.1)$$

Die Zeilen der Untergitter-Matrix spannen einen p -dimensionalen Unterraum des $D = 2^{(p-1)}$ dimensionalen UG-Raumes auf.

In endlichen Neuronalen Netzen müssen nicht alle UG-Klassen realisiert sein, aber im Falle unkorrelierter Zufallsmuster und im hier betrachteten Limes mit $N \rightarrow \infty$, aber p fest, treten alle Untergitter auf.

Die Indices $i \in [1, N]$ der Neuronen, die zur UG-Klasse a gehören, fassen wir in der Indexmenge I_a zusammen:

$$I_a = \left\{ i \mid \forall \mu : \xi_i^\mu = \xi_a^\mu \right\}.$$

Damit können wir eine Abbildung der Neuronenindices auf die Untergitterindices definieren:

$$ug(ij \dots m) = \left\{ ab \dots e \mid i \in I_a, j \in I_b, \dots, m \in I_e \right\}.$$

Die Mächtigkeit der Indexmengen I_a bezeichnen wir mit N_a . Für unkorrelierte Zufallsmuster sind die Untergitter-Mächtigkeiten alle gleich und es gilt

$$N_a \approx \frac{N}{2^{(p-1)}}.$$

Die Einteilung der Neuronen in disjunkte Untergitter-Mengen ermöglicht es nun, voneinander unabhängige dynamische Kenngrößen zu definieren. Es sind die zu den verschiedenen Untergittern gehörigen Magnetisierungen

$$m_a = \frac{1}{N_a} \sum_{i \in I_a} S_i.$$

Klarerweise gilt auch hier $m_a \in [-1, +1]$. Da aber im Gegensatz zu den Musterüberlappungen die Untergitter-Magnetisierungen verschiedener Untergitter unabhängig voneinander sind, stellt nun *jede* Position im $2^{(p-1)}$ -dimensionalen Hyperkubus der Untergitter-Magnetisierungen einen erlaubten Systemzustand dar. Umgekehrt hat jeder mögliche Systemzustand eindeutige Koordinaten innerhalb des $2^{(p-1)}$ -dimensionalen Untergitter-Raumes. Wir können also durch die Rückprojektion des Untergitter-Hyperkubus in den Überlapp-Raum die erlaubten Bereiche für die Musterüberlappungen finden.

2.2 Der Überlapp-Raum zu mehreren Mustern

2.2.1 Transformation zwischen Überlapp- und Untergitter-Raum

Beschränken wir uns zunächst auf unkorrelierte Zufallsmuster. Dann sind die Untergitter-Klassen alle gleichmächtig und es gilt

$$\begin{aligned} m_\mu &= \frac{1}{N} \sum_i \xi_i^\mu \cdot S_i = \frac{1}{N} \sum_i \xi_{ug(i)}^\mu \cdot S_i \\ &= \frac{1}{N} \left\{ \xi_a^\mu \sum_{i \in I_a} S_i + \dots + \xi_z^\mu \sum_{i \in I_z} S_i \right\} \\ &= \frac{1}{2^{(p-1)}} \left\{ \xi_a^\mu m_a + \dots + \xi_z^\mu m_z \right\} \end{aligned}$$

oder, in Matrixschreibweise,

$$\begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{pmatrix} = \frac{1}{2^{(p-1)}} \begin{pmatrix} \xi_a^1 & \xi_b^1 & \cdots & \xi_z^1 \\ \xi_a^2 & \xi_b^2 & \cdots & \xi_z^2 \\ \vdots & \vdots & \ddots & \vdots \\ \xi_a^p & \xi_b^p & \cdots & \xi_z^p \end{pmatrix} \cdot \begin{pmatrix} m_a \\ m_b \\ \vdots \\ m_z \end{pmatrix}.$$

Dies stellt eine lineare Projektion des $2^{(p-1)}$ -dimensionalen Untergitter-Raums in den p -dimensionalen Überlapp-Raum dar. Wegen der unterschiedlichen Dimensionen der beiden Räume gibt es i.a. *keine* inverse Transformation vom Überlapp-Raum in den Untergitter-Raum. Die lineare Projektion führt dazu, daß die Topologie des Untergitter-Hyperkubus erhalten bleibt; der konvexe Hyperkubus des Untergitter-Raums wird auf eine konvexe Menge des Überlapp-Raums abgebildet.

2.2.2 Erlaubter Überlapp-Raum bei zwei und drei Mustern

Zur Erläuterung der vorgestellten Konzepte soll nun die Gestalt des Überlapp-Raumes bei zwei und drei Mustern diskutiert werden. Dies bietet den Vorteil, die dadurch entstehenden Grenzflächen des Überlapp-Raumes noch grafisch dargestellt werden können. Zudem treten im Falle korrelierter Muster zusätzliche Schwierigkeiten für $p > 3$ auf (siehe Abschnitt 2.2.3).

Zweidimensionaler Überlapp-Raum

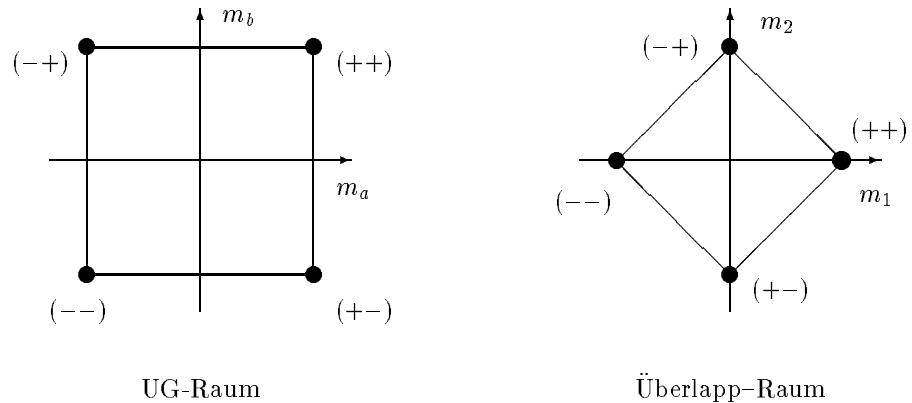


Abbildung 2.1: Aus der Transformation des UG-Quadrates erhält man die erlaubten Bereiche für die Musterüberlappungen. Die UG-Vektoren mit voller UG-Magnetisierung, $(++)$, $(+-)$, $(-+)$ und $(--)$, werden dabei auf die Muster und ihre Inversen abgebildet.

Bei zwei Mustern hat die Untergitter-Matrix die einfache Gestalt

$$\xi_a^\mu = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}.$$

Für unkorrelierte Zufallsmuster erhalten wir also die Musterüberlappungen durch

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix} \begin{pmatrix} m_a \\ m_b \end{pmatrix}. \quad (2.2)$$

Diese lineare Transformation stellt im wesentlichen eine Drehung des Untergitter-Quadrats um 45° im Uhrzeigersinn dar (vergl. Abbildung 2.1). Überlappungen, die außerhalb des gedrehten Quadrats liegen, sind nicht realisierbar.

Dreidimensionaler Überlapp-Raum

Um den erlaubten Überlapp-Raum für drei Muster zu ermitteln, muß man die vier Untergitter-Klassen

	1	2	3
a	+	+	+
b	+	+	-
c	+	-	+
d	+	-	-

berücksichtigen. Der zulässige Überlapp-Raum ergibt sich durch die Projektion des vierdimensionalen Hyperkubus der Untergitter-Magnetisierungen in den dreidimensionalen Überlapp-Raum. Konkret stellt sich diese Projektion — wieder im Falle unkorrelierter Zufallsmuster — durch

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \end{pmatrix} \cdot \begin{pmatrix} m_a \\ m_b \\ m_c \\ m_d \end{pmatrix}$$

dar. Es ergibt sich als erlaubter Überlappbereich ein rhombisches Dodekaeder mit 14 Ecken (Abbildung 2.2). Davon korrespondieren 6 Ecken zu den drei eingespeicherten Mustern bzw. deren invertierten Bitmustern. Die übrigen 8 Ecken sind Mischungen aus allen drei Mustern.

Betrachten wir nun die für Punktattraktoren maßgebliche Lypunov-Funktion $L_a[\{S_i(t)\}]$ (Gleichung (1.2) auf Seite 9). Im hier behandelten Limes, $\alpha \rightarrow 0$, stellt sich diese Funktion für alle drei bekannten Kopplungsmatrizen als quadratische Funktion der Musterüberlappungen dar:

$$\begin{aligned} L_a[\{S_i(t)\}] &= -\sum_{i,j} S_i J_{ij} S_j \\ &= -N^{-1} \sum_{i,j,\mu} S_i \xi_i^\mu \xi_j^\mu S_j \\ &= -N \sum_{\mu} (m_\mu)^2 \end{aligned}$$

Die Äquipotentialflächen der Lyapunov-Funktion $L_a[\{S_i(t)\}]$ sind also Kugeln um den Ursprung des Koordinatensystems, wobei die Funktion nach außen hin abnimmt. Damit sind aber alle Ecken des Dodekaeders Attraktoren der Dynamik, also auch die 8 Ecken, die nicht zu den Mustern korrespondieren! Dies ist das erste Auftreten dynamischer Attraktoren, die nicht explizit im Neuronalen Netz eingespeichert sind.

Insgesamt ergibt sich folgendes Bild: Neben den Bitkombinationen der zu den Mustern korrespondierenden Untergitter-Vektoren,

$$\left. \begin{array}{l} \mathbf{Muster\ 1} : + + + + \\ \mathbf{Muster\ 2} : + + - - \\ \mathbf{Muster\ 3} : + - + - \end{array} \right\} \text{Muster}$$

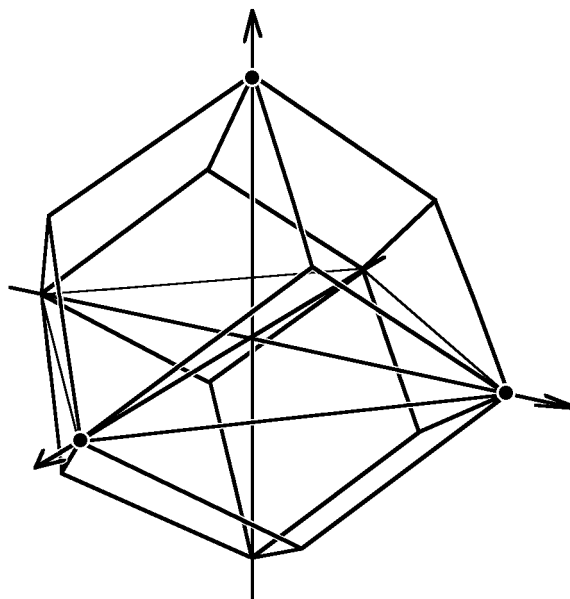


Abbildung 2.2: Der erlaubte Überlapp-Raum in drei Dimensionen ist ein rhombisches Dodekaeder. Die meisten der Untergitter-Vektoren mit voller Magnetisierung liegen an den Ecken des Dodekaeders. Lediglich die Untergitter-Vektoren $(+ - - +)$ und $(- + + -)$ liegen im Zentrum des Überlapp-Raums, dem Maximum der Lyapunov-Funktion $L_a[\{S_i(t)\}]$.

sind noch die Bitkombinationen mit Beimischungen aller drei Muster

$$\left. \begin{array}{l} \text{Meta 1} \quad : \quad + + + - \\ \text{Meta 2} \quad : \quad + + - + \\ \text{Meta 3} \quad : \quad + - + + \\ \text{Meta 4} \quad : \quad - + + + \end{array} \right\} \begin{array}{l} \text{3er-Mischzustände,} \\ \text{definiert durch } S_i = \text{sign}(\pm \xi_i^1 \pm \xi_i^2 \pm \xi_i^3) \end{array}$$

und die invertierten Bitkombinationen dynamisch stabile Konfigurationen. Keine Attraktoren sind die Ecken $(+ - - +)$ und $(- + + -)$ des UG-Hyperkubus, denn sie werden auf den Nullpunkt des Überlapp-Raumes, dem Maximum der Lyapunov-Funktion, abgebildet.

Die Koordinaten der 3er-Mischzustände im Überlapp-Raum erhalten wir aus der Transformation der Untergitter-Koordinaten. So folgt beispielsweise für die Koordinaten des 3er-Mischzustandes $(+ + + -)$,

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}_{(+ + + -)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Die Koordinaten der anderen 3er-Mischzustände folgen analog. Es sind gleichmäßige Mischungen aus allen drei Mustern.

2.2.3 Der Überlapp-Raum bei korrelierten Mustern

Auch bei korrelierten Mustern ergibt sich der zulässige Bereich der Muster-Überlappungen aus der Projektion des Untergitter-Hyperkubus in den Überlapp-Raum. Nur müssen jetzt die unterschiedlichen Mächtigkeiten der Untergitter

berücksichtigt werden. Man erhält

$$\begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_p \end{pmatrix} = N^{-1} \begin{pmatrix} N_a \xi_a^1 & N_b \xi_b^1 & \cdots & N_z \xi_z^1 \\ N_a \xi_a^2 & N_b \xi_b^2 & \cdots & N_z \xi_z^2 \\ \vdots & \vdots & \ddots & \vdots \\ N_a \xi_a^p & N_b \xi_b^p & \cdots & N_z \xi_z^p \end{pmatrix} \cdot \begin{pmatrix} m_a \\ m_b \\ \vdots \\ m_z \end{pmatrix}.$$

Die Situation wird allerdings für mehr als drei Muster kompliziert: aus der Korrelationsmatrix

$$C_{\mu\nu} = N^{-1} \sum_i \xi_i^\mu \xi_i^\nu$$

folgen nämlich nur $\frac{p(p-1)}{2}$ Bedingungen für die $2^{(p-1)}$ Mächtigkeiten der verschiedenen UG-Klassen. Eine zusätzliche Bedingung liefert noch $\sum_a N_a = N$. Damit werden die Untergitter-Mächtigkeiten aber nur im Falle von $p = 2$ oder 3 Mustern durch deren Korrelationen festgelegt. Schon bei vier Mustern fehlt eine Bedingung, um die Untergitter-Mächtigkeiten zu bestimmen. Wir beschränken uns deshalb auch in diesem Abschnitt auf Beispiele mit zwei oder drei Mustern.

Zwei Muster

Mit $c_{12} = N^{-1} \sum_i \xi_i^1 \xi_i^2$ folgt für die Untergitter-Mächtigkeiten

$$\begin{aligned} N_a &= \frac{N}{2}(1 + c_{12}) \\ N_b &= \frac{N}{2}(1 - c_{12}). \end{aligned}$$

Die Transformation (2.2) (Seite 17) geht damit über in

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = N^{-1} \begin{pmatrix} +N_a & +N_b \\ +N_a & -N_b \end{pmatrix} \cdot \begin{pmatrix} m_a \\ m_b \end{pmatrix}$$

und das gedrehte Überlapp-Quadrat entartet zu einer Raute (Abbildung 2.3, links).

Drei Muster

Bei drei Mustern ist die allgemeine Transformation aus dem Untergitter-Raum in den Überlapp-Raum durch

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = N^{-1} \begin{pmatrix} +N_a & +N_b & +N_c & +N_d \\ +N_a & +N_b & -N_c & -N_d \\ +N_a & -N_b & +N_c & -N_d \end{pmatrix} \cdot \begin{pmatrix} m_a \\ m_b \\ m_c \\ m_d \end{pmatrix}$$

gegeben.

Die UG-Mächtigkeiten können hierbei noch aus den Korrelationen der drei Muster bestimmt werden. Wegen der Definition der Untergitter folgt für die Korrelationen zwischen den Mustern zunächst

$$\begin{pmatrix} c_{12} \\ c_{13} \\ c_{23} \end{pmatrix} = N^{-1} \begin{pmatrix} +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{pmatrix} \cdot \begin{pmatrix} N_a \\ N_b \\ N_c \\ N_d \end{pmatrix}.$$

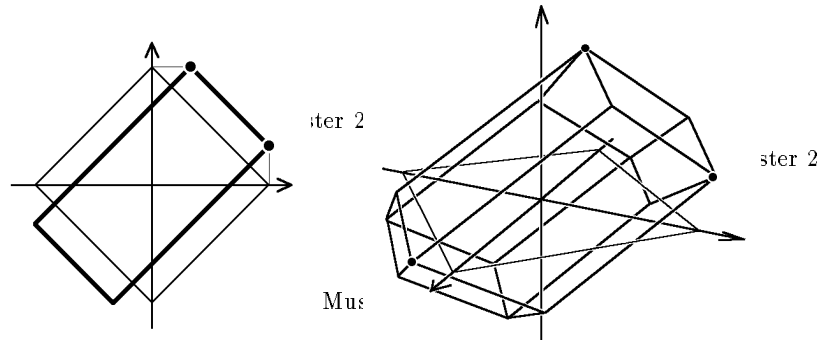


Abbildung 2.3: Der erlaubte Überlapp-Raum bei korrelierten Mustern. Bei zwei Mustern ergibt sich eine Raute (links), bei drei Mustern (rechte Abbildung) ein deformiertes Dodekaeder.

Mit der zusätzlichen Bedingung $\sum_a N_a = N$ kann diese Gleichung invertiert werden. Die UG-Mächtigkeiten folgen aus den Korrelationen durch

$$\begin{pmatrix} N_a \\ N_b \\ N_c \\ N_d \end{pmatrix} = \frac{N}{4} \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \end{pmatrix} \cdot \begin{pmatrix} c_{12} \\ c_{13} \\ c_{23} \\ 1 \end{pmatrix}.$$

Damit liegt die Transformation fest und wir sind in der Lage, bei vorgegebener Korrelationsmatrix die Koordinaten der Muster und der 3er-Mischzustände zu berechnen. Die Koordinaten der Muster ergeben sich der Einfachheit halber direkt aus der Korrelationsmatrix, beispielsweise gilt für das erste Muster:

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}_{\xi^1} = \begin{pmatrix} 1 \\ c_{12} \\ c_{13} \end{pmatrix}.$$

Analog folgen die Koordinaten der übrigen gelernten Muster.

Die Koordinaten der 3er Mischzustände erhält man aus der Transformation der entsprechenden UG-Koordinaten. So gilt beispielsweise

$$\begin{aligned} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}_{(++++)} &= N^{-1} \begin{pmatrix} +N_a & +N_b & +N_c & -N_d \\ +N_a & +N_b & -N_c & +N_d \\ +N_a & -N_b & +N_c & +N_d \end{pmatrix} \cdot \begin{pmatrix} +1 \\ +1 \\ +1 \\ -1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 + c_{12} + c_{13} - c_{23} \\ 1 + c_{12} - c_{13} + c_{23} \\ 1 - c_{12} + c_{13} + c_{23} \end{pmatrix}. \end{aligned}$$

Analog findet man die Koordinaten der anderen sieben Mischzustände. Es ergibt sich als erlaubter Überlapp-Bereich ein deformiertes Dodekaeder (Abbildung 2.3, rechts).

2.3 Untergitter–Raum und Dynamik

2.3.1 Interne Felder der Untergitter

Das dynamische Verhalten der einzelnen Neuronen wird von ihrem internen Feld h_i bestimmt. Wir betrachten nun die internen Felder in den einzelnen Untergittern. Für die Hopfield–Kopplungsmatrix kann das interne Feld des Neurons i als

$$h_i = \sum_{\mu} \xi_i^{\mu} m_{\mu} - \frac{p}{N} S_i$$

geschrieben werden. Dabei korregiert der rechte Term für die fehlende Diagonale der Kopplungsmatrix. Für den von uns betrachteten Grenzprozess, $N \rightarrow \infty$, aber p fest, verschwindet dieser Term und wir erhalten

$$h_i \rightarrow \sum_{\mu} \xi_i^{\mu} m_{\mu} .$$

Damit ist h_i als eine Funktion globaler Größen, den Musterüberlappungen m_{μ} , und von lokalen Größen, den Vorzeichenkombinationen am Platz i , ξ_i^{μ} , dargestellt. Letztere sind aber eine Funktion der Untergitter; die internen Felder innerhalb eines Untergitters sind also alle gleich. Dies gilt im hier betrachteten Limes auch für die beiden anderen Kopplungsmatrizen, pseudoinverse Kopplungsmatrix und Perceptron–Matrix, bei denen in der obigen Formel lediglich m_{μ} durch die orthogonalisierten Musterüberlappungen

$$a_{\mu} = \sum_{\nu} C_{\mu\nu}^{-1} m_{\nu}$$

zu ersetzen ist.

2.3.2 Stabilität von Mustern

Die Gleichheit der internen Felder innerhalb eines Untergitters führt zu einigen einfachen Folgerungen über die dynamische Stabilität eines Systemzustandes. Prinzipiell gilt: Ein Muster ist stabil, falls bei allen Neuronen S_i parallel zum internen Feld h_i steht. Da alle Neuronen eines Untergitters das gleiche interne Feld besitzen, müssen in einem stabilen Zustand die Untergitter notwendigerweise voll durchmagnetisiert, also $m_a = \pm 1$ sein:

$$\text{Muster stabil} \quad \iff \quad \forall_a : m_a = \pm 1$$

Damit auch wird klar, warum die oben gefundenen 3er–Mischzustände an den Ecken des Untergitter–Hyperkubus liegen. Positionen innerhalb des Untergitter–Hyperkubus haben ja $|m_a| < 1$ und können schon deshalb dynamisch nicht stabil sein. Allerdings sind auch nicht alle $2^{(2^{(p-1)})}$ möglichen Ecken des Hyperkubus stabile Muster. Die Äquipotentialflächen der Lyapunov–Funktionen sind nämlich im UG–Raum gedreht und gestaucht. Schon im Falle dreier Muster waren die UG–Vektoren $(+ - - +)$ und $(- + + -)$ keine stabilen Systemzustände, sondern lagen vielmehr im Maximum der Lyapunov–Funktion.

Die vlldurchmagnetisierten Zustände an den Ecken des Untergitter–Raums sind binäre Vektoren. Bezeichnen wir diese Ecken mit $\eta_a = \pm 1 \equiv m_a$, so läßt sich innerhalb des Untergitter–Raums ein Musterüberlapp

$$m_{\mu}^{u,g} = \frac{1}{2^{(p-1)}} \sum_a \eta_a \xi_a^{\mu}$$

definieren. Da der Untergitter-Raum $2^{(p-1)}$ -dimensional ist, nimmt der Überlapp m_μ^{ug} genau $2^{(p-1)} + 1$ verschiedene, im Intervall $[-1, +1]$ im Abstand von $\Delta = 2^{-(p-2)}$ liegende Werte an. Nun gilt aber für unkorrelierte Zufallsmuster, daß

$$m_\mu = \frac{1}{2^{(p-1)}} \sum_a m_a \xi_a^\mu = m_\mu^{ug}.$$

Damit folgt für die Musterüberlappungen von Mischzuständen die Abschätzung

$$\max_\mu m_\mu \leq 1 - \Delta = 1 - 2^{-(p-2)}.$$

Der Überlapp von Mischzuständen ist also immer kleiner als eins; der maximal mögliche Überlapp-Wert für Mischzustände hängt davon ab, wieviele Muster im Netzwerk gespeichert sind.

2.3.3 Attraktionsgebiete der Mischzustände

Mischzustände sind nur lokale Minima der Lyapunov-Funktion und es ist zunächst nicht klar, in welchem Maße dies die Dynamik der Neuronaler Netzwerke beeinflusst. Die Attraktionsgebiete der Mischzustände können klein sein, sodaß sie dynamisch keine Rolle spielen, oder sie können große Gebiete des Zustandsraumes abdecken und zu einer erheblichen Verschlechterung der Assoziationsfähigkeit des Neuronalen Netzes führen. Wir wollen dies nun am Beispiel der 3er-Mischzustände näher untersuchen.

Trennebenen

Die Dynamik des Netzwerkes kann sowohl im Überlapp-Raum als auch im Untergitterraum beschrieben werden. Entscheidend für den Zustand der Neuronen zum darauffolgenden Zeitpunkt sind die Vorzeichenkombinationen der internen Felder. Da die internen Felder der Neuronen wegen

$$h_i = \sum_\mu \xi_i^\mu m_\mu$$

bzw.

$$h_i = \sum_{\mu, a} \xi_i^\mu \xi_a^\mu m_a$$

lineare Funktionen der in den jeweiligen Räumen maßgeblichen dynamischen Variablen sind, werden Bereiche mit unterschiedlichen Vorzeichenkombinationen jeweils durch Hyperebenen voneinander getrennt.

Während die Trennebenen im Überlapp-Raum relativ einfach bestimmt werden können, ist die Lage der Ebenen im Untergitter-Raum i.a. schwieriger anzugeben. Leider ist aber gerade der Untergitter-Raum dynamisch gesehen der wichtigere Raum, da jedes Untergitter sein eigenes internes Feld besitzt und nur die Untergitter-Magnetisierungen freie dynamische Variable sind.

Die Trennebenen sind im allgemeinen nicht mit den Grenzen der Attraktionsgebiete identisch, denn die dynamische Evolution des Netzwerk-Zustandes hängt zusätzlich von der jeweils verwendeten Dynamik ab.

synchrone Dynamik

Am einfachsten zu diskutieren ist die synchrone Dynamik, bei der alle Neuronen parallel schalten. Sie stellt einen deterministischen Prozess dar. Bereits nach

dem ersten Zeitschritt sind alle Neuronen S_i parallel zu ihrem internen Feld h_i ausgerichtet, also alle Untergitter voll durchmagnetisiert. Dies bleibt auch in späteren Zeitschritten so; das Neuronale Netz bewegt sich dann nur noch auf den Ecken des Untergitter-Kubus.

Um Attraktionsgebiete zu bestimmen, muß man die Bereiche mit unterschiedlichen Vorzeichenkombinationen bestimmen. Dies kann bei mehreren Mustern ein hoffnungsloses Unterfangen sein, gelingt jedoch noch bei drei eingespeicherten Mustern. In diesem Falle haben wir vier Untergitter-Klassen und entsprechend viele interne Felder sind zu berücksichtigen. Der bereits bekannte Überlapp-Dodekaeder wird durch die vier Ebenen in mehrere Bereiche unterschiedlicher Vorzeichenkombinationen separiert. Die Trennebenen sind durch die Gleichungen

$$\begin{aligned} m_3(m_1, m_2) &= -m_1 - m_2 & (h_a = 0) \\ m_3(m_1, m_2) &= +m_1 + m_2 & (h_b = 0) \\ m_3(m_1, m_2) &= -m_1 + m_2 & (h_c = 0) \\ m_3(m_1, m_2) &= +m_1 - m_2 & (h_d = 0) \end{aligned}$$

definiert. Sie laufen alle durch den Ursprung und sind gegenüber den Koordinatenachsen um 45° geneigt.

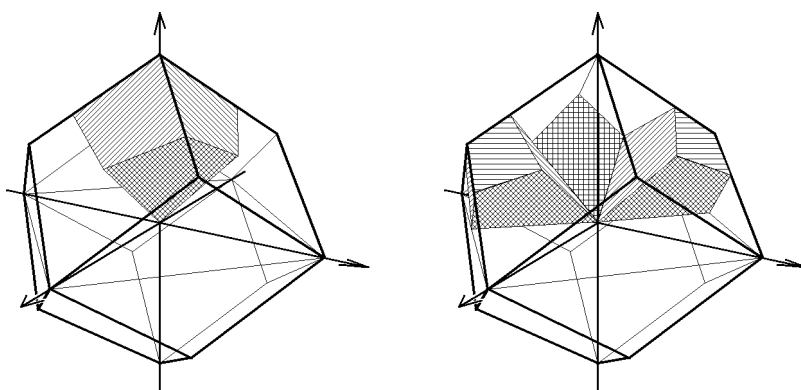


Abbildung 2.4: Die Attraktionsgebiete von Mustern (links) und 3er-Mischzuständen (rechte Abbildung) im Überlapp-Dodekaeder.

Durch die Trennebenen werden aus dem Überlapp-Dodekaeder konische Teilvolumina herausgeschnitten, innerhalb derer die internen Felder eine eindeutige Vorzeichenkombination haben. Wie sich herausstellt, sind die Vorzeichenkombinationen der internen Felder in jedem der Teilvolumina identisch mit einem der Muster, einem der 3er-Mischzustände bzw. den dazu inversen Bitmustern. Damit springt die synchrone Dynamik schon im ersten Zeitschritt auf den im Teilvolumen liegenden Attraktor. In diesem Fall sind also Trennebenen und die Grenzen der Attraktionsgebiete identisch. Die Vorzeichenkombinationen, die nicht direkt, d.h. in einem Zeitschritt, zu einem der Attraktoren führen (dies wären die Vorzeichenkombinationen $(+ - - +)$ und $(- + + -)$), existieren nicht.

Wie Abbildung 2.4 deutlich macht, haben die Muster entlang den Überlapp-Koordinatenachsen ein volles Attraktionsgebiet. Aber auch das Attraktions-

gebiet der 3er-Mischzustände (Abbildung 2.4, links) reicht bis zum Ursprung. Erhebliche Bereiche des Überlapp-Dodekaeders werden also durch die Attraktionsgebiete der Mischzustände beansprucht.

asynchrone Dynamik

Im Gegensatz zur synchronen Dynamik stellt die asynchrone Dynamik im allgemeinen, bedingt durch die zufällige Schaltreihenfolge der Neuronen, einen stochastischen Prozess dar.

Die Wahrscheinlichkeit, daß sich das Untergitter a um einen infinitesimalen Betrag dm_a ändert, ist dabei proportional der Anzahl der in diesem Untergitter dem internen Feld antiparallel stehenden Spins, also $\sim N_a(\text{sign}(h_a) - m_a)$. Damit erhalten wir bei geeigneter Skalierung der Zeit

$$\frac{dm_a}{dt} = \text{sign}(h_a) - m_a$$

und, wegen

$$dm_\mu = N^{-1} \sum_a N_a \xi_a^\mu dm_a,$$

schließlich als dynamische Gleichung für den Überlap

$$\frac{dm_\mu}{dt} = -m_\mu + N^{-1} \sum_a N_a \xi_a^\mu \text{sign}\left(\sum_\nu \xi_a^\nu m_\nu\right).$$

Diese Gleichung kann auch strenger hergeleitet werden [Co88]. Man erhält sie dann als Flußgleichung einer Fokker-Planck-Gleichung, bei welcher der dissipative Term fehlt. Die obige Gleichung beschreibt damit im Limes $N \rightarrow \infty$ die wahre Trajektorie des Systems, denn Fluktuationen finden, wegen dem fehlenden dissipativen Term in der Fokker-Planck-Gleichung, nicht statt.

Die asynchrone Dynamik wird nach obiger Gleichung durch eine exponentielle Relaxation in Richtung der Vorzeichenkombination der internen Felder gekennzeichnet. Im Überlapp-Raum ergeben sich damit als Trajektorien gerade Linien innerhalb der Bereiche gleichen Vorzeichens. Wird allerdings eine der Trennebenen durchschritten, ändert sich die Trajektorie abrupt.

In unserem Fall, bei drei gespeicherten Mustern, liegen die Trajektorien des Neuronalen Netzes so, daß beim Fluß der Dynamik keine Trennebenen durchkreuzt werden (siehe Abbildung 2.5). Die Attraktionsgebiete für synchrone und asynchrone Dynamik werden damit identisch.

Neben der traditionellen asynchronen Dynamik, in der die Reihenfolge, nach welcher die Neuronen schalten, zufällig bestimmt wird, wurden in der Literatur noch andere Dynamiken vorgeschlagen, die durch geschickte Auswahl der zu testenden Neuronen zu einer Verbesserung der Attraktionsgebiete führen sollen. Es sind dies der Gradientenabstieg [Ho87] und die sequentielle Dynamik mit bevorzugtem Testen der falschstehenden Neuronen [KA87]. Sie führen in der Tat, wie die numerisch bestimmten Attraktionsgebiete in Abbildung 2.6 zeigen, zu einer Verbesserung des Attraktionsgebietes der Muster.

Beim Gradientenabstieg bewegt sich das System immer in Richtung des stärksten internen Feldes. Die Lyapunov-Funktion $L_a[\{S_i(t)\}]$ (Gleichung (1.2) auf Seite 9) wird also in jedem Zeitschritt um den jeweils größten Betrag abgesenkt. Die Dynamik erfolgt entlang einer Untergitter-Koordinatenlinie, und zwar solange, bis diese Untergitter-Koordinate mit dem größten internen Feld gesättigt ist.

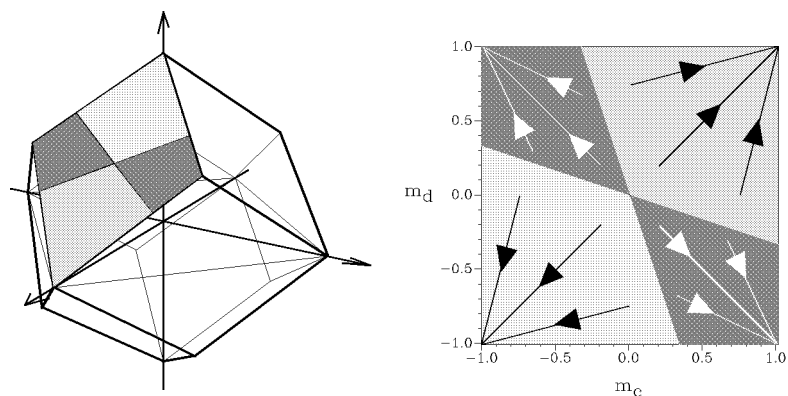


Abbildung 2.5: Auf den Facetten des Dodekaeders variieren lediglich zwei von vier der Untergitter-Magnetisierungen. Die Facette im linken Bild korrespondiert zum Untergitter-Diagramm auf der rechten Seite. Dort sind neben den Attraktionsgebieten auch exemplarisch einige Flusslinien der asynchronen Dynamik dargestellt. Sie konvergieren zu den bei $(+1, +1)$ bzw. $(-1, -1)$ liegenden Mustern (schwarze Pfeile) oder zu den zwei 3er-Mischzuständen bei $(-1, +1)$ und $(+1, -1)$ (weiße Pfeile).

Dann verändert sich die Untergitter-Koordinate mit dem zweitgrößten Feld, bis schließlich auch diese gesättigt ist. Letztlich endet die Dynamik wieder in einem Zustand, der voll durchmagnetisiert ist. In Abbildung 2.6, links, sind exemplarisch einige Flusslinien der Gradientendynamik eingezeichnet. Die gestrichelten Linien geben die Grenzen des Attraktionsgebietes bei normaler asynchroner Dynamik an.

Die sequentielle Dynamik mit dem bevorzugten Testen falschstehender Neuronen ist eine etwas artifizielle Dynamik, muß doch extern, also nicht vom Netzwerk, festgestellt werden, welche Neuronen am wahrscheinlichsten falsch stehen. Dies könnte aber z.B. bei der Bildrestauration zur Anwendung kommen, wo es Bereiche gibt, die besonders anfällig für Störungen durch Rauschen sind, wie die Ränder von zu erkennenden Zeichen. Testet man diese Bereiche zuerst, vergrößert sich das Attraktionsgebiet des dazugehörigen Musters, wie Abbildung

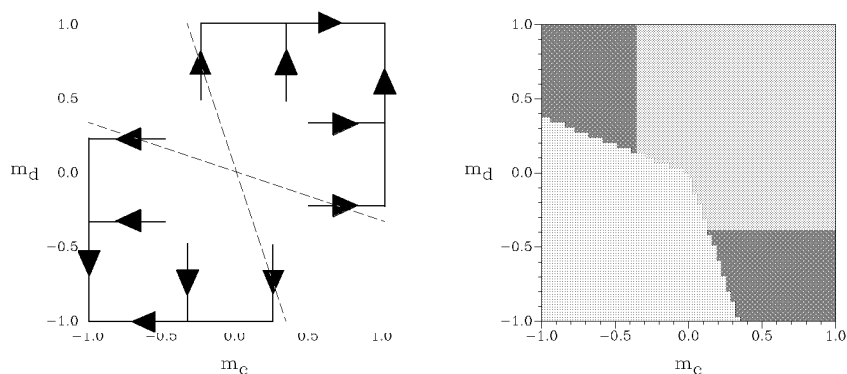


Abbildung 2.6: Die Attraktionsgebiete von Gradientendynamik und sequentieller Dynamik mit bevorzugtem Testen von falschstehenden Neuronen. Beide Diagramme sind das Ergebnis numerischer Untersuchungen.

2.6, rechts, für das bei $m_c = m_d = 1$ liegende Muster zeigt. Das Verfahren hat allerdings zwei Nachteile: es ist nicht besser als der Gradientenabstieg und es erlaubt nur die Vergrößerung des Attraktionsgebietes eines der Muster.

2.4 Exakte Enumeration von metastabilen Niveaus

Die 3er-Mischzustände sind nicht die einzigen metastabilen Niveaus, welche in Autoassoziativen Netzwerken die Dynamik behindern. Speichert man mehr als drei Muster in einem Netzwerk, treten eine Vielzahl zusätzlicher dynamischer Attraktoren auf, für die bislang keinerlei Bildungsgesetz bekannt ist.

Um alle diese metastabilen Zustände zu finden, müßte im Extremfall der komplette 2^N -dimensionale Konfigurationsraum des Neuronalen Netzwerkes durchsucht werden. Dies ist selbst bei relativ kleinen Netzwerkgrößen ein hoffnungsloses Unterfangen und Aussagen für den Limes $N \rightarrow \infty$ sind auf diese Weise überhaupt nicht zu erhalten. In der Literatur vorgeschlagene Methoden zur Verringerung der unerwünschten Attraktoren, bei denen störende metastabile Niveaus gezielt 'entlernt' werden sollen [HO83], erweisen sich schon aus diesem Grunde als nicht praktikabel. Wegen der großen Zahl der metastabilen Niveaus ist man kaum in der Lage, alle die Dynamik behindernden dynamischen Fallen zu finden.

Eine exakte Enumeration aller metastabiler Niveaus wird durch das Arbeiten im Untergitter-Raum ermöglicht. Jedes der metastabilen Niveaus liegt ja notwendigerweise auf einem der Eckpunkte des Untergitter-Hyperkubus. Dies bleibt auch für $N \rightarrow \infty$ so, dieser Grenzprozess ist also bei der Suche im Untergitter-Raum bereits implizit durchgeführt. Es macht allerdings keinen Sinn, selbst für niedrige Musterzahlen, alle $2^{(2^{p-1})}$ möglichen Ecken des Untergitter-Hyperkubus zu testen. Schon bei 7 Mustern wären dies bereits $2^{64} \approx 1.8 \cdot 10^{19}$ verschiedene Bitmuster. Zur exakten Enumeration aller metastabilen Niveaus werden deshalb noch folgende Tatsachen ausgenutzt:

- Die möglichen Musterüberlappungen m_μ der metastabilen Niveaus liegen im Intervall $[-1, +1]$ äquidistant im Abstand $\Delta = 2^{-(p-2)}$.
- Die Lyapunov-Funktion $L_a[\{S_i(t)\}]$ (Seite 9) ist eine quadratische Funktion der Musterüberlappungen. Ist also $(m_1, m_2, \dots, m_k, \dots, m_p)$ ein metastabiles Niveau, so auch $(m_1, m_2, \dots, -m_k, \dots, m_p)$. Ferner ist die Funktion invariant unter der Vertauschung von Musterüberlappungen. Es genügt damit, nur Überlapp-Kombinationen zu testen, für die $m_k \leq m_{k+1}$ gilt.
- Die Musterüberlappungen sind, wie dargestellt, nicht unabhängig voneinander. Allerdings werden die Abhängigkeit schnell zu komplex, um numerisch vorteilhaft ausgewertet zu werden. Konkret wurde nur ausgenutzt, daß $m_2 \leq 1 - m_1$ sein muß.
- Es genügt, für alle $2^{(p-1)}$ Untergitter das interne Feld h_a zu bestimmen. Andere interne Felder treten nicht auf.

Der implementierte Algorithmus nutzt die oben beschriebenen Tatsachen aus, um die Zahl der zu testenden Bitkombinationen erheblich zu reduzieren. Kurz skizziert wurde folgendes gemacht:

1. Für alle erlaubten Überlapp-Werte m_μ wurden die internen Felder $h_a = \sum_\mu \xi_a^\mu m_\mu$ der Untergitter bestimmt.

Tabelle 2.1: Die Werte der in der Tabelle aufgeführten Musterüberlappungen sind ebenso wie die Werte der internen Felder mit 64 multipliziert. Der Eintrag in der Symmetrie-Spalte gibt jeweils an, wie oft sich der Wert eines Musterüberlapps wiederholt. Dabei sind die Werte von links nach rechts in fallender Reihenfolge geordnet.

2. Die internen Felder h_a legen über $m_a = \text{sign}(h_a)$ die UG-Magnetisierungen m_a fest.
3. Falls die so erhaltene Bitkombination wieder über $m_\mu = \sum_a \xi_a^\mu m_a$ die ursprünglichen Überlapp-Werte reproduziert, wurde ein metastabiles Niveau gefunden.

Trotz der durch diesen Algorithmus erfolgten, erheblichen Reduzierung der Anzahl der zu testenden Bedingungen bleibt die Suche nach metastabilen Niveaus numerisch sehr aufwendig. Die zur Suche benötigte Zeit wächst exponentiell mit der Anzahl der Muster. Die exakte Enumeration aller metastabilen Niveaus wurde deshalb auch nur bis zu $p = 7$ Mustern durchgeführt. Insgesamt findet man bei 5 gelernten Mustern schon 1292 stabile Systemzustände, während bei 7 gelernten Mustern bereits 3 548 358 dynamische Attraktoren, welche in 40 verschiedene Klassen eingeteilt werden können (vergl. Tabelle), existieren!

Die Hoffnung, aus der Vielzahl der Attraktoren Bildungsgesetze herzuleiten, erfüllte sich leider nicht. Zwar lassen sich einige der metastabilen Niveaus als Derivate anderer metastabiler Niveaus interpretieren — so teilt sich der symmetrische 3er-Mischzustand in einen unsymmetrischen 23er- oder einen 25er-Mischzustand, der symmetrische 5er-Mischzustand in einen 43er-Mischzustand — aber eindeutige Bildungsgesetze sind trotzdem nicht zu identifizieren.

Man findet allerdings nur bestimmte Symmetrien bei den Mischzuständen realisiert. So können beispielsweise alle symmetrischen Mischzustände mit einer geraden Anzahl von Mustern, also die 2er-, 4er-, 6er- usw. Mischzustände dynamisch nicht stabil sein, denn hier gibt es immer ein Untergitter, in dem das interne Feld h_a identisch null wird. Aber auch ein unsymmetrischer 4er-Mischzustand existiert nicht, allerdings gibt es unsymmetrische 6er-Mischzustände (die Nummern 9, 11 und 13). Es gibt ferner noch einen total asymmetrischen Mischzustand, bei dem keiner der Musterüberlappungen gleich ist (die Nummer 35). Trotzdem scheint auch über die Symmetrie-Eigenschaften der metastabilen Niveaus kein Bildungsgesetz zu folgen.

Generell findet man mit der Erhöhung der Anzahl der gelernten Muster die Tendenz, daß sowohl relativ hohe als auch relativ niedrige Musterüberlappungen bei den metastabilen Niveaus zu beobachten sind. So findet man in der Tabelle beim Mischzustand Nr. 25 einen maximalen Musterüberlapp von 0.65625, während man bei den metastabilen Niveaus Nr. 15,16,18 und 21 bis 23 als minimalen Musterüberlapp 0.0625 feststellt. Dies legt die Vermutung nahe, daß sich bei wachsenden Musterzahlen die metastabilen Niveaus immer näher an die Muster bzw. an den Ursprung heranschieben. Letzteres kann man für die symmetrischen Mischzustände analytisch zeigen (vergl. Abschnitt 3.1.2 und [AM85]).

2.5 Ein Netzwerk ohne metastabile Niveaus

2.5.1 Prinzipielle Idee

Die große Zahl der metastabilen Niveaus schränkt das Attraktionsgebiet der wenigen gelernten Muster erheblich ein. Man kann aber die Tatsache, daß alle Mischzustände kleinere interne Felder als die Muster besitzen (siehe Tabelle 2.1), ausnutzen, um diese unerwünschten Attraktoren zu destabilisieren.

In der Literatur werden dazu im wesentlichen zwei verschiedene Methoden vorgeschlagen. Die in Abschnitt 1.3.4 definierte, stochastische Dynamik führt dazu, daß Neuronen mit kleinerem internen Feld mit größerer Wahrscheinlichkeit antiparallel zu diesem internen Feld eingestellt, also destabilisiert werden. Ist das

Rauschen stark genug, kann dadurch sogar der komplette Mischzustand destabilisiert werden. Wie eine genaue Analyse zeigt, sind bei geschickt gewählter, genügend hoher Temperatur im Netzwerk nur noch die gelernten Muster als dynamische Attraktoren zu finden [PE89]. Allerdings hat diese Methode der Destabilisierung zwei entscheidende Nachteile — zum einem ist unter stochastischer Dynamik der Assoziationsvorgang erheblich langsamer als bei einer $T = 0$, also deterministischen Dynamik, zum anderen werden durch das Rauschen auch die Muster destabilisiert: die Attraktoren sind nicht mehr mit den Mustern identisch, sondern das Netzwerk läuft auf einer stochastischen Trajektorie in der Nähe der Muster durch den Zustandsraum.

Eine andere Methode der Destabilisierung versucht, die unerwünschten Attraktoren aus der Kopplungsmatrix zu entfernen, ein Verfahren, welches generell als 'Entlernen' (Unlearning) bezeichnet wird [HO83]. Diese Vorgehensweise zeichnet allerdings ein prinzipieller Nachteil aus: man kann nur solche metastabilen Zustände entlernen, die man *kennt*. Ein angesichts der schon bei nur 7 Mustern gefundenen großen Anzahl von metastabilen Zuständen ein eher hoffnungsloses Unterfangen.

Hier soll nun eine alternative Methode vorgestellt werden, welche die Konstruktion einer Kopplungsmatrix ermöglicht, bei der wirklich nur die gelernten Muster als dynamische Attraktoren auftreten. Auch hier wird die Tatsache ausgenutzt, daß metastabile Zustände betragsmäßig immer kleinere interne Felder als die Muster haben. Zur Destabilisierung wird aber verwendet, daß Mischzustände, die ja keine echten Linearkombinationen der Muster sind, Komponenten in dem zum Raum der Muster aufgespannten orthogonalen Raum besitzen. Diese Anteile können benutzt werden, um die internen Felder der metastabilen Niveaus entsprechend zu modifizieren.

2.5.2 Konstruktion der Kopplungsmatrix

Diskussion im Untergitter-Raum

Der Projektor in den von den gelernten Mustern aufgespannten Raum ist im Untergitter-Formalismus einfach zu bestimmen. Da die zu den Mustern korrespondierenden Untergitter-Vektoren orthogonal zueinander sind, ist der Projektor \mathbb{P} einfach durch

$$\mathbb{P}_{ab} = D^{-1} \sum_{\nu} \xi_a^{\nu} \xi_b^{\nu}$$

gegeben. Dies ist nichts anderes als die Hopfield-Matrix im Untergitter-Raum. Dabei ist $D = 2^{(p-1)}$ die Dimension des Untergitter-Raumes.

Wir untersuchen nun die internen Felder der Kopplungsmatrix

$$\begin{aligned} \mathcal{C} &= \mathbb{P} - \lambda \mathbb{P}_{\perp} \\ &= \mathbb{P} - \lambda(\mathbb{1} - \mathbb{P}) . \end{aligned}$$

Diese Kopplungsmatrix hat die Komponenten

$$\mathcal{C}_{ab} = (1 + \lambda) \left\{ D^{-1} \sum_{\nu} \xi_a^{\nu} \xi_b^{\nu} - J \delta_{ab} \right\} ,$$

wobei wir

$$J = \frac{\lambda}{1 + \lambda}$$

eingeführt haben.

Ein beliebiger Untergitter-Vektor $\eta_a = \pm 1$ besitzt mit dieser Kopplungsmatrix das interne Feld

$$h_a = \sum_a C_{ab} \eta_a = (1 + \lambda) \left\{ \sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug} - J \eta_a \right\}.$$

Für dynamisch stabile Zustände, also für $\eta_a = \text{sign}(h_a)$ muß damit gelten:

$$\eta_a = \text{sign} \left(\sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug} - J \eta_a \right).$$

Da $J > 0$ ist, geht dies nur, falls

$$\forall_a : \quad \left| \sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug} \right| > J \quad (2.3)$$

gilt.

Nun unterscheiden sich die Mischzustände von den Mustern durch die kleineren Werte ihrer internen Felder. Bei geeignet gewähltem J kann also Bedingung (2.3) für die metastabilen Niveaus gezielt verletzt und diese dadurch destabilisiert werden.

Dazu zeigen wir zunächst, daß für metastabile Niveaus mindestens ein Untergitter $a \in \{0, \dots, p\}$ existiert, für das

$$\left| \sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug} \right| \leq 1 - \Delta$$

ist. Wir nutzen dazu die Tatsache, in ξ_a^{ν} alle möglichen Vorzeichenkombinationen zu finden (vergl. Seite 15), also in $|\sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug}|$ beliebige Summen und Differenzen der Musterkoordinaten bilden zu können.

Wir ordnen die Musterkoordinaten zunächst so an, daß

$$|m_k| \leq |m_{k+1}|$$

gilt. Sind nun die zwei letzten Koordinaten betragsmäßig gleich, so verschwindet entweder deren Summe oder deren Differenz. Sind sie ungleich, dann lassen sie sich zumindest so zusammenfassen, daß

$$\tilde{m}_k \equiv |m_k \pm m_{k+1}| \leq \max(|m_k|, |m_{k+1}|)$$

ist. Wir ersetzen jetzt m_k durch \tilde{m}_k ; die Anzahl der Musterkoordinaten hat sich damit um eins erniedrigt. Wir fahren jetzt mit dieser Konstruktion fort, fassen also m_{k-1} und \tilde{m}_k zusammen, bis nur nur eine einzige Musterkoordinate übrigbleibt. Für diese gilt dann

$$|\tilde{m}_1| \leq \max(|m_1|, \dots, |m_p|) < 1 - \Delta.$$

Damit ist obige Behauptung gezeigt. Sie erlaubt uns nun, gezielt *alle* Mischzustände zu destabilisieren. Für die gelernten Muster haben wir nämlich

$$\sum_{\nu} \xi_a^{\nu} m_{\nu}^{ug} = 1.$$

Wählen wir deshalb $1 - \Delta < J < 1$, sind nur noch die reinen Muster stabil!

2.5.3 Transformation in Zustandsraum

Unsere Diskussion wurde im Untergitter-Raum durchgeführt. Gelingt es uns, im Zustandsraum des Neuronalen Netzes eine Kopplungsmatrix zu konstruieren, welche dort dieselben internen Felder erzeugt, werden auch hier die metastabilen Niveaus destabilisiert. Eine zudem symmetrische Kopplungsmatrix sorgt im Zusammenspiel mit der durch die Lyapunov-Funktion $L_a[\{S_i(t)\}]$ getriebenen Dynamik für optimale Attraktionsgebiete der gelernten Muster, da diese als einzige noch die Stabilitätsforderung (2.3) erfüllen.

Betrachten wir dazu das interne Feld des Neurons i . Gehört es der Untergitter-Klasse $ug(i)$ an, so sollte es den Wert

$$h_i = (1 + \lambda) \left\{ \sum_{\nu} \xi_{ug(i)}^{\nu} m_{\nu}^{ug} - J m_{ug(i)} \right\} \quad (2.4)$$

besitzen.

Dies kann nun durch die symmetrische Kopplungsmatrix

$$J_{ij} = \frac{C_{ug(ij)}}{N_{ug(i)} N_{ug(j)}}$$

erreicht werden; wir erhalten für die internen Felder

$$\begin{aligned} h_i &= (1 + \lambda) N_{ug(i)}^{-1} \left\{ \left\{ D^{-1} \sum_{j,\nu} N_{ug(j)}^{-1} \xi_{ug(i)}^{\nu} \xi_{ug(j)}^{\nu} S_j - J \sum_j N_{ug(j)}^{-1} \delta_{ug(ij)} S_j \right\} \right\} \\ &= (1 + \lambda) N_{ug(i)}^{-1} \left\{ \left\{ D^{-1} \sum_{\nu} \xi_{ug(i)}^{\nu} \sum_{a \in ug} N_a^{-1} \xi_a^{\nu} \sum_{ug(j)=a} S_j \right\} - J m_{ug(i)} \right\} \\ &= (1 + \lambda) N_{ug(i)}^{-1} \left\{ \left\{ D^{-1} \sum_{\nu} \xi_{ug(i)}^{\nu} m_{\nu}^{ug} - J m_{ug(i)} \right\} \right\}. \end{aligned} \quad (2.5)$$

Dies ist, bis auf den Skalierungsfaktor $N_{ug(i)}^{-1} > 0$, identisch mit Gleichung (2.4). Die symmetrische Kopplungsmatrix J_{ij} erzeugt also im Zustandsraum des Neuronalen Netzwerkes die gleiche Vorzeichenverteilung wie C_{ab} im Untergitter-Raum. Sie ist eine optimale Kopplungsmatrix, bei der nur die gelernten Muster als Attraktoren der Dynamik auftreten.

Für die Gültigkeit dieser Konstruktion wurden stillschweigend zwei Bedingungen vorausgesetzt. So müssen

1. alle Untergitter realisiert sein, sonst wird die Orthogonalitätsrelation zwischen den Mustervektoren im Untergitter-Raum (Gleichung (2.1) auf Seite 15), die zur Konstruktion des Projektionsoperators benutzt wurde, verletzt.
2. für die Untergittermächtigkeiten muß $N_a \gg 1$ gelten, sonst reduziert sich in Gleichung (2.5) der Term $J \sum_j N_{ug(j)}^{-1} \delta_{ug(ij)} S_j$ nicht auf $J m_{ug(i)}$.

Beide Forderungen werden durch $N \gg 2^{(p-1)}$, also Anzahl der Neuronen sehr viel größer als Anzahl der Untergitter, sichergestellt. Dies ist leider auch eine deutliche Einschränkung der Anwendbarkeit dieser Konstruktion, denn bei vorgegebener Neuronenzahl N können dann nur noch $\sim \ln(N)$ viele Muster optimal eingespeichert werden. Aber es bleibt immerhin festzuhalten, daß es zumindest in diesen Fällen möglich ist, eine Kopplungsmatrix mit optimalem dynamischen Verhalten zu konstruieren.

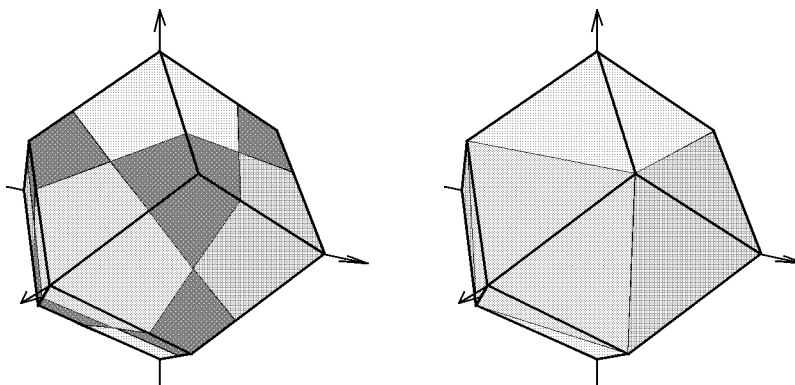


Abbildung 2.7: Ein Vergleich der Attraktionsgebiete von Hopfield-Modell (links) und optimalem Modell (rechts). Beim optimalen Modell sind die unerwünschten Mischzustände (links dunkel dargestellt) verschwunden. Die Muster haben jeweils das maximal mögliche Attraktionsgebiet. Die Schemazeichnung wurde anhand numerischer Simulationen der beiden Modelle angefertigt.

Die numerischen Simulationen mit der oben definierten optimalen Kopplungsmatrix ergeben in der Tat das größtmögliche Attraktionsgebiet für die eingespeicherten Muster. Bei geeigneter Wahl von J können sogar Mischzustände mit niedriger Symmetrie selektiv destabilisiert werden. In Abbildung 2.7 sind schematisch die Resultate von numerischen Simulationen für Hopfield-Modell und optimales Modell dargestellt. Interessanterweise wurden optimale Attraktionsgebiete bei 6 eingespeicherten Mustern schon ab Neuronenzahlen um die 100 erreicht. Die Zahl der Untergitter lag hier bei $D = 2^5 = 32$, es galt also keineswegs die Bedingung $N \gg 2^{(p-1)}$. Die Untergitter-Mächtigkeiten lagen typischerweise bei $N_a \approx 3$. Eine Verletzung der oben erwähnten, impliziten Annahmen bei der Ableitung der Kopplungsmatrix, also auch die Forderung $N \gg 2^{(p-1)}$, scheint deshalb nicht besonders kritisch zu sein. Selbst für noch niedrigere Neuronenzahlen, bei denen dann nicht einmal alle Untergitter realisiert wurden, konnte durch die optimale Kopplungsmatrix noch eine deutliche Verbesserung der Attraktionsgebiete der Muster erreicht werden.

2.5.4 Destabilisierung der 3er-Mischzustände

Die Destabilisierung der Mischzustände durch die optimale Kopplungsmatrix kann gut am Beispiel der 3er-Mischzustände diskutiert werden. In einem Netzwerk, das drei Muster gespeichert hat, hat die optimale Kopplungsmatrix eine einfache Gestalt. Der Untergitter-Raum ist nämlich lediglich vierdimensional und drei der Raumrichtungen werden bereits von den Untergitter-Vektoren der drei Muster aufgespannt. Damit reduziert sich der Projektor in den Orthogonalraum der Muster auf das dyadische Produkt des einen, zu den Mustervektoren orthogonalen Vektors. Wir werden die Stabilität der 3er-Mischzustände mit Hilfe einer Mittleren-Feld-Rechnung, die im hier betrachteten Limes, $\alpha \rightarrow 0$, exakt ist, analysieren.

Konkret haben wir im Untergitter-Raum die folgende Situation vorliegen, wobei

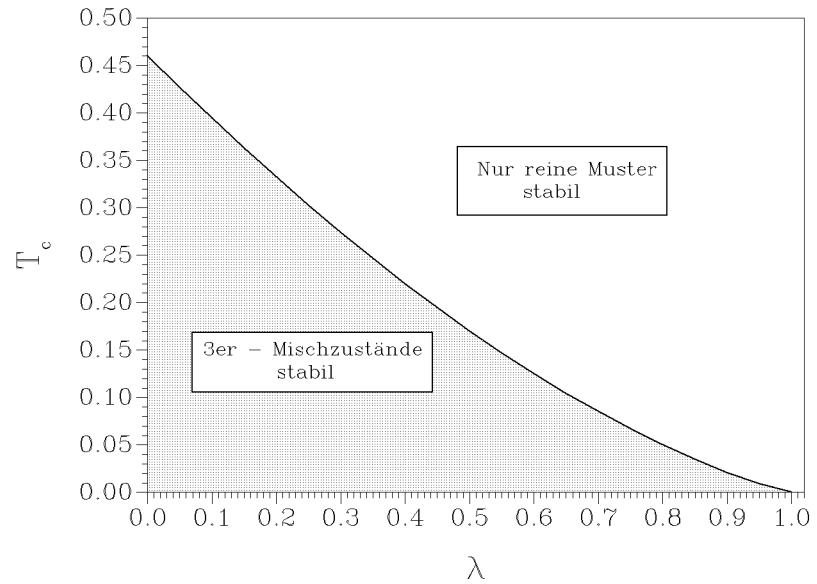


Abbildung 2.8: Das Stabilitätsgebiet der 3er-Mischzustände als Funktion von Temperatur und Destabilisierungsparameter λ . Nur im schattierten Gebiet sind die 3er-Mischzustände stabile Konfigurationen des Netzwerkes.

bestimmt. Hierbei ist q durch

$$q = \langle\langle \tanh^2 \left(\sum_{\mu=1}^3 m_{\mu} \xi^{\mu} - i\sqrt{\lambda} m_4 \xi^4 \right) \rangle\rangle$$

und die q_{ik} durch

$$q_{ik} = \langle\langle \xi^i \xi^k \tanh^2 \left(\sum_{\mu=1}^3 m_{\mu} \xi^{\mu} - i\sqrt{\lambda} m_4 \xi^4 \right) \rangle\rangle$$

gegeben.

Eine Auswertung der Stabilitätsmatrix liefert das in Abbildung 2.8 dargestellte Phasendiagramm. Bei genügend hoher Temperatur und/oder genügend großem Destabilisierungsparameter λ sind die 3er-Mischzustände nicht mehr stabil. Obwohl das Phasendiagramm von Abbildung 2.8 eine gewisse Äquivalenz zwischen Destabilisierungsparameter und Temperatur nahelegt, ist die Situation qualitativ völlig verschieden. Eine Destabilisierung durch Temperatur, wie von manchen Autoren vorgeschlagen [PE89, AM89], führt zu einem langsamen und, durch die hohe Temperatur bedingt, nie vollständigen Restaurierungsprozess. Im Extremfall werden sogar Muster erkannt, von denen im verrauschten Startmuster gar keine Anteile zu finden waren (ein Beispiel findet sich, allerdings ohne einen diesbezüglichen Kommentar, in [AM89], Abbildung 4.2, Seite 165). Im Gegensatz dazu ändert die hier vorgeschlagene, synaptische Destabilisierung nichts am Wiedererkennungsprozess — wir haben immer noch die schnelle $T = 0$ -Dynamik — und führt zu einer vollständigen Restauration der gelernten Muster.

2.6 Zweierzyklen

Die synchrone Dynamik besitzt außer den metastabilen Zuständen noch die Zweierzyklen als dynamische Attraktoren. Hier wechseln periodisch, von einem Zeitschritt zum nächsten, die Gesamtheit oder Teile der Neuronen ihren Zustand. Auch über Zweierzyklen ist wenig bekannt, deshalb sollen an dieser Stelle die Ergebnisse einiger numerischer Untersuchungen präsentiert werden, die ein wenig die Art und Weise, wie Zweierzyklen dynamisch entstehen, erhellen können.

Da in einem Zweierzyklus die Zustandsvariablen $S_i(t)$ einiger Neuronen oszillieren, sollte dies eigentlich zu ebenfalls oszillierenden Musterüberlappungen $m_\mu(t)$ führen. Die numerischen Untersuchungen, die im wesentlichen ähnlich den Untersuchungen bei der exakten Enumeration der metastabilen Zustände durchgeführt wurden, zeigen allerdings ein anderes, typisches Szenario eines 2er-Zyklus, bei dem die Musterüberlappungen $m_\mu(t)$ zeitlich konstant sind.

Betrachten wir dazu das interne Feld der Neuronen in einem endlichen neuronalen Netzwerk:

$$h_i = \sum_{\mu} \xi_i^{\mu} m_{\mu} - \frac{p}{N} S_i$$

Wir setzen wieder voraus, daß N so groß ist, daß alle Untergitter mit genügend großer Untergitter-Mächtigkeit N_a realisiert sind. In einem Zweierzyklus sind dann zwei verschiedene Gruppen von Untergittern zu finden. Für die eine Gruppe gilt

$$\left| \sum_{\mu} \xi_i^{\mu} m_{\mu} \right| > \frac{p}{N},$$

der Untergitterterm dominiert also im internen Feld und damit wird der Zustand der Neuronen eindeutig festgelegt; diese Untergitter sind voll durchmagnetisiert und ihre Neuronen ändern ihren Zustand zeitlich nicht.

In der anderen Gruppe gilt

$$\left| \sum_{\mu} \xi_i^{\mu} m_{\mu} \right| < \frac{p}{N},$$

damit folgt für die Zustandsvariablen der Neuronen

$$S_i(t+1) = \text{sign}(h_i(t)) = \text{sign}\left(-\frac{p}{N} S_i(t)\right) = -S_i(t).$$

Die Neuronen dieser Untergitter wechseln also ihre Vorzeichen im jedem Zeitschritt. Damit ändern sich natürlich auch die Magnetisierungen m_a dieser Untergitter. Dies müßte normalerweise zu oszillierenden Musterüberlappungen $m_\mu(t)$ führen, es stellt sich jedoch heraus, daß die oszillierenden Untergitter gerade verschwindende UG-Magnetisierungen, $m_a = 0$, haben. Damit ändern sich natürlich auch die Musterüberlappungen zeitlich nicht, und die Einteilung in oszillierende und nichtoszillierende Untergitter bleibt erhalten.

Im nächsten Zeitschritt wechseln also wieder nur die Zustände der Neuronen das Vorzeichen, welche bereits im Zeitschritt vorher ihren Zustand invertiert hatten. Damit sind wir wieder beim Ausgangspunkt der Betrachtung. Dies ist der typische 2er-Zyklus, wie er bei numerischen Simulationen gefunden wurde. Komplexere Szenarien sind zwar denkbar, wurden aber nicht angetroffen.

Die numerischen Untersuchungen konnten bis $p = 4$ Muster durchgeführt werden. Dabei wurden folgende typische 2er-Zyklen gefunden:

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix},$$

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix},$$

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \\ \frac{1}{4} \end{pmatrix},$$

und schließlich

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{pmatrix} = \begin{pmatrix} \frac{3}{8} \\ \frac{3}{8} \\ \frac{3}{8} \\ \frac{3}{8} \end{pmatrix}.$$

Es sei angemerkt, daß sich in der obigen Auflistung auch die symmetrischen 2er- und 4er-Mischzustände finden. Generell gilt, daß diese, bezüglich der asynchronen Dynamik instabilen, symmetrischen Mischzustände mit der Beimischung einer geraden Anzahl von Mustern, in der synchronen Dynamik zu Zweierzyklen werden. Außerdem sind noch *alle* Zustände, die vollständig im Orthogonalraum der Muster liegen, für die also $m_\mu = 0$ gilt, Zweierzyklen. Daraus resultiert eine sehr große Anzahl von Zweierzyklen. Dynamisch wichtiger aber, dies zeigen die numerischen Simulationen, sind zumindest für $\alpha \rightarrow 0$ die Punktattraktoren der metastabilen Niveaus. Die Zweierzyklen scheinen in diesem Falle keine Rolle zu spielen (vergl. auch Abbildung 4.12 auf Seite 74).