Kapitel 3: Netzwerke bei hoher Speicherdichte

Die komplexe Struktur des Zustandraumes Neuronaler Netzwerke wurde im vorherigen Kapitel im Limes $\alpha \to 0$ untersucht. Von größerem praktischen Interesse ist aber der Limes $\alpha > 0$, bei dem das Netzwerk unter wesentlich höherer Speicherdichte operiert. Die Analysetechniken des vorherigen Kapitels sind in diesem Falle nur noch bedingt anwendbar, da nur wenige der $2^{(p-1)}$ vielen möglichen Untergitter realisiert sind. Zudem liegen die Untergitter–Mächtigkeiten der dann noch realisierten Untergitter typischerweise bei $N_a \approx 1$, d.h., maximal ein Neuron ist in einer der Untergitter–Klassen zu finden. Hier müssen also modifizierte Analysetechniken zur Anwendung kommen.

Die Erhöhung der Speicherdichte wird zu einem starken Anwachsen der Zahl der metastabilen Niveaus führen. Es stellt sich die Frage, welche Auswirkungen dies auf die Struktur des Zustandsraumes und die Dynamik Neuronaler Netze hat. Dies soll exemplarisch durch Analyse mit Mittleren–Feld–Methoden, approximativen dynamischen Beschreibungen und durch Resultate numerischer Simulationen erörtert werden.

3.1 Statische Resultate

Das mit Mitteln der statistischen Physik am besten erforschte Neuronale Netzwerk ist das Hopfield–Modell. In diesem Abschnitt wird eine kurze Zusammenfassung von Resultaten gegeben, die aus Untersuchungen Neuronaler Netzwerke im thermischen Gleichgewicht stammen und für unsere folgende dynamische Analyse von Interesse sind.

3.1.1 Signal/Rauschanalyse

In einem vorgegebenen Systemzustand haben typischerweise nicht alle der $p = \alpha N$ Muster–Überlappungen makroskopische Werte. So findet man in einem 3er–Mischzustand drei kondensierte Muster, mit Überlappungen von etwa 0.5, während die Anteile der übrigen Muster lediglich von der Ordnung $\mathcal{O}(\frac{1}{\sqrt{N}})$ sind.

Berücksichtigt man diese Separation der Muster in kondensierte und unkondensierte Muster, lassen sich die Untergitter–Techniken des vorigen Kapitels auf die kondensierten Muster wieder anwenden. Sind die ersten k Muster kondensiert, so können wir das interne Feld als Summe zweier unterschiedlicher Terme, eines Signal– und eines Rauschterms, darstellen:

$$h_i = \underbrace{\sum_{\mu \leq k} \xi_i^\mu m_\mu}_{\text{Signal}} + \underbrace{\sum_{\mu > k} \xi_i^\mu m_\mu}_{\text{Rauschen}}$$

Führen wir jetzt eine Untergitter–Klasseneinteilung bezüglich der kondensierten Muster ein, so wird der Signalterm wieder eine Funktion der Untergitter–Klasse. Innerhalb der Untergitter hat er jeweils einen festen Wert. Da die Anzahl der kondensierten Muster, k, für $N \to \infty$ konstant bleibt, werden auch alle Untergitter realisiert. Der Signalterm kann hierbei $2^{(k-1)}$ verschiedene Werte annehmen.

Die Überlappungen mit den übrigen, nichtkondensierten Muster erzeugen im internen Feld der Neuronen den Rauschterm, der sich dem Signalterm überlagert. Ist er klein genug, stört er das Signal nicht, und wir haben eine der Analyse des zweiten Kapitels äquivalente Situation vorliegen. Alle der dort gefundenen Mischzustände sind also prinzipiell auch im Falle $\alpha > 0$ realisierbar.

Wird das Rauschen jedoch zu groß, können einzelne Neuronen antiparallel zum Signalterm eingestellt werden. Dies veringert die Korrelationen des Systemzustandes an die kondensierten Muster, d.h., die kondensierten Musterkoordinaten erniedrigen sich, sodaß auch der Signalterm betragsmäßig kleiner wird. Dadurch ist es möglich, daß sich noch mehr Neuronen antiparallel zum Signalterm einstellen, ab einem gewissen, kritischen α können sich die zwei Effekte gegenseitig so aufschaukeln, daß der Systemzustand spontan destabilisiert wird.

In einer ersten Approximation wollen wir aber den Rauschterm

$$\sum_{\mu>k} \xi_i^\mu m_\mu = N^{-1} \sum_{\mu>k} \sum_{j\neq i} \xi_i^\mu \xi_j^\mu S_j$$

als eine Summe von $(N-1)(p-k)$ unabhängigen Zufallsvariablen, die ± 1–verteilt sind, ansehen. Für jedes Neuron wird damit der Rauschterm zu einer gaußverteilten Zufallsvariable mit der Varianz

$$\sigma = \sqrt{\frac{p}{N}} = \sqrt{\alpha}\ .$$

Damit erhält man für den thermischen Mittelwert der Musterüberlappungen

$$\begin{aligned} \overline{m_\mu} &= N^{-1} \sum_i \xi_i^\mu \overline{S_i} \\ &= N^{-1} \sum_i \int \xi_i^\mu \tanh(h_i) P_{h_i} dh_i \\ &= \langle\langle \xi^\mu \tanh \beta (\sum_{\nu=1}^{k} \xi^\nu m_\nu + \sqrt{\alpha} z) \rangle\rangle \end{aligned}$$

Hierbei stehen die doppelten eckigen Klammern für die durch

$$\langle\langle f(\vec{\xi}, z) \rangle\rangle = 2^{-k} \sum_{\xi^\nu = \pm 1} \int_{-\infty}^{+\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) f(\vec{\xi}, z)$$

definierte Mittelung. Die Summe über alle Bitkombinationen ξ^μ sorgt dabei für die Mittelung über die Untergitter, während die Integration über die Variable z eine Mittelung über das gaußsche Rauschen darstellt. Für $\beta \to \infty$ und $\alpha \to 0$ erhält man natürlich wieder Stabilitätsgleichungen, die jenen des vorherigen Kapitels äquivalent sind.

Prinizpiell reagieren die Neuronen mit den betragsmäßig kleinsten internen Felder am kritischsten auf eine Erhöhung der Speicherdichte. Durch den kleinen Signalterm kann hier das musterinduzierte Rauschen am ehesten angreifen, um die Neuronen antiparallel zum Signalterm einzustellen. Bezüglich der Speicherdichte α wird hierdurch eine Hierachie der Systemzustände induziert, wobei mit steigender Speicherdichte zunächst diejenigen Mischzustände destabilisiert werden, deren interne Felder betragsmäßig am kleinsten sind. Zum Schluß verlieren als letzte und stabilste Mischzustände die 3er–Mischzustände ihre Stabilität (vergl. Tabelle 2.1 auf Seite 28). Danach sind, bis zur Netzwerkkapazität α_c, nur noch die gelernten Muster dynamische Attraktoren.

3.1.2 Die Spinglas–Phase

Die Annahme unkorrelierter Zufallsvariabler im Rauschterm ist eine recht grobe Näherung, die bei einer genaueren Analyse revidiert werden muß. Für die asynchrone Dynamik wurde eine Meanfield–Rechnung durchgeführt [Am85], die auf der in Abschnitt 1.3.4 definierten Freien Energien basiert und für Neuronenzahlen $N \rightarrow \infty$ im wesentlichen exakt ist. Sie liefert Gleichungen für drei Typen von Variablen: Zunächst für die k kondensierten Musterkoordinaten m_μ die Gleichungen

$$m_\mu = \langle\langle \xi^\mu \tanh \beta(\sum_{\nu=1}^{k} \xi^\nu m_\nu + \sqrt{\alpha r} z)\rangle\rangle \ .$$

Die internen Felder, das Argument des tanh, sind also auch in dieser Näherung eine Summe aus dem Untergitter–Signalterm und einem gaußschem Rauschen z. Dieses gaußsche Rauschen hat aber nicht mehr die Varianz der einfachen Signal/Rauschanalyse, vielmehr gilt

$$\sigma = \sqrt{\alpha r}$$

mit

$$r = \frac{q}{(1-\beta+\beta q)^2}$$

Der in obiger Formel auftretende neue Ordnungsparameter q ist der sogenannte Spinglas–Ordnungsparameter. Er ist definiert durch

$$q = \overline{N^{-1} \sum_i < S_i >^2} \ .$$

Hierbei stehen die Klammern "$< >$" für ein thermisches Mittel, der Mittelungsstrich "$\overline{\ }$" für die Mittelung über die Kopplungen.

In der hier diskutierten Meanfield–Analyse ist q durch

$$q = \langle\langle \tanh^2 \beta(\sum_{\nu=1}^{k} \xi^\nu m_\nu + \sqrt{\alpha r} z)\rangle\rangle$$

gegeben. Der Parameter q beschreibt das Einfrieren einzelner Neuronen in zufällige Richtungen. Zusammen mit den Musterüberlappungen m_μ kann er benutzt werden, um verschiedene thermodynamische Phasen des Neuronalen Netzwerkes zu unterscheiden.

Bei hoher Temperatur existiert innerhalb des Neuronalen Netzwerkes keine Ordnung und die Neuronen wechseln völlig zufällig ihren Zustand. Wir können hier von einer paramagnetischen Phase sprechen. In dieser Phase sind die Ordnungsparameter m_μ und q identisch null.

Bei hinreichend niedriger Temperatur erscheinen die Attraktionsgebiete der Muster oder Mischzustände. Sie entsprechen einer ferromagnetischen Phase. Hier ist q und mindestens ein m_μ ungleich null.

Daneben existiert noch eine dritte Phase, welche insbesondere bei hohen Speicherdichten α zu finden ist, in der zwar alle $m_\mu = 0$ sind, jedoch $q \neq 0$ gilt. Hier sind also die Neuronen in zufällige, nicht zu den Mustern korrelierte Konfigurationen eingefroren. Dieses zufällige Einfrieren wurde zuerst bei den magnetischen Spingläsern gefunden, deshalb wird diese Phase als Spinglas–Phase bezeichnet.

In der Spinglas–Phase gibt es eine große Anzahl von Konfigurationen, in denen das Netzwerk einfrieren kann. Dies sind alles metastabile Niveaus, die auch hier als Fallen der Dynamik wirken. Sie sind natürlich genauso unerwünscht wie die Mischzustände. Im reinen Spinglas, also ohne ferromagnetische Musterphasen, werden exponentiell viele metastabile Niveaus gefunden, für die Anzahl N_s von metastabilen Niveaus gilt das Gesetz [TA80]

$$N_s \approx e^{0.1992\,N}$$

Dies divergiert stark für $N \to \infty$. Ähnlich starke Abhängigkeiten der Anzahl von metastabilen Niveaus von der Anzahl der Neuronen kann man auch bei Neuronalen Netzen feststellen (siehe z.B. [GA86, KU90]).

Zwischen den im vorherigen Kapitel, für $\alpha = 0$, diskutierten Mischzuständen und der Spinglas–Phase wird ein Zusammenhang vermutet. Zunächst verhalten sich die Musterüberlappungen m_μ der symmetrischen Mischzustände mit k beigemischten Mustern für $T = 0$ wie

$$m_\mu \approx \sqrt{\frac{2}{k\pi}}\ ,$$

was für $k \to \infty$ gegen $m_k \to 0$ geht. Diese Mischzustände nähern sich also im Überlapp der Spinglas–Phase. Ferner ist die Energie der Spinglas–Phase durch

$$E_{SG} = -\frac{1}{\pi} - \sqrt{\frac{2\alpha}{\pi}}$$

gegeben, was für $\alpha \to 0$ gegen die Grenzenergie $-1/\pi$ der symmetrischen Mischzustände geht. Dies legt nahe, daß die symmetrischen Mischzustände beim Erhöhen der Speicherdichte α zur Spinglas–Phase degenerieren [AM85].

Ebenso wie die Mischzustände ist die Spinglas–Phase bei niedrigem α metastabil, die gelernten Muster die globalen Minima der Freien Energie. Beim Erhöhen von α wird die Spinglas–Phase aber immer dominanter, was sich in einer Vergrößerung des Rauschterms in den internen Feldern bemerkbar macht.

Beim Hopfield–Modell findet sich, für $T = 0$, folgendes Szenario, welches man durch Analyse der obigen Meanfield–Gleichungen erhält: Zwischen $\alpha = 0.0$ und $\alpha = 0.051$ sind die gelernten Muster die globalen Minima der Energie. Oberhalb von 0.051 tauschen Spinglas–Phase und gelernte Muster die Rollen: die Spinglas–Phase wird globales Minimum, während die Muster nur noch lokale Minima der Freien Energie darstellen. Oberhalb von $\alpha = 0.138$ findet sich dann nur noch die Spinglas–Phase als thermodynamisch stabile Phase. Die Muster spielen keine Rolle mehr.

Ein ähnliches Verhalten wie die Muster zeigen die Mischzustände. Sie sind zwar auch bei $\alpha = 0$ lediglich lokale Minima der Freien Energie, dominieren aber, wie wir gesehen haben, große Bereiche des Zustandsraumes. Bei Erhöhung der Speicherdichte α werden, wie schon bei der Signal/Rauschanalyse diskutiert, die Mischzustände nacheinander destabilisiert. Dabei werden zuerst die Mischzustände mit niedriger Symmetrie destabilisiert, da diese die betragsmäßig kleinsten internen Felder haben (es sei wieder auf die Tabelle 2.1, Seite 28, verwiesen). Als letztes verlieren die 3er–Mischzustände ihre Stabilität, beim Hopfield–Modell ist dies bei $\alpha = 0.03$ der Fall.

Zum Abschluß dieses Abschnittes noch eine Bemerkung. Die obigen Meanfield–Gleichungen wurden als 'im wesentlichen exakt' bezeichnet. Eine genauere Ana-

lyse führt zu einer Modifikation der Theorie, die unter dem Namen Replika–Symmetrie–Brechung bekannt ist [Pa80]. Bezüglich der ferromagnetischen Phasen der Muster sind die Modifikationen gering, das kritische α wächst dabei von 0.138 auf 0.145 , aber sie haben Auswirkungen auf die Spinglas–Phase [Cr86, Am89]. Dies manifestiert sich u.a. in einer modifizierten Feldverteilung für den Rauschterm, der dann nicht mehr gaußverteilt ist. Dies sollte aber nicht überraschen, denn der gaußverteilte Rauschterm war schon in obiger Meanfield–Gleichung kein Indiz für die Addition vieler unkorrelierter Zufallsgrößen — sonst wäre ja die einfache Signal/Rauschanalyse korrekt gewesen.

3.2 Dynamische Resultate

Ist die Analyse Neuronaler Netzwerke in thermodynamischen Gleichgewichtszuständen schwierig, erfordert die exakte Behandlung der Dynamik noch größere Anstrengungen. Wir haben es im allgemeinen mit einem komplizierten stochastischen Prozess zu tun, in dem sich zweifelsohne die Komplexität der Thermodynamik des Neuronalen Netzes wiederspiegelt. Man kann allerdings auch hier verschiedene Näherungen einführen, die dann die Dynamik des Neuronalen Netzwerkes in verschiedener Qualität beschreibbar machen. Zwei dieser Näherungen sollen im folgenden vorgestellt werden.

Wir beschränken uns für den Rest dieser Arbeit auf die synchrone Dynamik von Netzwerken, da die asynchrone Dynamik, neben der durch die Temperatur implizierten Stochastizität, eine analytisch schwer zu handhabende Stochastizität bezüglich der Schaltreihenfolge der Neuronen einführt.

3.2.1 Definition des stochastischen Prozesses

An sich ist die dynamische Evolution eines Neuronalen Netzwerkes, zumindest, wenn kein Temperaturrauschen vorliegt, unter der synchronen Dynamik ein völlig deterministischer Prozess. Jeder vorgegebener Startzustand $S_i(t = 0)$ resultiert in einer eindeutigen Trajektorie durch den Zustandsraum, und, für $t \to \infty$, im Erreichen eines der dynamischen Attraktoren. Man ist allerdings nicht an einer solchen mikroskopischen Beschreibung des Assoziationsprozesses interessiert, da bei einer solchen Analyse sowohl der konkrete Startzustand $S_i(0)$ als auch die aktuell gelernten Muster ξ_i^μ eine wichtige Rolle spielen. Vielmehr interessiert man sich für das *typische* Verhalten des Neuronalen Netzwerkes, was eine geeignete Mittelung über Startzustand und gelernte Muster impliziert. Dadurch wird die Dynamik des Netzwerkes ein stochastischer Prozess, mit Startzuständen und gelernten Mustern als den ursächlichen Zufallsgrößen.

Betrachten wir zunächst den einfachen Fall eines einzigen kondensierten Musterüberlapps $m_1 = \mathcal{O}(1)$. Wir setzen das dazugehörige Muster $\xi_i^1 = 1$, was durch Umeichung immer erreichbar ist (Anhang C). Da wir die Startzustände $S_i(0)$ nicht weiter spezifizieren wollen, stellt

$$S_i(0) = \begin{cases} +1 & \text{mit der Wahrscheinlichkeit } \frac{1+m(0)}{2} \\ -1 & \text{mit der Wahrscheinlichkeit } \frac{1-m(0)}{2} \end{cases}$$

eine geeignete Anfangsbedingung dar. Es gilt damit

$$\overline{m_1(0)} = m(0) \qquad \text{und, für } \mu > 1, \qquad \overline{m_\mu(0)} = \mathcal{O}(\frac{1}{\sqrt{N}}) \; .$$

Der Musterüberlapp $m(t+1)$ zu späteren Zeitpunkten wird damit

$$\begin{aligned}\Big[m(t+1)\Big]_{S_i(0)} &= \Big[N^{-1}\sum_i \xi_i^1 S_i(t+1)\Big]_{S_i(0)} \\ &= \Big[N^{-1}\sum_i \mathrm{sign}(h_i(t))\Big]_{S_i(0)} \\ &= \int dh\; P_t(h)\,\mathrm{sign}(h)\;.\end{aligned}$$

womit die gemittelte Wahrscheinlichkeitsverteilung der inneren Felder

$$P_t(h) = \frac{1}{N}\sum_i \Big[\delta(h-h_i(t))\Big]_{S_i(0)} \tag{3.1}$$

definiert worden ist. Hierbei ist wieder $\xi_i^1 = 1$ verwendet worden.

Die Wahrscheinlichkeitsverteilung der internen Felder, $P_t(h)$ ist für unsere weitere Diskussion von entscheidender Bedeutung. Vergleicht man die dynamische Evolutionsgleichung

$$m(t+1) = \int dh\; P_t(h)\,\mathrm{sign}(h) \tag{3.2}$$

mit der korrespondierenden Meanfield–Gleichung für den Überlapp im thermischen Gleichgewicht,

$$m = \int \frac{dz}{\sqrt{2\pi}}\;\exp\Big(-\frac{z^2}{2}\Big)\;\mathrm{sign}(m_1+\sqrt{\alpha r}z)\;,$$

so erkennt man, daß in den Momenten der Feldverteilung $P_t(h)$, für $t\to\infty$, alle Ordnungsparameter der Meanfield–Theorie versteckt sind: in den stationären Zuständen korrespondiert der Mittelwert der internen Felder mit dem Überlapp zum kondensierten Muster und die Varianz der Feldverteilung hängt eng mit dem Spinglasparameter q zusammen.

Um $P_t(h)$ zu bestimmen, müssen wir die Kopplungen J_{ij} festlegen. Diese hängen wiederum von den gewählten Mustern ab. Da wir an dem typischen Systemverhalten interessiert sind, wählen wir einfach

$$\xi_i^\mu = \pm 1 \qquad \text{mit der Wahrscheinlichkeit} \qquad p = \frac{1}{2}\;.$$

Die Wahl solcher, keinen der Neuronenorte auszeichnenden Muster hat weitreichende Konsequenzen. Die internen Felder der einzelnen Neuronen werden nämlich dadurch selbstmittelnd, d.h., wir können das Ortsmittel in Gleichung (3.1) durch eine Mittelung über die gelernten Muster ersetzen.

3.2.2 Die ersten Zeitschritte — Cavity–Argumente

Für die ersten zwei Zeitschritte können die Feldverteilungen der beiden wichtigsten Neuronalen Netzwerke, des Hopfield–Modell und des Netzwerk mit pseudoinverser Kopplungsmatrix, durch relativ einfache Argumente bestimmt werden.

Das Feld zum ersten Zeitschritt

Das interne Feld zum ersten Zeitschritt

$$h_i(0) = \sum_j J_{ij} S_j(0)$$

ist eine einfache gaußverteilte Zufallsvariable. Um dies zu sehen, schreiben wir das Feld als

$$\begin{aligned} h_i(0) &= \sum_j J_{ij}\{m_0 + \delta S_j(0)\} \\ &= h_i^1 m_0 + \underbrace{\sum_j J_{ij}\delta S_j(0)}_{\equiv u} , \end{aligned} \tag{3.3}$$

womit das interne Feld im ersten Muster, $h_i^1 = \sum_j J_{ij}$, und eine neue Zufallsvariable δS_i, deren Mittelwert bezüglich der Anfangsbedingung verschwindet, definiert worden ist.

Wir ersetzen jetzt das Ortsmittel in (3.1) durch ein Mittel über die gelernten Muster. Der Term h_i^1 ist eine Zufallsvariable bezüglich dieser Muster. Bei der pseudoinversen Kopplungsmatrix ergibt sich eine δ–Verteilung bei $(1-\alpha)$ (siehe Anhang B), beim Hopfield–Modell erhalten wir eine Gaußverteilung mit Mittelwert

$$\overline{h_i^1} = \overline{1 + \sum_{\nu>1}\sum_j \xi_i^\nu \xi_j^\nu} = 1$$

und Varianz

$$\begin{aligned} \overline{\Delta(h_i^1)^2} &= \overline{\left(1 + N^{-1}\sum_{\nu>1}\sum_j \xi_i^\nu \xi_j^\nu - \overline{h_i^1}\right)^2} \\ &= N^{-2}\sum_{\nu>1}\sum_j (\xi_i^\nu \xi_j^\nu)^2 \\ &= \frac{(p-1)N}{N^2} = \alpha . \end{aligned} \tag{3.4}$$

Da, wegen ihrer Konstruktion, die $\delta S_j(0)$ nicht mit den Mustern, und damit auch nicht mit den Kopplungen J_{ij} korreliert sind, ist der zweite Term in Gleichung (3.3), u, ebenfalls eine gaußverteilte Zufallsvariable. Damit wird das gesamte interne Feld zum ersten Zeitschritt gaußisch, wie behauptet.

Die Größe u ist eine Zufallsvariable bezüglich der gelernten Muster und der Anfangsbedingungen. Mitteln wir über letztere, so erhalten wir

$$\overline{u} = \sum_j J_{ij}\,\overline{\delta S_j(0)} = 0$$

und

$$\begin{aligned} \overline{\Delta u^2} &= \sum_{jk} J_{ij}J_{ik}\,\overline{\delta S_j(0)\delta S_k(0)} \\ &= \sum_{jk} J_{ij}J_{ik}\,\overline{(S_j(0)-m_0)(S_k(0)-m_0)} \\ &= \sum_{j\neq k} J_{ij}J_{ik}\{\overline{(S_j(0)\ S_k(0)} - m_0^2\} + \sum_j J_{ij}^2\{1-m_0^2\} \\ &= \sum_j J_{ij}^2\{1-m_0^2\} . \end{aligned} \tag{3.5}$$

Die Größe $J = \sum_j J_{ij}^2$ ist nun aber für Hopfield– und pseudoinverse Kopplungsmatrix selbstmittelnd, wir erhalten beispielsweise für die Hopfield–Matrix (für

den Wert dieser Größe bei der pseudoinversen Kopplungsmatrix siehe Anhang B)

$$J = \overline{N\frac{1}{N^2}\sum_{\nu}(\xi_i^\nu \xi_j^\nu)^2} = \frac{Np}{N^2} = \alpha \ .$$

Damit ergibt sich für die Varianz der Feldverteilung zum ersten Zeitschritt

$$\begin{aligned} \overline{\Delta h_i^2(0)} &= \overline{(h_i^1)^2}(m_0)^2 + \overline{\Delta u^2} \\ &= \overline{(h_i^1)^2}\, m_0^2 + J(1-m_0^2) \ , \end{aligned} \tag{3.6}$$

also für das Hopfield–Modell

$$\overline{\Delta h_i^2(0)} = \alpha m_0^2 + \alpha(1-m_0^2) = \alpha \ .$$

Daraus folgt für den Musterüberlapp zum zweiten Zeitschritt die Gleichung

$$m(1) = \int \frac{dy}{\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big) \tanh\beta(m(0) + \sqrt{\alpha}y) \ .$$

Die Feldverteilung für die pseudoinverse Kopplungsmatrix wird in Abschnitt 4.1.4 ausführlich diskutiert werden.

Nichtgaußsches Feld beim zweiten Zeitschritt

Beim zweiten Zeitschritt finden wir bereits eine Rückkopplung der Neuronen auf sich selber. Jedes Neuron S_i polarisiert seine Umgebung, die Neuronen S_j, über die Kopplungen J_{ji} an deren interne Felder h_j. Die Neuronen S_j wirken aber andererseits über die Kopplung J_{ij}, im nächsten Zeitschritt, wieder auf das Neuron S_i zurück. Wie die nachfolgende Analyse zeigen wird, hat diese, wenn auch kleine Rückkopplung, einen entscheidenden Einfluß auf die Struktur des internen Feldes und damit auf die Dynamik des Netzwerkes.

Die Feldverteilung läßt sich mit Hilfe eines Arguments erhalten, welches in der Literatur als 'Cavity'–Methode bekannt ist [ME87]. Die generelle Idee ist, das Verhalten eines N–Neuronen–Systems mit einem System bestehend aus $N+1$ Neuronen zu vergleichen. Man analysiert dabei die kleinen Änderungen, denen das alte, N–Neuronen–System durch das Hinzufügen eines zusätzlichen Neurons S_0 unterworfen ist. Dabei werden diese Einflüße auch für das Neuron S_0 selbstkonsistent berücksichtigt.

In unserem Falle ändert das Hinzufügen des Neurons S_0, zusammen mit seinen Kopplungen J_{0i} und J_{i0}, die internen Felder der übrigen Neuronen zum Zeitpunkt $t=0$ in

$$h_i(0) \rightarrow \widetilde{h}_i(0) = h_i(0) + J_{i0}S_0(0) \ .$$

Im nächsten Zeitschritt haben wird dann für das Neuron S_0 das interne Feld

$$\begin{aligned} h_0(1) &= \sum_j J_{0j}\mathrm{sign}(\widetilde{h}_j(0)) \\ &= \sum_j J_{0j}\mathrm{sign}(h_j(0) + J_{j0}S_0(0)) \ . \end{aligned}$$

Wir entwickeln nun das Argument der Vorzeichenfunktion nach der kleinen Größe $J_{i0}S_0(0)$ und mitteln über die Feldverteilung $P_0(h)$ des ursprünglichen N–Neuronensystems.

Mit

$$\mathrm{sign}(h+\epsilon) = \mathrm{sign}(h) + 2\delta(h)\epsilon + \mathcal{O}(\epsilon^2)$$

erhalten wir für das interne Feld zum zweiten Zeitschritt die Gleichung

$$h_0(1) = \sum_j J_{0j} S_j(1) + 2P_{h_i}(0) \sum_j J_{0j} J_{j0} \cdot S_0(0) \; . \tag{3.7}$$

Der rechte Term ist die erwähnte Rückwirkung des Neurons S_0 auf sich selbst.

Der linke Term von (3.7) ist wieder eine gaußverteilte Zufallsvariable. Zerlegen wir wieder $S_j(1)$ in einen Anteil in Richtung des kondensierten Musters $\xi_i^1 = 1$ und einen dazu orthogonalen Anteil,

$$S_j(1) = m(1) + \delta S_j(1) \; ,$$

so erhalten wir für den Term $\sum_j J_{0j} m(1)$ im Falle der pseudoinversen Kopplungsmatrix $(1-\alpha)m(1)$ und im Falle des Hopfield–Modells wieder einen gaußverteilten Rauschterm. Im Term $\sum_j J_{0j}\delta S_j(1)$ sind die Anteile J_{0j} und $\delta S_j(1)$ nur schwach über die gelernten Muster am Platze "0" korreliert, sodaß auch für diesen Term eine Gaußverteilung zu erwarten ist. Dies kann für das Hopfield–Modell und das Netzwerk mit pseudoinverser Kopplungsmatrix durch eine genauere Analyse auch bestätigt werden ([GA87] und Abschnitt 4.1.4).

Alles in allem ergibt sich für die Struktur des internen Feldes zum zweiten Zeitschritt die Form

$$h_1 = w_1 + d_1 \cdot S_0 \; ,$$

also eine Summe aus einem gaußschen Rauschterm w_1 und einer diskret verteilten Zufallsvariablen S_0. Damit ist das interne Feld im zweiten Zeitschritt bereits deutlich *nichtgaußisch*. Für das Hopfield–Modell findet man zum zweiten Zeitschritt [GA87]

$$\begin{aligned} \overline{w_1} &= m_1 \; , \\ \overline{\Delta w_1^2} &= \alpha + \frac{2}{\pi} \exp\left(-\frac{m(0)^2}{\alpha}\right) + 2m(0)m(1)\sqrt{\frac{2\alpha}{\pi}} \exp\left(-\frac{m(0)^2}{2\alpha}\right) \; , \end{aligned}$$

und der Vorfaktor von S_0 hat die Größe

$$d_1 = \sqrt{\frac{2\alpha}{\pi}} \exp\left(-\frac{m(0)^2}{2\alpha}\right) \; .$$

Für die Feldverteilungen der Pseudoinverse wird wieder auf Abschnitt 4.1.4 verwiesen.

Nebenbei sei bemerkt, daß d_1 für beliebige Kopplungsmatrizen durch

$$d_1 = 2P_{h_i}(0) \sum_j J_{ij} J_{ji}$$

gegeben ist, diese Größe also proportional zur Symmetrie der Kopplungsmatrix, definiert durch

$$\eta = \frac{\sum_j J_{ij} J_{ji}}{\sum_j J_{ij}^2}$$

wird. Für den Fall verschwindender Symmetrie, also $\eta = 0$, haben wir zum zweiten Zeitschritt eine dem ersten Zeitschritt äquivalente Situation vorliegen: das interne Feld besteht nur aus einer gaußverteilten Zufallsvariablen und die Rückkopplung $d_1 \cdot S_0$ fehlt. Korrelationen bauen sich damit erst zu späteren Zeitpunkten auf. Dies kann zu einem verbesserten Netzwerkverhalten für Kopplungsmatrizen mit niedrigerer Symmetrie η führen.

3.3 Approximative Beschreibungen der Dynamik Neuronaler Netzwerke

Eine Analyse der Dynamik für längere Zeiten ist schwierig, da sich, analog zum zweiten Zeitschritt, in späteren Zeiten immer höhere Korrelationen aufbauen (siehe z.B. [GA87]). Es liegt deshalb nahe, ähnlich wie bei den statischen Untersuchungen approximative Beschreibungen der Dynamik zu verwenden.

Wie dort kann man auch bei der dynamischen Analyse Näherungen machen, die die Untersuchung des Netzwerkverhaltens vereinfachen, aber auch hier müssen die Näherungen gegebenenfalls auf ihre Validität überprüft werden. Die hier vorgestellten Näherungen sind in der Literatur als exakte Beschreibungen der Dynamik von Neuronaler Netzwerke mit einfacher Topologie eingeführt worden, wir bezeichnen sie deshalb mit den Namen der entsprechenden Topologien.

3.3.1 extrem verdünnte Netzwerke

Als die einfachste Approximation zur Dynamik vollvernetzter Netzwerke läßt sich die Dynamik extrem verdünnter Netzwerke [DE87] interpretieren. Bei extrem verdünnten Netzwerken fehlen, wegen der Verdünnung, die Korrelationen zwischen den Neuronen (bzw. den internen Feldern) an verschiedenen Plätzen. Dies gilt allerdings nur dann, falls die Verdünnung asymmetrisch erfolgt, sodaß Schleifen der Form $J_{ij}J_{ji}$ nicht auftreten.

Es sei angemerkt, daß die fehlende Korrelation zwischen den Neuronen bzw. die dazu notwendige, extreme Verdünnung die praktische Anwendung der verdünnten Netzwerke als auch deren Übertragbarkeit auf biologische Netzwerke sehr einzuschränkt. So ist bislang kein Versuch gemacht worden, extrem verdünnte Netzwerke numerisch zu simulieren, was auch angesichts der dafür notwendigen Systemstruktur wohl kaum möglich sein wird. Andererseits weiß man aus neurophysiologischen Untersuchungen, daß Konzeptbildung in biologischen Nervennetzen mit der Synchronisation von Neuronen einhergeht [GR89], also gerade die Korrelation zwischen Neuronen in biologischen Netzen wichtig ist.

Wie auch immer, die fehlende Korrelation zwischen den Neuronen in verdünnten Netzwerken ermöglicht die Anwendung des zentralen Grenzwertsatzes — es treten nur noch gaußverteilte Felder auf. Wie die obige Diskussion zeigt, ist dies für vollvernetzte Neuronale Netze im ersten Zeitschritt noch der Fall, wird jedoch ab dem zweiten Zeitschritt falsch.

Die Berechnung des Mittelwertes und der Varianz der internen Felder erfolgt analog Abschnitt 3.2.2, die dynamische Gleichung der verdünnten Netzwerke ist damit, im Falle der Hopfield–Kopplungsmatrix, durch

$$m(t+1) = \int \frac{dy}{\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big) \tanh\beta\big(m(0) + \sqrt{\alpha}y\big)$$

gegeben. Ein Vergleich mit den Formeln der statischen Analyse zeigt, daß die Iterationsformel der verdünnten Netzwerke in den Fixpunkten äquivalent der Formeln der statischen Signal/Rauschanalyse ist, mit denselben falschen Ergebnissen. So wird bezüglich des Musterüberlappes ein 2.Ordnungsphasenübergang gesehen, während das Hopfield–Modell eigentlich einen starken 1.Ordnungsphasenübergang besitzt (vergl. Abbildung 3.1 auf Seite 50).

3.3.2 Feedforward–Netzwerke

Eine andere Netzwerkstruktur, für die eine analytische Lösung der Dynamik existiert, sind die sogenannten Feedforward–Netzwerke (FF–Netzwerke,[DO89]).

In ihnen laufen die synaptischen Kopplungen nur in Vorwärtsrichtung, dies ist eine Topologie, die auch in vielen biologischen Strukturen gefunden wird. Die FF–Netzwerke sind also für sich gesehen schon interessant, sollen aber hier lediglich als eine verbesserte Approximation zur Dynamik vollvernetzter Netzwerke verwendet werden. Eine Schicht der FF–Netzwerke entspricht dabei einem bestimmten Zeitpunkt im vollvernetzten Modell. Wir werden deshalb in diesem Abschnitt die Begriffe 'Schicht' und 'Zeitpunkt' synonym verwenden.

In jeder neuen Schicht der FF–Netzwerke werden die Bitmuster neu gewählt, was sich in einer modifizierten, zeitabhängigen Kopplungsmatrix niederschlägt:

$$J_{ij}^{t+1} = N^{-1} \sum_{\nu} \xi_{i;\nu}^{t+1} \xi_{j;\nu}^{t} \ .$$

Hierbei zeigt der Index t bzw. $t+1$ die jeweilige Schicht an. Da dies in der dynamischen Interpretation der Zeitindex ist, bedeutet dies, daß in jedem Zeitschritt die Muster neu bestimmt werden.

Die Unordnung bezüglich der Muster ist also während der Dynamik nicht eingefroren und die analytische Behandlung vereinfacht sich dadurch erheblich. So hängt der Zustand des i–ten Neurons zum Zeitpunkt t,

$$S_i^t = \mathrm{sign}(h_i^{t-1}) = \mathrm{sign}\Big(\sum_{j,\nu} \xi_{i;\nu}^{t-1} \xi_{j;\nu}^{t-2} S_j^{t-2} \Big) \ ,$$

nur von Mustern $\xi_{i;\nu}^{\tau}$ ab, für die $\tau < t$ gilt. Deshalb sind z.B. die unkondensierten Musterüberlappungen

$$m_\mu^t = N^{-1} \sum_i \xi_{i;\mu}^t S_i^t$$

untereinander unkorreliert, was wir im folgenden mehrfach ausnutzen werden. Dies gilt selbstverständlich nicht für vollvernetzte Neuronale Netze.

Uns interessiert die Projektion des internen Feldes h_i^t auf die Muster $\xi_{i;1}^t$, welche als einzige einen makroskopischen Musterüberlapp besitzen sollen. Über die verbleibenden, nichtkondensierten Muster wird wie üblich gemittelt. Setzen wir wieder durch Umeichen $\xi_{i;1}^t = 1$, erhalten wir für die Projektion des internen Feldes

$$\begin{aligned} h_i^{t+1} &= \sum_j J_{ij}^{t+1} S_j^t \\ &= N^{-1} \sum_{j,\nu} \xi_{i;\nu}^{t+1} \xi_{i;\nu}^t S_j^t \\ &= m_1^t + \sum_{\nu>1} \xi_{i;\nu}^{t+1} m_\nu^t \ . \end{aligned}$$

Nun sind aber die $\xi_{i;\nu}^{t+1}$ unkorreliert zueinander und zu den Musterüberlappungen m_ν^t der vorherigen Schicht, sodaß sich nach dem zentralen Grenzwertsatz ein gaußverteiltes internes Feld ergibt. Da der Mittelwert der rechten Summe verschwindet, erhalten wir für das interne Feld, gemittelt über Anfangsbedingungen und unkondensierte Muster,

$$\overline{h_i^{t+1}} = m_1^t \ .$$

Die Bestimmung der Varianz des internen Feldes erfordert eine etwas längere Rechnung. Wir führen zunächst die Abkürzung $(\Delta^{t+1})^2 := \overline{\Delta(h_i^{t+1})^2}$ ein und

erhalten

$$\begin{aligned}(\Delta^{t+1})^2 &= \overline{\sum_{\mu,\nu>1} \xi_{i;\mu}^{t+1} m_\mu^t \, \xi_{i;\nu}^{t+1} m_\nu^t} \\ &= \sum_{\nu>1} \overline{(m_\mu^t)^2} + \sum_{\substack{\mu,\nu>1 \\ \mu\neq\nu}} \overline{\xi_{i;\mu}^{t+1} m_\mu^t \, \xi_{i;\nu}^{t+1} m_\nu^t} \\ &= \alpha N \overline{(m_\mu^t)^2} \; . \end{aligned} \tag{3.8}$$

Hierbei wurde berücksichtigt, daß $\xi_{i;\mu}^{t+1}$ und $\xi_{i;\nu}^{t+1}$ für $\mu \neq \nu$ unabhängige Zufallsvariable sind, und anschließend die Unkorreliertheit der unkondensierten Musterüberlappungen ausgenutzt.

Nun gilt weiter

$$\begin{aligned}\alpha N \overline{(m_\mu^t)^2} &= \alpha N^{-1} \overline{\sum_{i,j} \xi_{i;\nu}^t S_i^t \, \xi_{j;\nu}^t S_j^t} \\ &= \alpha (1 + N^{-1} \sum_{i\neq j} \overline{\xi_{i;\nu}^t S_i^t \, \xi_{j;\nu}^t S_j^t}) \; . \end{aligned}$$

Bei den verdünnten Netzwerken würde die rechte Summe der letzten Gleichung wegfallen, sodaß sich als zeitlich konstante Varianz für die internen Felder einfach $\overline{\Delta(h_i^{t+1})^2} = \alpha$ ergeben würde. Bei den FF–Netzwerkes ist dies jedoch nicht der Fall; hier haben S_i^t und S_j^t gemeinsame Vorfahren, mit dem Resultat einer schwachen Korrelation zwischen den Neuronen.

Für einen einzelnen Term in der Summe erhalten wir zunächst

$$\overline{\xi_{i;\nu}^t S_i^t \, \xi_{j;\nu}^t S_j^t} = \overline{\mathrm{sign}(\widetilde{h}_i^t + m_\nu^t) \cdot \mathrm{sign}(\widetilde{h}_j^t + m_\nu^t)}$$

mit

$$\widetilde{h}_i^t = \xi_{i;\nu}^t m_1^t + x_i$$

und

$$\widetilde{h}_j^t = \xi_{j;\nu}^t m_1^t + x_j \; .$$

Hierbei wurden neue Größen x_k eingeführt,

$$x_k = \xi_{k;\nu}^t \cdot \sum_{\mu\neq 1,\nu} \xi_{k;\mu}^t m_\mu^t \; .$$

Die x_k sind wieder voneinander unabhängige, gaußverteilte Zufallsvariable mit Mittelwert null und derselben Varianz $(\Delta^t)^2$ wie die internen Felder zum Zeitpunkt t.

Wir entwickeln nun beide sign–Funktionen nach dem kleinen Argument m_ν^t und erhalten

$$\overline{\mathrm{sign}(\widetilde{h}_i^t + m_\nu^t) \cdot \mathrm{sign}(\widetilde{h}_j^t + m_\nu^t)} \approx (m_\nu^t)^2 \overline{\delta(\widetilde{h}_i^t) \cdot \delta(\widetilde{h}_j^t)} \; .$$

Während auf der linken Seite über die x_k und m_ν^t gemittelt wird, erfolgt die Mittelung auf der rechten Seite nur über die unabhängig verteilten x_k. Mit

$$\overline{\delta(\widetilde{h}_i^t)} = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{(m_1^t)^2}{2\Delta^t}\right)$$

erhalten wir dann aus Gleichung (3.8) eine Rekursionsformel für die Varianz der internen Felder:

$$(\Delta^{t+1})^2 = \alpha + \frac{2}{\pi} \exp\left(-\frac{(m_1^t)^2}{\Delta^t}\right) \; .$$

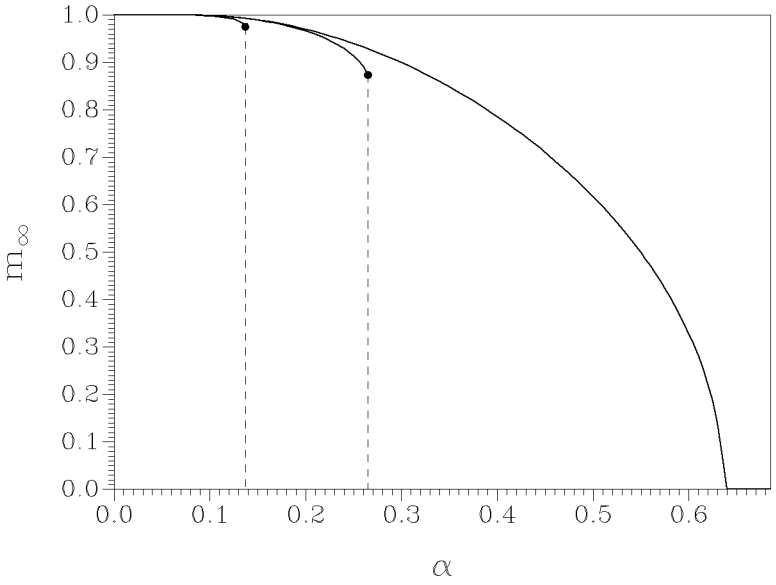

Abbildung 3.1: Ein Vergleich der stationären Lösungen vom Hopfield–Modell, FF–Netzwerken und verdünntem Modell (von links nach rechts).

Die Rekursionsgleichung für den Überlapp ergibt sich natürlich zu

$$\begin{aligned} m_1^{t+1} &= \int dh\; P_t(h)\,\mathrm{sign}(h) \\ &= \mathrm{erf}\left(\frac{m_1^t}{\sqrt{2}\Delta^t}\right) . \end{aligned}$$

Es ist interessant, sich die stationären Lösungen von verdünntem Netzwerk und FF–Netzwerk als Funktion der Speicherdichte α anzusehen. Beide stimmen mit der statischen Lösung des Hopfield–Modells für kleine α überein, liefern aber nicht das korrekte kritische α_c. Lediglich das FF–Netzwerk zeigt den diskontinuierlichen Phasenübergang des vollvernetzten Hopfield–Modells, allerdings für ein fast doppelt so großes α_c.

3.4 Das Verschwinden der 3er–Mischzustände

Wir verwenden nun die approximative Beschreibung der Dynamik der vollvernetzten Neuronalen Netze durch die FF–Netzwerke zur Untersuchung des dynamischen Verschwindens der 3er–Mischzustände. Dies wird mit numerischen Simulationen des vollvernetzten Modells verglichen.

Falls mehrere Muster makroskopischen Überlapp haben, sind die oben abgeleiteten dynamischen Gleichungen der FF–Netzwerke zu modifizieren. Die kondensierten Muster sorgen jetzt in jeder Schicht für eine Einteilung in Untergitter–Klassen, wobei jede Untergitter–Klasse eine andere Feldverteilung besitzt. Wir können also zunächst nur die Untergittermagnetisierungen zum nächsten Zeitschritt berechnen:

$$m_a^{t+1} = \int dh_a\; P_t^a(h_a)\,\mathrm{sign}(h_a) \; . \tag{3.9}$$

Hierbei ist h_a das interne Feld im Untergitter a. Die Musterüberlappungen folgen dann aus

$$m_\mu^{t+1} = 2^{(k-1)} \sum_a \xi_{a;\mu}^{t+1} m_a^{t+1} \ . \tag{3.10}$$

Um die Feldverteilung $P_t^a(h)$ innerhalb eines Untergitters zu berechnen, stellen wir das interne Feld wieder als Summe eines Signalterms und eines Rauschterms dar. Für das Neuron $i \in I_a$ gilt

$$\begin{aligned} h_i^{t+1} &= N^{-1} \sum_{j,\nu} \xi_{i;\nu}^{t+1} \xi_{i;\nu}^t S_j^t \\ &= \sum_{\nu<k} \xi_{ug(i);\nu}^{t+1} m_\nu^t + \sum_{\nu \geq k} \xi_{i;\nu}^{t+1} m_\nu^t \ . \end{aligned}$$

Der erste Term,

$$\sum_{\nu<k} \xi_{ug(i);\nu}^{t+1} m_\nu^t = \sum_{\nu<k} \xi_{a;\nu}^{t+1} m_\nu^t \ ,$$

hat innerhalb des Untergitters a einen festen Wert, während der zweite Term fluktuiert. Er ist eine gaußverteilte Zufallsvariable mit Mittelwert null. Wir erhalten damit zunächst

$$\overline{h_i^{t+1}} = \overline{h_{ug(i)}^{t+1}} = \sum_{\nu<k} \xi_{ug(i);\nu}^{t+1} m_\nu^t \ .$$

Zur Bestimmung der Varianz des internen Feldes ist eine analoge Rechnung wie in Abschnitt 3.3.2 durchzuführen. Man erhält zunächst wieder

$$\overline{\Delta(h_i^{t+1})^2} = \alpha \Big(1 + N^{-1} (m_\mu^t)^2 \sum_{i \neq j} \overline{\delta(\widetilde{h}_i^t) \cdot \delta(\widetilde{h}_j^t)} \Big) \ . \tag{3.11}$$

Nun gilt aber

$$\widetilde{h}_i^t = \sum_{\nu<k} \xi_{ug(i);\nu}^{t+1} m_\nu^t + x_i$$

und

$$\widetilde{h}_j^t = \sum_{\nu<k} \xi_{ug(j);\nu}^{t+1} m_\nu^t + x_j \ ,$$

wir haben also die unterschiedlichen Untergitter zu berücksichtigen. Zunächst erhalten wir

$$\begin{aligned} \overline{\delta(\widetilde{h}_i^t)} &= \sqrt{\frac{2}{\pi}} \exp\left(- \frac{(\sum_{\nu<\mu} \xi_{ug(i);\nu}^{t+1} m_\nu^t)^2}{2\Delta^t} \right) \\ &\equiv D_{ug(i)} \ . \end{aligned}$$

Die Summe $\sum_{i \neq j}$ in Gleichung (3.11) kann jetzt umgeschrieben werden:

$$\begin{aligned} N^{-2} \sum_{i \neq j} D_{ug(i)} D_{ug(j)} &= N^{-2} \left(\frac{N}{2^{(k-1)}} \right)^2 \sum_{i \neq j} D_{ug(i)} D_{ug(j)} \\ &= 2^{-2(k-1)} \sum_{a,b} D_a D_b \\ &= 2^{-2(k-1)} \left(\sum_a D_a \right)^2 \ . \end{aligned}$$

Damit erhalten wir als Rekursionsgleichung für die Varianz der internen Felder

$$(\Delta^{t+1})^2 = \alpha + 2^{-2(k-1)} \cdot \frac{2}{\pi}\Big\{ \sum_a \exp\Big(- \frac{(\sum_{\nu<\mu} \xi^t_{a;\nu} m^{t-1}_\nu)^2}{\Delta^t} \Big) \Big\}^2 . \qquad (3.12)$$

Die Gleichungen (3.9), (3.10) und (3.12) stellen die dynamischen Evolutionsgleichungen des FF–Netzwerkes im Falle mehrerer kondensierter Muster dar [Me88]. Wir benutzen sie nun im folgenden, um das Verschwinden der 3er–Mischzustände beim Erhöhen der Speicherdichte α zu untersuchen. Parallel dazu werden die Resultate numerischer Simulationen des *vollvernetzten* Neuronalen Netzes mit pseudoinverser Kopplungsmatrix präsentiert.

Für die Untersuchung der Wechselwirkung zwischen Mischzuständen und Mustern bei der Erhöhung der Speicherdichte α bietet sich wieder eine der Facetten des Überlapp–Dodekaeders an, da hier die Attraktionsgebiete von jeweils zwei Mustern und zwei 3er–Mischzuständen aneinandergrenzen. Die Facetten sind dabei identisch mit den Begrenzungsflächen des vierdimensionalen Hyperkubus der Untergitter–Magnetisierungen (vergl. Abbildung 2.5 auf Seite 26).

Betrachten wir zunächst den Limes $\alpha \to 0$ (Abbildung 3.2). Hier haben wir die schon aus dem zweiten Kapitel bekannte Situation vorliegen. Sowohl Muster als auch 3er–Mischzustände haben große Attraktionsgebiete, welche die Dodekaeder–Facette vollständig ausfüllen. Analytik (links) und numerische Simulation (rechts) stimmen vollständig überein.

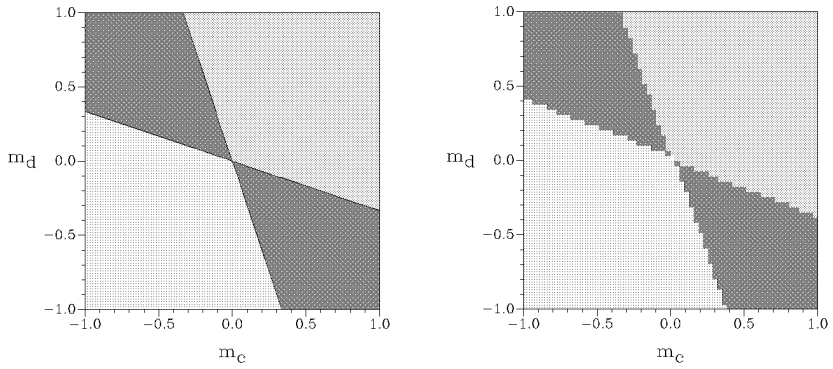

Abbildung 3.2: Attraktionsgebiete der Muster und der 3er–Mischzustände für $\alpha \to 0$. Links das FF–Netzwerk bei $\alpha = 10^{-6}$, rechts die numerische Simulation des vollvernetzten Pseudoinversen–Netzwerkes mit 256 Neuronen bei $\alpha = 0.012$ (3 Bilder).

Erhöhen wir jetzt α, (Abbildung 3.3), tritt der bemerkenswerte Effekt auf, daß die Muster ihre Attraktionsgebiete auf Kosten der Mischzustände ausdehnen! Da mit steigender Speicherdichte α die Spinglas–Phase immer dominanter wird, hätte eigentlich erwartet werden können, daß die Spinglas–Phase die durch die reduzierten Attraktionsgebiete der 3er–Mischzustände freiwerdenden Bereiche auffüllt. Dies ist jedoch nicht der Fall, wie auch die numerische Simulation des vollvernetzten Modells deutlich zeigt. Zwar laufen bei der numerischen Simulation einige der Startzustände (die weißen Flächen in Abbilung 3.3, rechts) weder in ein Muster noch in einen der 3er–Mischzustände, doch bleibt der wesentliche Effekt die Ausdehnung der Attraktionsgebiete der Muster.

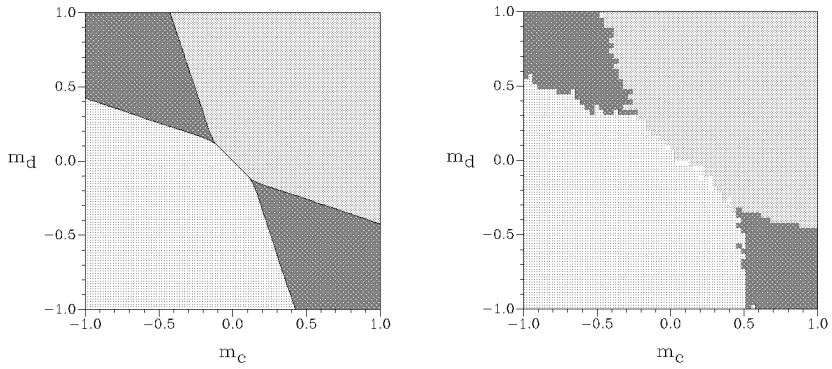

Abbildung 3.3: Die Muster vergrößern ihr Attraktionsgebiet auf Kosten der 3er–Mischzustände. Links ist das FF–Netzwerk bei $\alpha = 0.01$, rechts die numerische Simulation des vollvernetzten Pseudoinversen–Netzwerkes mit 256 Neuronen bei $\alpha = 0.059$ (15 Bilder) zu sehen. Die weißen Flächen im rechten Diagramm gehören zu Startzuständen, die weder in ein Muster noch in einen der 3er–Mischzustände relaxierten.

Erst bei weiterer Erhöhung von α schiebt sich ein Teil der Spinglas–Phase zwischen 3er–Mischzustand und Musterattraktionsgebiet (Abbildung 3.4, links). Dieser schmale, in der Abbildung weiße Bereich ist auch in der numerischen Simulation (Abbildung 3.4, rechts) ansatzweise zu sehen.

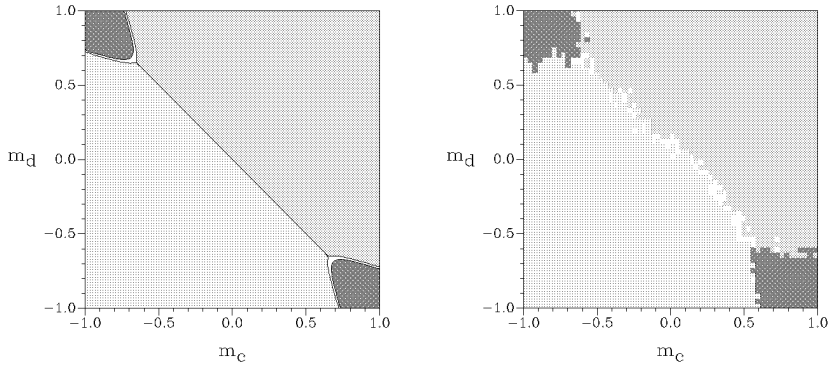

Abbildung 3.4: Die Situation kurz vor dem Verschwinden der 3er–Mischzustände. Links sind die Daten für das FF–Netzwerk bei $\alpha = 0.05954$, rechts die Daten aus der numerischen Simulation des vollvernetzten Pseudoinversen–Netzwerkes mit 256 Neuronen bei $\alpha = 0.117$ (30 Bilder) dargestellt.

In Abbildung 3.4, links, operiert das FF–Netzwerk unmittelbar vor dem α–Wert, $\alpha = 0.05954$, bei dem die 3er–Mischzustände destabilisiert werden. Die folgende Abbildung, 3.5, links, zeigt die Situation bei einem nur unwesentlich höheren α–Wert, $\alpha = 0.06$. Die 3er–Mischzustände sind plötzlich verschwunden, ihre Attraktionsgebiete werden jetzt vollständig vom Spinglas–Zustand übernommen. In den numerischen Simulationen ist dieses plötzliche Verschwinden der 3er–Mischzustände mit ihren Attraktionsgebieten nicht gut zu sehen. Allenfalls kann man es in Abbildung 3.6, rechts, beim 3er–Mischzustand, der bei den Koordinaten $m_c \approx 1$, $m_d \approx -1$ liegt, erahnen.

Es ist an dieser Stelle interessant, die Relaxation des Systems etwas genauer zu studieren. Von einem Startpunkt an der Grenze zum Attraktionsgebiet eines Musters (Abbildung 3.5, links) relaxiert das System zuerst zum virtuell noch vorhandenen 3er–Mischzustand , läuft dann jedoch von ihm weg, um schließlich im Muster zu landen. Eine genauere Analyse ermöglicht Abbildung 3.5, rechts. Hier ist der zeitliche Verlauf aller dynamischen Parameter für diese Trajektorie zu sehen. Die oberste Graphik zeigt den zeitlichen Verlauf der Rauschamplitude, $q_t = \alpha^{-1}\Delta_t$, die anderen Graphiken das zeitliche Verhalten der Untergitter–Magnetisierungen. Deutlich zu erkennen ist, daß das System bis etwa $t = 20$ zum 3er–Mischzustand relaxiert, während die Rauschamplitude langsam ansteigt. Dies führt im weiteren Verlauf der Relaxation zur Destabilisierung des Mischzustandes und zur Drift in Richtung des Ursprungs. Die Rauschamplitude steigt dann immer weiter an, bis die Untergitter–Magnetisierung m_d bei etwa $t = 80$ das Vorzeichen wechselt. Ab dann gewinnen die Muster und das System re[illegible]ergitt

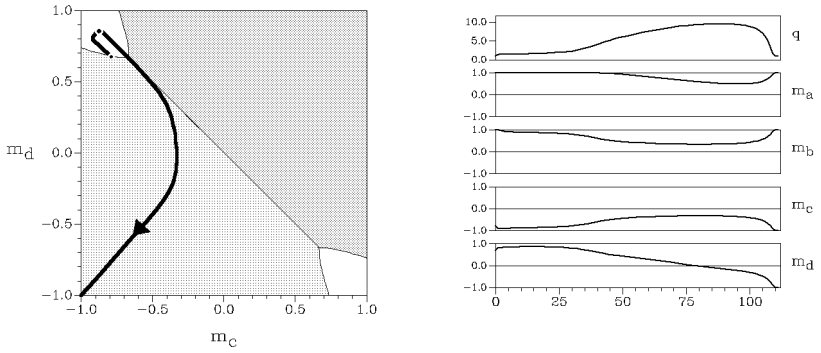

Abbildung 3.5: Das FF–Netzwerk bei $\alpha = 0.06$ (links). Eingezeichnet ist die Trajektorie von einem Startzustand an der Grenze des Attraktionsgebietes eines der Muster. Die rechte Abbildung zeigt den zeitlichen Verlauf aller dynamischen Parameter für diese ausgesuchte Trajektorie.

Wie man sieht, werden die Grenzen der Attraktionsgebiete also durch recht komplexe dynamische Prozesse bestimmt. Für einen benachbarten, aber knapp außerhalb des Attraktionsgebietes der Muster liegenden Startpunkt erfolgt die Relaxation bis zum Zeitpunkt $t \approx 80$ fast identisch. Dann aber ist das Rauschen zu groß geworden und die Untergitter–Magnetisierung m_d bleibt auf null. In der Folge relaxieren auch die übrigen Untergitter–Magnetisierungen gegen null, während die Rauschamplitude ihren Maximalwert $q = 1 + 2/(\alpha\pi)$ annimmt.

Kehren wir zurück zur Betrachtung der Attraktionsgebiete. Bei weiterer Erhöhung der Speicherdichte schieben sich die Muster fast vollständig in das ursprüngliche Attraktionsgebiet der Mischzustände. Es gibt ein optimales α, bei dem die Attraktionsgebiete der Muster maximal werden; es liegt für die FF–Netzwerke bei etwa 0.07 (Abbildung 3.6). Der gleiche Effekt ist auch in der numerischen Simulation des vollvernetzten Modells zu sehen, hier liegt das optimale α etwa bei 0.16.

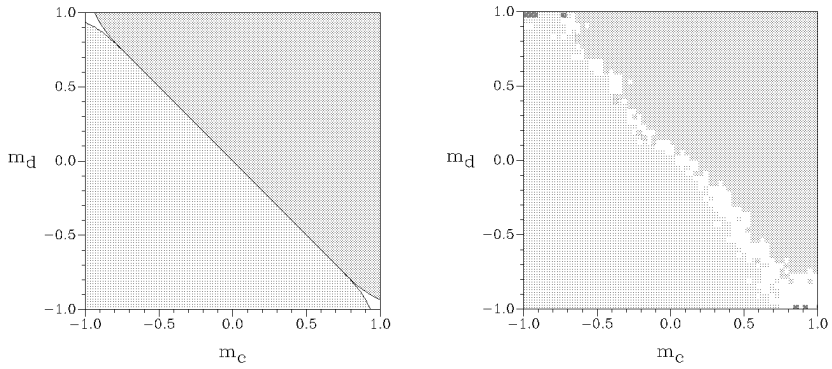

Abbildung 3.6: Das FF–Netzwerk bei $\alpha = 0.07$ (links) zeigt ein maximales Attraktionsgebiet für die Muster. Auch in der numerischen Simulation (256 Neuronen bei $\alpha = 0.156$ (40 Bilder)) wird ein solcher Effekt gesehen.

Bei weiterer Erhöhung der Speicherdichte überwiegt schließlich auch gegenüber den Mustern die Spinglas–Phase. Das Attraktionsgebiet der Muster schrumpft zusammen (Abbildung 3.7), bis schließlich, oberhalb von α_c, auch die Muster nicht mehr dynamisch stabil sind. Die numerischen Simulationen des vollvernetzten Modells zeigen das gleiche Verhalten.

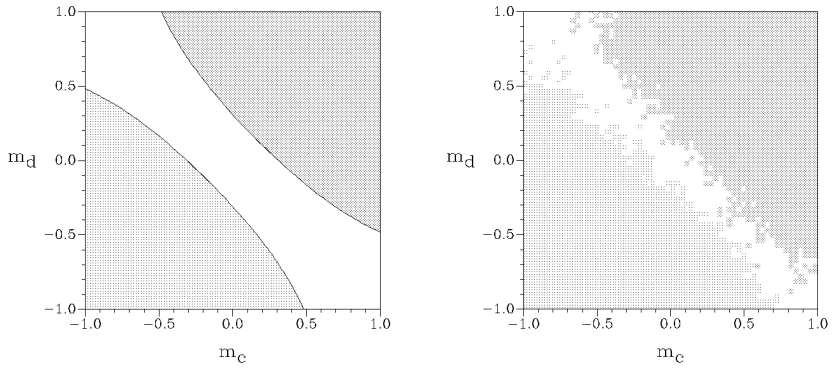

Abbildung 3.7: Bei $\alpha = 0.2$ sind die Attraktionsgebiete der Muster im FF–Netzwerk schon erheblich geschrumpft (links). Die numerische Simulation zeigt die Pseudoinverse bei $\alpha = 0.234$ (60 Bilder)).