Kapitel 5: Zusammenfassung und Diskussion

Es war Ziel der vorliegenden Arbeit, einen Einblick in die Dynamik Autoassoziativer Neuronaler Netzwerke zu geben. Bei diesem Typ von Netzwerken steht praktisch jedes Neuron mit allen anderen Neuronen des Netzwerkes in Wechselwirkung; verbunden mit dem nichtlinearen Übertragungsverhalten der einzelnen Neuronen ergibt sich damit für das gesamte Netzwerk ein komplexes dynamisches Verhalten während des Assoziationsvorgangs.

Die Arbeit ist in zwei Bereiche aufgeteilt. Der erste Teil untersucht den Zustandsraum Neuronaler Netzwerke für die Grenzfälle $\alpha \to 0$ und $\alpha > 0$. Es wird das Entstehen metastabiler Niveaus und die Auswirkung dieser dynamischer Fallen auf die Dynamik des Neuronalen Netzwerkes untersucht. Im zweiten Teil der Arbeit wird dann, ausgehend von einer exakten dynamischen Theorie, eine approximative Beschreibung der Dynamik Neuronaler Netzwerke entwickelt, mit der es erstmals möglich wird, auch Aussagen für das Verhalten Neuronaler Netze über längere Zeiträume zu erhalten.

Netzwerke bei niedriger Speicherdichte

Selbst im Grenzfall verschwindender Speicherdichte, also $\alpha \to 0$, zeigt sich eine komplexe Struktur des Zustandsraumes Neuronaler Netzwerke aus Attraktionsgebieten von gelernten Mustern und von nichtgelernten, aber sehr zahlreich auftretenden Mischzuständen. Die genaue Form der Attraktionsgebiete stellt sich als dynamikabhängig heraus. Geeignete Dynamiken können zu einer partiellen Vergrößerung der Attraktionsgebiete der Muster führen, lösen aber prinzipiell nicht das Problem der Mischzustände, die als dynamische Fallen einen erheblichen Einfluß auf die Dynamik Neuronaler Netzwerke haben.

Exakte Enumeration metastabiler Niveaus

Für niedrige Musterzahlen ist es möglich, durch Anwendung entsprechender Analysetechniken, alle metastabile Niveaus oder Mischzustände des Netzwerkes zu finden und aufzulisten. Die Zahl der nichtgelernten, aber trotzdem eingespeicherten Muster wächst sehr stark mit der Erhöhung der Zahl der gelernten Muster. Sind bei drei gelernten Muster lediglich 8 zusätzliche, nichtgelernte Muster im Netzwerk gespeichert, findet man bei 5 Mustern bereits 1282 metastabile Niveaus. Bei 7 Mustern sind dann schon 3 548 344 zusätzliche, nicht gelernte dynamische Attraktoren im Zustandsraum des Neuronalen Netzwerkes zu finden!

Die metastabilen Niveaus haben makroskopische Überlappungen mit den Mustern und sind deshalb nicht identisch mit der für $\alpha > 0$ in Erscheinung tretenden Spinglas–Phase. Lediglich bestimmte, hochsymmetrische Mischzustände scheinen für $\alpha > 0$ zur Spinglas–Phase zu degenerieren.

Optimale Kopplungsmatrix

Die große Zahl der unerwünschten Attraktoren schränkt das Attraktionsgebiet der wenigen gelernten Muster erheblich ein. Es ist aber möglich, eine optimale Kopplungsmatrix zu konstruieren, die zur Destabilisierung der Mischzustände und damit zu optimalen Attraktionsgebieten der gelernten Muster führt. Dabei wird ausgenutzt, daß Mischzustände immer Komponenten im Orthogonalraum der Muster besitzen, welche zu einer gezielten Reduzierung der Stabilität dieser unerwünschten dynamischen Attraktoren herangezogen werden können.

Interessanterweise werden bei der optimalen Kopplungsmatrix auch die Stabilitäten der gelernten Muster verringert. Daß dabei gleichzeitig deren Attraktionsgebiete vergrößert werden, steht im Widerspruch zu der in [GA88, KR88] geäußerten Vermutung, kleinere Stabilitäten würden auch reduzierte Attraktionsgebiete der Muster implizieren. Möglicherweise stimmt diese Vermutung nur im Limes hoher Speicherdichte, bei dem durch die Muster induziertes Rauschen zu einer Destabilisierung von Systemzuständen mit niedriger Stabilität führt.

Netzwerke bei hoher Speicherdichte

Metastabile Zustände haben immer kleinere interne Felder als die gelernten Muster, was im Limes $\alpha > 0$, wenn also die Zahl der gelernten Muster proportional zur Zahl der Neuronen erhöht wird, zur sukzessiven Destabilisierung dieser dynamischen Fallen durch das musterinduzierte Rauschen führt. Gleichzeitig entwickelt sich beim Erhöhen der Speicherdichte eine Spinglas–Phase, die für genügend hohe α–Werte das globale Minimum der Freien Energie wird. Trotzdem werden die Attraktionsgebiete der destabilisierten Mischzustände zunächst nicht von der Spinglas–Phase, sondern von den gelernten Mustern in Anspruch genommen. Damit bietet sich die Möglichkeit, durch eine geschickt gewählte Speicherdichte α, die Attraktionsgebiete der Muster, und damit das Verhalten des gesamten Neuronalen Netzes zu optimieren.

Eine Analyse der Dynamik nahe der Grenze der Attraktionsgebiete zeigt, daß dies ein komplizierter dynamischer Prozess ist, bei dem alle kondensierten Musterüberlappungen und die Amplitude des musterinduzierten Rauschen eine Rolle spielen; die Bestimmung der Grenzen von Attraktionsgebieten wird damit höchst nichttrivial.

Statische und dynamische Analysen, Approximationen

Die Komplexität des Zustandsraumes Autoassoziativer Netzwerke spiegelt sich auch im komplexem dynamischen Verhalten der Netzwerke wieder. Bislang wurde aber weitgehend, von [GA87] einmal abgesehen, wo die Dynamik des Hopfield–Modells zwei Zeitschritte weit exakt berechnet wurde, das Verhalten Neuronaler Netze mit Hilfe von Gleichgewichtstheorien der statistischen Physik behandelt [AM85, KA87]. Bei einer solchen Analyse wird natürlich der eigentliche Assoziationsvorgang des Neuronalen Netzes explizit ausgeklammert. So treten die dynamischen Fallen, die im niedrigen α–Bereich das Netzwerk kurz vor dem Erreichen des Musterattraktors stoppen können (vergl. Abbildung 4.5 auf Seite 68) in den Gleichgewichtstheorien nicht auf. Deswegen ist es auch zweifelhaft, ob von Gleichgewichtstheorien abgeleitete Näherungen zur Dynamik [KO89, HO89] korrekte Beschreibungen des Verhaltens Neuronaler Netzwerke liefern. Andere Näherungen [AM88, PA89] vernachläßigen wichtige Korrelationen und scheinen deshalb ebenfalls wenig geeignet zur Analyse des Langzeitverhaltens Neuronaler Netzwerke.

In der vorliegenden Arbeit wird dieser Weg vermieden; vielmehr bildet die in Anhang A entwickelte, exakte Theorie der Dynamik den Ausgangspunkt für die Entwicklung der approximativen Double–Peak–Dynamik *(DPD)*, welche den wesentlichen physikalischen Grundgehalt der exakten Theorie erhält, aber gleichzeitig die Zahl der dynamischen Parameter erheblich reduziert. Mit Hilfe der DPD ist es erstmals möglich, das Verhalten Neuronaler Netze über längere Zeiträume zu studieren. Angewendet wurde die DPD in dieser Arbeit auf die synchrone Dynamik des Netzwerkes mit pseudoinverser Kopplungsmatrix, aber es dürfte ohne größere Schwierigkeiten möglich sein, sie auch auf andere Netzwerke und Spingläser zu übertragen.

Synchrone Dynamik des Netzwerkes mit pseudoinverser Kopplungsmatrix

Die Analyse der synchronen Dynamik der Pseudoinversen zeigt ein differenziertes Erscheinungsbild des dynamischen Verhaltens dieses Neuronalen Netzwerkes. Im Vergleich mit den numerischen Ergebnissen für die asynchrone Dynamik des Netzwerkes kann folgendes festgestellt werden: Im hohen α–Bereich ($\alpha > \frac{1}{2}$) zerfällt der Überlapp eines Musters im Falle des Nichterkennens bei der synchronen Dynamik schnell nach null, ohne in den dynamischen Fallen hängenzubleiben, während bei der asynchronen Dynamik, die langsamer durch die Zustandsraum fließt, die vielen Punktattraktoren einen solchen Zerfall des Musterüberlapps verhindern (Abbildungen 4.5 und 4.6). Aber beide Dynamiken liefern im hohen α–Bereich die gleichen Attraktionsgebiete, sodaß vermutet werden darf, daß hier die Stabilität der eingespeicherten Muster der wesentliche, die Attraktionsgebiete bestimmende Faktor ist. Im niedrigen α–Bereich wird hingegen das Attraktionsgebiet durch dynamische Fallen begrenzt. Das Netzwerk läuft immer in Richtung des Musters, wird aber, bei zu niedrigem Startüberlapp, kurz vor Erreichen der Musterattraktoren gestoppt. Da für $\alpha \to 0$ bei asynchroner als auch bei synchroner Dynamik nur noch die Punktattraktoren der asynchronen Dynamik eine Rolle spielen (Abbildung 4.12), folgt auch in diesem α–Bereich für beide Dynamiken ein im wesentlichen identisches Attraktionsgebiet.

Die Erweiterung der DPD auf die temperaturabhängige Dynamik erlaubt schließlich die Analyse des Systemverhaltens unter externem Rauschen. Wie erwartet, verschwindet dann der Einfluß der metastabilen Zustände auf die Dynamik des Netzwerkes, weil das System jetzt durch niedrige Energiebarrieren 'hindurchtunneln' kann. Trotzdem ist nur eine marginale Verbesserung der Assoziationsfähigkeit des Neuronalen Netzes durch externes Rauschen festzustellen. Dies steht im Gegensatz zu in der Literatur geäußerten Vermutungen [Am89, Pe89].

Die Beschreibung der temperaturabhängigen Dynamik durch die DPD liefert ferner die Möglichkeit, das thermodynamische Verhalten des Neuronalen Netzwerkes unter synchroner Dynamik zu studieren. Hier existiert bislang lediglich eine replika–symmetrische Näherung von Kanter/Sompolinsky [Ka87], welche die thermodynamisch stabilen Systemzustände unter asynchroner Dynamik untersucht. Interessanterweise ist eine weitgehende Übereinstimmung zwischen dieser statischen Analyse und den dynamischen Resultaten der DPD festzustellen. Dies ist insofern erstaunlich, da die Meanfield–Gleichung aus [Ka87] in keinem Fall in die Fixpunktgleichung der DPD umgewandelt werden kann.

Die Analyse des thermodynamischen Verhaltens eines Neuronalen Netzwerkes mit Hilfe einer approximativen dynamischen Theorie ist neu, liefert aber durchaus vernünftige Ergebnisse. Dies ist ein Indiz für die Güte der DPD, mit der es erstmals möglich wurde, das Verhalten Autoassoziativer Netzwerke auch über lange Zeiträume zu studieren.