Kapitel 6

Attraktornetzwerke

6.1 Architektur und Dynamik des Netzwerkes

6.1.1 Definition des Netzwerkes

Wir betrachten zunächst Netzwerke mit einer deterministischen Update-Regel (s.u.).

Betr. vollständig verkoppeltes Netze aus N bipolaren Modellneuronen $i = 1, \ldots, N$, d.h. Aktivitätszustand des Neurons i mit $\xi_i \in \{-1, +1\}$ codiert. Aktivitätszustand des Netzwerkes zur Zeit t durch den

$$\text{Zustandsvektor} \quad \vec{\xi}(t) = (\xi_1(t), ..., \xi_N(t)) \tag{6.1}$$

gegeben, Beispiel:

$$\text{Zustandsvektor des Netzwerkes:} \quad \vec{\xi}(t) = (-1, 1, 1, 1, \ldots, -1, 1) \tag{6.2}$$

Die Aktivität (Output) des Neurons i ist durch

$$\xi_i = sign(z_i) \tag{6.3}$$

als Funktion des postsynaptischen Potentials (PSP) z_i und damit der Inputs von allen anderen Neuronen bestimmt, wobei das PSP sich aus der gewichteten Summe der Outputs aller anderen Neuronen ergibt:

$$\begin{gathered}\text{Postsynaptisches Potential (PSP) des Neurons } i: \\ z_i = \sum_{j=1}^{N} w_{i,j}\xi_j - h_i = \vec{w}_i \cdot \vec{\xi} - h_i\end{gathered} \tag{6.4}$$

mit

$$w_{ij} \quad = \text{Synaptische Kopplungsstärke zwischen Neuron } i \text{ und } j$$

wobei i. a. $w_{i,i} = 0$ (keine Selbstkoppelung). Die Schwelle h_i kann wie gehabt durch Einführung einer Zusatzkomponente (fiktives Neuron) $\xi_0 = 1$ und $w_{i,0} = -h_i$ automatisch mitgeführt werden.

Bei den betrachteten statischen Neuronen ist aber die Regel 6.3 wegen der Rückkopplungsschleifen nicht eindeutig, die outputs können sich erst nach einer minimalen Verzögerungsszeit aus den Inputs bilden. Eine eindeutige Festlegung ergibt sich durch Einführung eines

$$\text{Zeittaktes } t = 0, 1, 2, \ldots.$$

Aktualisierung (Updaten) der neuronalen Aktivitäten entweder für alle Neuronen gleichzeitig (parallel, synchron) oder in systematischer oder stochastischer Abfolge (sequentiell, asynchron). Durch das Aktualisieren kann sich der Netzwerkzustand verändern, d. h. wir erhalten einen zeitlich veränderlichen Netzwerkzustand $\vec{\xi}(t)$, der die Dynamik des Netzwerkes repräsentiert.

Algorithmus für asynchrones "Updaten":

1.	$t = 0:$	Initialisierung: Lege Ausgangszustand $\vec{\xi}(0)$ des Netzwerkes fest.
2.	$t = t + 1:$	Wähle ein $i \in \{1, 2, \ldots, N\}$ aus. Bestimme das aktuelle (effektive) PSP $z_i = \vec{w}_i \cdot \vec{\xi} - h_i$ Aktualisiere den Zustand ξ_i dieses Neurons gemäß $\xi_i^{neu} = sign\,(z_i)$
3.		Abbruch? Sonst GOTO 2

(6.5)

Bemerkung: ξ_i^{neu} ist der Zustand des Neurons zur neuen Zeit.

Die Dynamik kann auch im Sinne einer iterierten Abbildung geschrieben werden

$$\xi_i := sign\left(\vec{w}_i \cdot \vec{\xi} - h_i\right)$$

oder

$$\xi_i := sign\left(\sum_{j \neq i} w_{ij} \cdot \xi_j - h_i\right) \tag{6.6}$$

wobei beim asynchronen Updaten das Neuron i etwa zufällig ausgewählt wird. Offensichtlich bildet diese iterierte Abbildung den Zustandsraum, d. h. die Menge $\{-1, +1\}^N$ auf sich selbst ab.

6.1.2 Formale Eigenschaften

Darstellung der Dynamik im Konfigurationsraum

Die Dynamik 6.5 definiert ein zeitdiskretes dynamisches System. Darstellung des Zustandes $\vec{\xi}(t)$ im **Konfigurations- oder Zustandsraum**, d. h. nach (6.2)

als Eckpunkt des N-dimensionalen Hyperwürfels. Eine (echte, d. h. $\xi_i^{neu} \neq \xi_i^{alt}$) Aktualisierung gemäß (**??**) bedeutet Übergang zu einem benachbarten Eckpunkt (ein Bit geflippt), Zeitverlauf $\vec{\xi}(t)$ des Zustandes durch Trajektorie, d. h. Abfolge von solchen Sprüngen gegeben.

Die Energiefunktion. Symmetrische Kopplungen.

Die Definition der Dynamik (6.5) gilt ganz allgemein für beliebige Kopplungen. Besonderheit bei symmetrischen Kopplungsstärken

$$w_{i,j} = w_{j.i}$$

ist die Exisenz einer Energiefunktion $H(t)$, die durch die Dynamik (6.5) minimiert wird. Betr.

$$H(t) = H\left[\vec{\xi}\right] = -\frac{1}{2}\sum_{i\neq j}\xi_i w_{i,j}\xi_j - \sum_{i=1}^{N}\xi_i\, h_i, \quad \text{wobei} \quad \xi_i = \xi_i(t) \tag{6.7}$$

wobei $\sum_{i\neq j}$ auf den Ausschluß der Selbstkopplungen hinweist. Unter der Dynamik (6.5) gilt

$$\frac{d}{dt}H(t) \leq 0 \tag{6.8}$$

so daß $H(t)$ monoton fallend (oder konstant) ist.

Beweis: Nehme an, Neuron k werde gerade aktualisiert, die Änderung des Netzwerkzustandes infolge der Aktualisierung ist

$$\delta\xi_k = \xi_k^{neu} - \xi_k^{alt} = sign(z_k) - \xi_k^{alt}, \quad \text{und} \quad \delta\xi_i = 0 \quad \forall i \neq k$$

Wegen $\xi_k^{neu} = sign(z_k)$ ist $\delta\xi_k$ entweder 0 oder $2\,sign(z_k)$. Für die Änderung $\Delta H = H\left[\vec{\xi}^{neu}\right] - H\left[\vec{\xi}^{alt}\right]$ von H infolge der Aktualisierung gilt folglich (verwende $w_{i,j} = w_{j,i}$)

$$\Delta H = -\sum_j \delta\xi_k w_{k,j}\xi_j - \delta\xi_k h_k = -\delta\xi_k\, z_k \leq 0 \tag{6.9}$$

da entweder $\delta\xi_k = 0$ oder $\delta\xi_k = 2\,sign(z_k)$.

Fixpunkte der Dynamik

Damit nähert sich der Zustand des Netzwerkes zumindest asymptotisch einem festen Wert an, da ja die Energie nicht beliebig klein werden kann. Der angestrebte Zustand $\vec{\xi}^{Fix}$ ist Fixpunkt der Dynamik (Endpunkt der Trajektorie), denn für ihn gilt

$$\xi_i^{Fix} = sign\left(\vec{w}_i \cdot \vec{\xi}^{Fix} - h_i\right)$$

bzw.

$$\xi_i^{Fix} = sign\left(\sum_j w_{ij} \cdot \xi_j^{Fix} - h_i\right)$$

Lage und Zahl der Fixpunkte hängen bei gegebenener Dynamik (6.5) entscheidend von den Kopplungen $\{\vec{w}_i\}$ ab (s. u.). Welcher Fixpunkt angelaufen wird bestimmt der Startpunkt $\vec{\xi}(0)$. Das macht diese Systeme zu Kandidaten für Assoziativspeicher, falls es gelingt, die Fixpunkte, die dann die Speicherzustände sein sollten, durch geeignete Wahl der Kopplungsstärken $\{w_{ij}\}$ vorzugeben (s. u.).

6.1.3 Optimierung frustrierter Systeme als Netzwerkdynamik

Das Auffinden der Fixpunkte der Dynamik entspricht der Minimierung der Energiefunktion. In diesem Sinne löst die Dynamik ein Optimierungsproblem. Bei Abbildung eines Optimierungsproblems auf eine Energiefunktion mit symmetrischen Kopplungen und mit einem binären Lösungsraum kann die obige Dynamik die Lösung liefern. Bei zufällig verteilten $w_{i,j}$ mit unterschiedlichen Vorzeichen ist das Problem des Auffindens eines optimalen $\vec{\xi}^{opt}$, das also die Energiefunktion minimiert, ein NP-vollständiges Problem (nicht in polynomialer Zeit lösbar). Grund: das System ist frustriert. Zur Diskussion betr. wir den Fall gleichstarker Koplungen $w_{i,j} = w$ mit unterschiedlichen Vorzeichen. Betr.zunächst Kopplung zwischen zwei Neuronen. Für (verwende $\Leftrightarrow$ für positive und $\rightarrow$ für negative Kopplungen, Pfeilrichtung spielt keine Rolle)

$$\oplus \Leftrightarrow \ominus \quad , \quad \text{oder} \qquad \ominus \Leftrightarrow \oplus \tag{6.10}$$

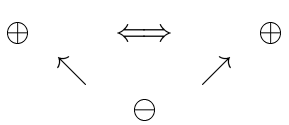
ergibt sich ein positiver Beitrag $|w|$, für

$$\oplus \Leftrightarrow \oplus \quad , \quad \text{oder} \qquad \ominus \Leftrightarrow \ominus \tag{6.11}$$

ein negativer Beitrag $-|w|$ zur Energiefunktion (6.7), d. h. bei (6.11), ist die Verbindung durch die gefundenen Konfiguration "befriedigt", weil sie zu einer Energieverringerung führt. Bei negativer Kopplung ist die befriedigende Konfiguration entsprechend

$$\oplus \rightarrow \ominus \quad , \quad \text{oder} \qquad \ominus \rightarrow \oplus$$

Bei drei Neuronen kann auch im Fall rein positiver Kopplungen eine befriedigende Lösung gefunden werden, bei gleichsinniger Aktivität aller Neuronen ergibt sich die maximale Energieminimierung $-3\,|w|$, ebenso wie bei



Sonst ist, z. B. bei rein negativen Kopplungen, nur eine von mehreren gleichberechtigten Kompromißlösungen als suboptimale Variante realisierbar, z. B.

mit Beitrag $-|w|$ zur Energiefunktion. Hier bleibt immer eine Verbindung ($\oplus \rightarrow \oplus$) unbefriedigt oder "frustriert".

Bei größerer Neuronenzahl und zufällig verteilten Vorzeichen der Kopplungen steigt der Frustrationsgrad (Zahl der frustrierten Kopplungen) schnell an, so daß die optimale Konfiguration immer nur eine von vielen etwa gleichberechtigten Kompromißlösungen ist. Die Zahl dieser Kompromißlösungen steigt mit N wie $\exp(cN)$, wobei c eine problemspezifische Konstante ist, an[1] und damit ist die systematische Suche nach dem optimalen $\vec{\xi}$ nicht in polynomialer Zeit durchführbar⇒ NP-vollständiges Problem.

Abbildung von Optimierungsproblemen auf symmetrische Attraktornetzwerke

Ein Zweiklassenproblem Betr. eine Menge von Personen $i = 1, 2, \ldots, N$, die in zwei Klassen eingeteilt werden sollen. Beziehungen zwischen den Personen (Sympathie, Antipathie) durch Wechselwirkungsterme $w_{i,j}$ bewertet. Ziel: Einteilung der Personen in zwei Klassen mit möglichst geringer Frustration, d. h. Personen mit positiver (negativer) Wechselwirkung sind möglichst in die gleiche (unterschiedliche) Klasse(n) einzuordnen. Klassenlabel ξ_i: $\xi_i = 1\,(-1)$ heißt Person $i \in$Klasse $1\,(2)$. Wenn $-\xi_i w_{i,j} \xi_j$ den Beitrag der Personen i, j zur Gesamtspannung in dem System mißt, so ist die zu minimierende objektive Funktion des Optimierungsproblems

$$H\left[\vec{\xi}\right] = -\frac{1}{2}\sum_{i \neq j} \xi_i w_{i,j} \xi_j$$

in Übereinstimmung mit (6.7). Die Dynamik (6.5) minimiert damit diese Funktion, zumindest wird ein lokales Minimum erreicht, s.u.

Das Graphenpartitionierungsproblem

Der reisende Handelsmann (travelling salesman)

6.2 Das Hopfieldnetzwerk

Wegen der Existenz von Fixpunkten mit endlichen Einzugsgebieten können Attraktornetzwerke allgemein als Assoziativspeicher verwendet werden, allerdings

[1] Es gilt $c \prec \log 2$, da die Gesamtzahl der Zustände gleich $2^N = \exp(N \log 2)$ ist.

müssen dazu die Kopplungen entsprechend eingestellt werden, so daß die Fixpunkte mit den zu speichernden Mustern übereinstimmen. Das gelingt mit folgendem Verfahren

6.2.1 Die Hebbsche Lernregel

Eine Einstellung der Kopplungen, so daß die Fixpunkte vorgegebenen Mustern

$$\vec{\xi}^{\mu}, \quad \mu = 1, 2, \ldots, P \tag{6.12}$$

entsprechen, gelingt zumindest näherungsweise mit dem Ansatz

$$w_{i,j} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^{\mu} \xi_j^{\mu} \tag{6.13}$$

(Hebb- bzw. Hopfield-Matrix).Die Vorschrift kann als Resultat der Anwendung folgender Lernregel betrachtet werden:

1. Initialisierung $w_{i,j} = 0, \quad \forall i, j$ (Tabula rasa)

2. For $\mu = 1$ to P do

For $i = 1$ to N do
For $j = 1$ to N do
$w_{i,j} = w_{i,j} + \frac{1}{N} \xi_i^{\mu} \xi_j^{\mu}$

Diese entspricht dem Hebbschen Paradigma des Lernens, wonach synaptische Verbindungen zwischen Neuronen im gleichen (ungleichen) Aktivitätszustand verstärkt (geschwächt) werden.

Ein vollständig verkoppeltes neuronales Netz mit der Dynamik (6.5) und den synaptischen Kopplungen (6.13) bildet das eigentliche Hopfieldmodell.

6.2.2 Eigenschaften des Hopfieldmodells

Vorbemerkung – Eigenschaften hochdimensionaler binärer Zufallsvektoren

Seien $\vec{\xi}, \vec{\eta}, \quad \xi_i, \eta_i \in \{-1, +1\}$ zwei Zufallsvektoren, d. h. ξ_i, η_i seien unabhängige, gleichverteilte Zufallszahlen mit Mittelwert Null,

$$\langle \xi_i \rangle = 0, \quad \langle \eta_i \rangle = 0, \quad \langle \xi_i \xi_j \rangle = \delta_{i,j}, \quad \langle \xi_i \eta_j \rangle = 0 \quad \forall i, j$$

und Streuung

$$\sigma = \sqrt{\langle \xi_i^2 \rangle} = 1$$

Das Skalarprodukt

$$\left(\vec{\xi}, \vec{\eta}\right) = \frac{1}{N} \sum_{i=1}^{N} \xi_i \eta_i = \frac{1}{N} \vec{\xi} \cdot \vec{\eta}$$

ist gleich 1 für $\vec{\xi} = \vec{\eta}$ (da dann $\xi_i \eta_i = 1$) bzw. für $\vec{\xi} \neq \vec{\eta}$ gleich einer Zufallszahl ζ mit $\langle \zeta \rangle = \left\langle \left(\vec{\xi}, \vec{\eta} \right) \right\rangle = 0$ und

$$\sqrt{\langle \zeta^2 \rangle} = \frac{1}{\sqrt{N}} \tag{6.14}$$

da $\zeta = \left(\vec{\xi}, \vec{\eta} \right) = \frac{1}{N} \sum_{i=1}^{N} \omega_i$, mit ω_i =random$\in \{-1, +1\}$. Für $N \to \infty$ genügt $\bar{\zeta} = \sqrt{N}\zeta$ einer Gaußverteilung (zentraler Grenzwertsatz) mit Streuung 1, vgl. 6.14

$$p\left(\bar{\zeta}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\zeta}^2}{2}} \tag{6.15}$$

Mit (6.14) folgt, daß zwei im obigen Sinne unabhängige binäre Zufallsvektoren für $N \to \infty$ immer orthogonal zueinander sind. Für endliche N gilt der Satz bis auf Abweichungen der Ordnung $\frac{1}{\sqrt{N}}$

$$\left(\vec{\xi}, \vec{\eta} \right) = O\left(1/\sqrt{N}\right) \tag{6.16}$$

Wichtige Kenngrößen des Netzwerkes

- Die **Überlappung** (Skalarprodukt) eines gegebenen Netzwerkzustandes $\vec{\xi}$ mit einem Muster $\vec{\xi}^{\mu}$

  $$m^{(\mu)}(t) = \left(\vec{\xi}(t), \vec{\xi}^{\mu} \right) \tag{6.17}$$

  mißt die Übereinstimmung des Netzwerkzustandes mit dem Muster. Zusammenhang mit Hammingdistanz:

  $$D\left[\vec{\xi}(t), \vec{\xi}^{\mu} \right] = \frac{N}{2} \left(1 + m^{(\mu)}(t) \right) \tag{6.18}$$

  mißt die Zahl der Bits in denen $\vec{\xi}(t)$ und $\vec{\xi}^{(\mu)}$ übereinstimmen. Damit ist $\left(1 + m^{(\mu)}(t) \right)/2$ die relative Zahl der übereinstimmenden Bits. Die gespeicherten Zustände (Speicherinhalte) entsprechen Fixpunkten $\vec{\xi}^{Fix} = \lim_{t \to \infty} \xi(t)$, ihre Überlappung mit den Mustern sei

  $$\left(\vec{\xi}^{Fix}, \vec{\xi}^{\mu} \right) = m^{(\mu)} = \lim_{t \to \infty} m^{(\mu)}(t)$$

  (Achtung: Nicht alle Fixpunkte sind Speicherinhalte, s. u.).

- Das System befindet sich in einem **idealen Speicherzustand** (retrieval state) zum Muster $\vec{\xi}^{\mu}$ falls

  $$\vec{\xi}^{Fix} = \vec{\xi}^{\mu} \tag{6.19}$$

  und damit

  $$m^{(\mu)} = \delta_{\mu, \mu^F}$$

Ideale Speicherung heißt also daß die Überlappung eines Speicherinhaltes mit genau einem der Muster gleich eins und mit allen anderen Mustern gleich Null ist. Dies wird bei hoher Beladung des Netzes nicht erreicht, $m^{(\mu)}$ definiert deshalb die sog.

- **Wiedergabequalität** (retrieval quality): Ist das System in einem retrievfal statezum Muster μ, so gilt im allgemeinen, d. h. für $m^\mu < 1$

$$\vec{\xi}^{Retr.} = \vec{\xi}^\mu + \text{"cross talk"} \quad = \quad \text{Muster} + \text{"cross talk"} \tag{6.20}$$

m^μ definiert dann die **Wiedergabequalität** und $D^\mu = (1 + m^\mu)/2$ ist die relative Zahl der Bits, in denen die Speicherinhalte und die zu speichernden Muster übereinstimmen. (s. o.) Wie wir sehen werden ist die Zahl der speicherbaren Muster proportional zur Neuronenzahl N wir definieren deshalb den

- **Load parameter** α

$$\alpha = \frac{P}{N} \tag{6.21}$$

der die Zahl der im Netzwerk gespeicherten Muster angibt.

Speicherung eines einzigen Musters (Mattismodell)

Vorbemerkung: Rechnen mit der *sign*–Funktion
Es gilt (mit $|\xi_i| = 1$)

$$\begin{gathered} sign\,(ax) = sign\,(a)\,sign\,(x)\,, \quad sign\,(a\xi_i) = \xi_i sign\,(a) \\ sign\,(x+a) = sign\,(x) \quad \text{falls} \quad |x| > |a| \end{gathered} \tag{6.22}$$

Definition des Mattismodells: Das Mattismodell entspricht der Speicherung nur eines einzelnen Musters $\vec{\xi}^1$, d.h.

$$w_{i,j} = \frac{1}{N}\xi_i^1\xi_j^1, \qquad w_{ii} = 0$$

Eigenschaften: Das Muster $\vec{\xi}^1$ ist ein Fixpunkt der Dynamik, da

$$z_i = \sum_{j\neq i} w_{ij}\xi_j^1 = \xi_i^1\left(\vec{\xi}^1, \vec{\xi}^1\right) - \frac{1}{N}\xi_i^1 = \left(1 - \frac{1}{N}\right)\xi_i^1 \tag{6.23}$$

(der $\frac{1}{N}$ Term folgt aus der wegen $i \neq j$ beschränkten Summation). Sei hier und im folgenden immer

$$\left|m^1\right| \geq |h_i| + \frac{1}{N}$$

dann folgt mit 6.22

$$\xi_i^1 = sign\left(\xi_i^1\left(\vec{\xi}^1, \vec{\xi}^1\right) - \frac{1}{N}\xi_i^1 + h_i\right) = sign\left(\xi_i^1\right) \tag{6.24}$$

Eine analoge Überlegung gilt für das Inverse $-\vec{\xi}^1$ des Musters. Am Fixpunkt ist folglich

$$\vec{\xi}^{Fix} = m^{(1)}\vec{\xi}^1, \qquad m^{(1)} = \pm 1$$

wobei $m^{(1)} = 1$ bei Speicherung des Musters und $m^{(1)} = -1$ für sein Inverses gilt.

Dynamik: Bei Start von einem beliebigen Anfangszustand $\vec{\xi}(0)$ ist man bei parallelem Updaten in einem Schritt am Fixpunkt (Muster oder sein Inverses), denn

$$\xi_i(1) = sign\left(\xi_i^1\left(\vec{\xi}^1, \vec{\xi}(0)\right) + h_i\right)$$

$$= sign\left(m^{(1)}(0)\,\xi_i^1\right) = \xi_i^1 sign\left(m^{(1)}(0)\right)$$

(falls nur $|m^1(0)| \geq |h_i| + \frac{1}{N}$). Für den Fall des asynchronen Updatens beachten wir, daß für das PSP

$$z_i(t) = \xi_i^1 m^{(1)}(t) - \frac{1}{N}\xi_i(t) = \xi_i^1 m^{(1)}(t) - \frac{1}{N}\xi_i(t)$$

gilt und damit

$$\xi_i(t+1) = \xi_i^1 sign\left(m^{(1)}(t)\right) \tag{6.25}$$

Eine Änderung von ξ_i und damit von $m^{(1)}$ findet nur statt falls (vergl. Kap. 6.1.2) $0 \neq \delta\xi_i = \xi_i^{neu} - \xi_i^{alt} = 2\xi_i^{neu} = 2sign\left(m^{(1)}\right)$. Die **dabei** stattfindende Änderung $\Delta m^{(1)} = m^{(1)}(t+1) - m^{(1)}(t)$ ist (mit $m^{(1)}(t) \geq \frac{1}{N}$)

$$\Delta m^{(1)} = \frac{1}{N}\left\{\xi_i(t+1) - \xi_i(t)\right\}\xi_i^1 = \frac{2}{N}sign\left(m^{(1)}(t)\right) \tag{6.26}$$

so daß also bei jeder Zustandsänderung eines Neurons die Überlappung mit dem Muster (oder seinem Inversen) immer weiter vergrößert wird oder konstant bleibt – jedes fehlerhafte Bit wird durchh einmaliges Updaten korrigiert. Für $\left|m^{(1)}\right| = 1$ kommt der Vorgang zum Stehen, da dann keine Zustandsänderung mehr erfolgt (Fixpunkt erreicht). Das ist genau dann der Fall, wenn jedes Neuron mindestens einmal aktualisiert wurde.

Damit ist dieses Netzwerk ein Assoziativspeicher, der bei Anlegen eines beliebigen Anfangszustand $\vec{\xi}(0)$ das Muster $\vec{\xi}^1$ (oder $-\vec{\xi}^1$) als Speicherinhalt ausgibt. Der Konfigurationsraum zerfällt in zwei Teilräume, die die Einzugsgebiete der beiden Speicherzustände sind.

Ein analoges Ergebnis folgt für mehrere Muster, falls diese ein **Orthogonalsystem** bilden. Das ist im allgemeinen nicht der Fall und die Speicherungseigenschaften hängen im allgemeinen stark von der Struktur des Mustersatzes ab. Analytische Ergebnisse allgemeinerer Natur können für den **Fall zufällig gewählter Muster** gewonnen werden.

Signal-Rauschanalyse für den Fall vieler Muster

Für Zufallsmuster können die Ergebnisse des Mattismodelles verwendet werden, indem das gesamte PSP in einen Signal- und einen Rauschterm zerlegt wird, der Signalterm entspricht dem Beitrag des einzelnen Musters im Mattismodell. Betr. Netzwerkzustand $\vec{\xi}$, dieser sei hinreichend nahe an einem der Muster, o.B. d. A. Muster 1

$$\left(\vec{\xi},\vec{\xi}^1\right) = m^{(1)} \approx 1, \quad \left(\vec{\xi},\vec{\xi}^\nu\right) = m^{(\nu)} = O\left(1/\sqrt{N}\right), \quad \text{für } \nu \neq 1$$

entsprechend Zerlegung des PSP

$$z_i = \vec{w}_i \cdot \vec{\xi} = \sum_{j,(j\neq i)} w_{i,j}\xi_j = \frac{1}{N}\sum_{j,(j\neq i)}\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \xi_j = \xi_i^1 m^{(1)} + \sum_{\nu=2}^{P} \xi_i^\nu m^{(v)} = \text{Signal+"cross talk"} \tag{6.27}$$

Die $m^{(\nu)}$ und damit die $\xi_i^\nu m^{(v)}$ sind (Gaußverteilte) Zufallszahlen mit Streuung $1/\sqrt{N}$. Um einen Eindruck von der Stärke des cross–talk Termes zu bekommen, setzen wir $\vec{\xi} = \vec{\xi}^1$ und schreiben

$$z_i = \xi_i^1 + \zeta$$

Für den cross–talk Term ζ findet man elementar

$$\left\langle \zeta^2 \right\rangle = \frac{P}{N} = \alpha$$

falls alle Muster unkorreliert sind (wie bisher immer angenommen). Damit muß die Speicherung zusammenbrechen, wenn α in die Größenordnung von 1 kommt. Für $N \to \infty$ nähert sich ζ einer Gaußverteilung mit Breite α

$$p\left(\zeta\right) = \sqrt{\frac{N}{2\pi P}} e^{-\frac{N}{2P}\zeta^2}$$

Damit ist die Wahrscheinlichkeit P_{Fehler}, daß für ein Neuron das Rauschen größer als das Signal ist (was zu einer fehlerhaften Speicherung dieses Bits führt)

$$P_{Fehler} = \sqrt{\frac{N}{2\pi P}} \int_1^\infty d\zeta\, e^{-\frac{N}{2P}\zeta^2} \simeq \sqrt{\frac{P}{2\pi N}} e^{-\frac{N}{2P}}$$

Die mittlere Anzahl der fehlerhaft geseicherten Bits ist $P_{Fehler}N$. Damit ist eine korrekte Speicherung des Musters mit einer Wahrscheinlichkeit nahe 1 nur dann möglich, wenn (in der Grenze großer N)

$$P \leq N/\ln N$$

so daß für große N der kritische Wert von α mit N wie

$$\alpha_c \sim \frac{1}{\ln N} \tag{6.28}$$

geht.

Der Rauschterm entspricht dem Einfluss der restlichen Muster $\nu \neq \mu$ auf die Speicherung des betrachteten Musters μ, reflektiert also eine mangelnde Trennschärfe vergleichbar dem Effekt des Übersprechens zwischen verschiedenen Nachrichtenkanälen. Wenn der Rauschterm in die gleiche Größenordnung kommt wie der Signalterm, d. h.

$$P \sim N, \quad \text{d.h. } \alpha > 0 \quad \text{für } N \to \infty$$

(so daß α =endlich für $N \to \infty$) werden die Muster nicht mehr exakt gespeichert, d. h. eine bestimmte Zahl von Bits in dem gespeicherten Muster (retrieval state) ist gegenüber dem Muster "geflippt", falls überhaupt eine Speicherung möglich ist.

Für P endlich und $N \to \infty$, d.h-

$$\alpha = 0 \quad \text{für } N \to \infty$$

folgt allerdings, daß für $0 << |m^1| \leq 1$ der Signalterm beliebig stark überwiegt, so daß die gleichen Verhältnisse wie bei nur einem gespeicherten Muster vorliegen, d. h. **jedes Muster entspricht exakt einem Fixpunkt der Dynamik**, und jeder dieser Fixpunkte hat ein endliches Einzugsgebiet. Beachte, daß daneben noch weitere Zustände, z. B. **Mischzustände** (Überlagerung der Muster) als Fixpunkte auftreten können, z. B.

$$\xi_i^{mix} = sign\left(a_1\xi_i^1 + a_2\xi_i^2 + a_3\xi_i^3\right)$$

mit geeignet gewählten Koeffizienten a_i.

Für eine endliche Speicherkapazität α

$$\frac{P}{N} = \alpha$$

d. h. $P \sim N$, so daß $\alpha > 0$ für $N \to \infty$ sind die "konspirativen" Korrelationen im Rauschterm nicht mehr vernachlässigbar, ihr Verhalten verleiht dem Hopfieldnetzwerk folgende Eigenschaften:

- Eine genauere Analyse zeigt, daß mit steigendem α der Rauschterm von der gleichen Ordnung wie der Signalterm wird, für

  $$\alpha > \alpha_c \approx 0.138 \tag{6.29}$$

  erfolgt ein plötzliches Zusammenbrechen der Speicherung, d.h. wenn für unterkritisches α noch eine befriedigende Speicherleistung erreicht wird ($|m^\mu| \approx 1$ für das gespeicherte Muster μ), ist die Zahl der speicherbaren Muster unmittelbar oberhalb von α_c gleich Null, ($m^\mu \approx 0 \quad \forall\mu$). Das ist ein Phasenübergang erster Ordnung ("Konfusionsübergang"). Die genauen

Verhältnisse ergeben sich aus der Diskussion der Überlappung als Funktion der Kapazität α, s. Abbildung. (Der Wert für das kritische α gilt für $N \to \infty$ und ist das Ergebnis einer Replica-Rechnung ohne Brechung der Replica-Symmetrie. Mit Symmetriebrechung, also als realistischerer Wert folgt $\alpha_c = 0,142$).

- Diese zeigt auch, daß die Fixpunkte nur näherungsweise mit den Mustern übereinstimmen, d. h. die Speicherung ist nur näherungsweise korrekt, für $\alpha \nearrow \alpha_c$ werden nur noch ca. 98% der Bits richtig gespeichert (ist für die meisten Zwecke, z. B. Bildspeicherung, ohne Bedeutung).

- Neben den Mustern (bzw. ihren Inversen) und ihren Mischungen treten noch sog. Geisterzustände (*spurious states*) als Fixpunkte auf. Insgesamt ist die Zahl der Fixpunkte (Minima der Energiefunktion) von der Ordnung $\exp(cN)$ mit $c \approx 0,2$.

6.2.3 Stochastische Netzwerke – das Hopfield Modell bei endlichen Temperaturen

Wegen der vielen Nebenminima ist es zum Auffinden der Speicherinhalte günstig, eine "simulated annealing" Strategie zur Verfügung zu haben. Das gelingt über die Einführung einer stochastischen Update–Regel, d. h. die Vorschrift bei der Aktualisierung des Neurons i, den neuen Zustand des Neurons nach dem Vorzeichen des PSP auszurichten, $\xi_i(t+1) = sign(z_i(t))$, (Anpassungsregel), wird nur mit einer gewissen Wahrscheinlichkeit befolgt, mit der komplementären Wahrscheinlichkeit wird das Gegenteil dieser Vorschrift ausgeführt.

Einen geeigneten Ansatz für diese Wahrscheinlichkeit findet man auf folgende Weise. Die deterministische Update–Regel kann man auch so interpretieren, daß die Wahrscheinlichkeit $P(\xi_i = +1) = \theta(z_i)$ ist, d. h. die Wahrscheinlichkeit, nach dem Updaten das Neuron im Zustand $\xi_i = 1$ zu finden, ist gleich 1 falls $z_i \geq 0$ und gleich Null sonst. Analog ist $P(\xi_i = -1) = \theta(-z_i)$. Das stochastische Neuron ist durch eine aufgeweichte Sprungfunktion charakerisiert, allgemein üblich ist, die Sprungfunktion $\theta(z)$ durch die Fermifunktion $f_\beta(z)$ mit

$$f_\beta(z) = \frac{1}{1+e^{-2\beta z}} \tag{6.30}$$

zu ersetzen, d.h. die stochastische Update–Regel lautet

$$\begin{aligned} P(\xi_i^{neu} = 1) &= \tfrac{1}{1+e^{-2\beta\, z_i}}, \\ P(\xi_i^{neu} = -1) &= \tfrac{1}{1+e^{2\beta\, z_i}} \end{aligned} \tag{6.31}$$

mit $\beta = 1/T$ und T einer fiktiven Temperatur (die natürlich nichts mit der biologischen Temperatur des Nervengewebes zu tun hat). Dies entspricht der Glauberdynamik eines Spinsystems. Mit

$$f_\beta(-z) = 1 - f_\beta(z)$$

gilt auch

$$P(\xi_i^{neu} = \pm 1) = \frac{1}{1+e^{-(\pm)2\beta\, z_i}} \tag{6.32}$$

Die Wahrscheinlichkeit für die Übereinstimmung des neuen Zustandes mit dem Vorzeichen des PSP (Anpassung) kann andererseits in Termen der Energieänderung ΔH, vgl. 6.9 ausgedrückt werden (beachte $\Delta H \leq 0$):

$$P(\text{Anpassg}) = \frac{1}{1+e^{\beta \Delta H}} \tag{6.33}$$

und

$$P\left(\text{Anti-Anpssg.}\right) = 1 - P\left(\text{Anpassg}\right) = \frac{1}{1+e^{-\beta\Delta H}} \tag{6.34}$$

Das Verhältnis der Wahrscheinlichkeiten ist dementsprechend durch einen Boltzmannfaktor gegeben

$$\frac{P\left(\text{Anpassg}\right)}{P\left(\text{Anti-Anpssg.}\right)} = e^{-\beta\Delta H} \tag{6.35}$$

Wichtigste Folge dieser Regel: Die Dynamik konvergiert nicht, da bei $T \neq 0$ die Wahrscheinlichkeit für Zustandsänderungen nie Null ist. Allerdings werden bei nicht zu hoher Temperatur die Trajektorien vornehmlich in der Nähe der Fixpunkte der $T = 0$ Dynamik bleiben, wenn Sie diese einmal angelaufen haben.

Die Speicherinhalte sind auf diese Weise bei endlicher Temperatur nicht mehr eindeutig indentifizierbar, Aussagen etwa über die Wiedergabegenauigkeit (retrieval quality) nur im statistischen Mittel sinnvoll: Definiere Mittelwerte über den Netzwerkzustand als zeitlichen Mittelwert:

$$\left\langle \vec{\xi}\left(t\right)\right\rangle^{th} = \frac{1}{t_0}\sum_{t'=0}^{t_0}\vec{\xi}\left(t-t'\right), \quad t >> t_0, \qquad t_0 >> 1 \tag{6.36}$$

Alle obigen Formeln bleiben erhalten, wenn Zustand $\vec{\xi}\left(t\right)$ generell durch $\left\langle \vec{\xi}\left(t\right)\right\rangle^{th}$ ersetzt wird, speziell ist die Wiedergabegenauigkeit (retrieval quality) durch die Überlappung mit dem gemittelten Zustandsvektor des Netzwerkes gegeben

$$m^{(\mu)}\left(t\right) = \frac{1}{N}\sum_{i=1}^{N}\xi_i^{\mu}\left\langle \xi_i\left(t\right)\right\rangle^{th} \tag{6.37}$$

wobei die Überlappung auch bei ansonsten idealer Speicherung mit steigender Temperatur betragsmäßig abnehmen wird. Ebenso

$$\begin{aligned}\left\langle z_i\right\rangle^{th} &= \vec{w}_i\cdot\left\langle \vec{\xi}\right\rangle^{th} = \sum_{j,(j\neq i)} w_{i,j}\left\langle \xi_j\right\rangle^{th} = \frac{1}{N}\sum_{j,(j\neq i)}\sum_{\mu=1}^{P}\xi_i^{\mu}\xi_j^{\mu}\left\langle \xi_j\right\rangle^{th} \\ &= \xi_i^1 m^{(1)} + \sum_{\nu=2}^{P}\xi_i^{\nu}m^{(v)} = \text{Signal+"cross talk"}\end{aligned} \tag{6.38}$$

mit

$$m^{(\mu)} = \frac{1}{N}\vec{\xi^{\mu}}\cdot\left\langle \vec{\xi}\left(t\right)\right\rangle^{th}$$

Für die Existenz eines retrieval states (RS) muß schon bei $T = 0$ immer Signal ¿¿ "cross talk" gelten, gleichzeitig aber auch $0 << m^{(1)} < 1$. Für den retrieval state zum Muster $\vec{\xi}_{\mu}$ macht es folglich Sinn zu schreiben

$$\left\langle \xi_i^{retr}\right\rangle^{th} = m^{\mu}\xi_i^{\mu} + \text{"cross talk"} \tag{6.39}$$

Im folgenden sollen nun Formeln für die Existenz des RS abgeleitet werden.

Die Methode des mittleren Feldes

Speicherzustand für den Fall $P << N$:

Wir betrachten zunächst den Fall $\alpha \to 0$ für $N \to \infty$, d.h. eine endliche Anzahl P von Mustern, so daß der "cross talk" vernachlässigt werden kann. Aus 6.38 folgt, daß – bis auf das Vorzeichen – das PSP für alle Neuronen im thermischen Mittel gleich groß ist. Wir machen deshalb den Ansatz

$$\langle z_i \rangle = m \xi_i^\mu \tag{6.40}$$

Mit 6.38 liefert die Kombination beider Gleichungen m, wenn $\langle \xi_i^{retr} \rangle$ als Funktion von $\langle z_i \rangle$ bekannt. Ansatz des mittleren Feldes:

$$\langle z_i \rangle = z_i \tag{6.41}$$

(das entspricht der Annahme des Verschwindens der Fluktationen für $N \to \infty$). Bestimmung von $\langle \xi_i \rangle$ mittels **??**:

$$\langle \xi_i \rangle = \textstyle\sum_{\xi_i = \pm 1} \xi_i P(\xi_i) = P(\xi_i = +1) - P(\xi_i = -1) =$$

$$\frac{1}{1+e^{-2\beta z_i}} - \frac{1}{1+e^{+2\beta z_i}} = \frac{e^{\beta z_i} - e^{-\beta z_i}}{e^{\beta z_i} + e^{-\beta z_i}} = \tanh(\beta z_i)$$

Mit $\tanh(x) = -\tanh(-x)$ und $m = \langle \xi_i^\mu \xi_i \rangle$ folgt

$$m = \tanh(\beta m) \tag{6.42}$$

Die (grafische) Lösung dieser transzendenten Gleichung für m zeigt die Existenz einer kritischen Temperatur T_c oberhalb derer die Überlappung $m = 0$ (keine Speicherung) und unterhalb derer $m \neq 0$ (Speicherung). Bei Erhöhung der Temperatur gibt es also einen Phasenübergang an T_c (Zusammenbruch des Speicherverhaltens). T_c entspricht der Curietemperatur des Magnetismus.

Der Speicherzustand für den Fall endlicher Beladung ($\alpha > 0$ für $N \to \infty$):

Im Ansatz 6.39

$$\left\langle \vec{\xi}(t) \right\rangle = m^{(\mu)} \vec{\xi}^\mu + \text{"cross talk"} \tag{6.43}$$

weist der "cross talk"–Term auf die auch bei $T = 0$ noch fehlerhaft gespeicherten Bits hin. Die Stabilität der Speicherung eines jeden Bits hängt nun sowohl von der Stärke des cross talk ab als auch von der Stärke der Überlappung $m^{(\mu)}$, die ihrerseits durch die thermischen Fluktuationen reduziert wird. Beide Effekte bedingen im Wechselspiel die Speicherleistung des Netzwerkes.

Die genaue Bestimmung des Verhaltens bei endlicher Temperatur ist sehr schwierig und Gegenstand zahlreicher Untersuchungen. Verhalten inzwischen, zumindest in der replicasymmetrischen Näherung (s.u.), fast restlos aufgeklärt, insbesondere ex. ein detailliertes Phasendiagramm, d. h. Darstellung der Eigenschaften des Netzes in der α, T- Ebene, in der scharfe Grenzen die Existenzgebiete der

einzelnen Phasen, die das System annehmen kann, markieren. Unterschieden wird speziell die Spinglasphase, in der eine exponentiell grosse Zahl von stationären Zuständen existieren, die aber keine Überlappung mit den Mustern mehr haben, von der retrieval Phase, bei der die Speicherinhalte stationären Zuständen entsprechen. Hier funktioniert das Netz als Assoziativspeicher auch bei höheren Temperaturen.

Zusammenhang mit Isingmodell des Ferromagnetismus:

Wir führen eine Umeichung des Netzwerkzustandes $\vec{\xi}(t)$ bezüglich eines der Muster, o. B. d. A. Muster 1 durch, d. h. definieren neue Zustandsvariable

$$\tilde{\xi}_i(t) = \xi_i^1 \xi_i(t) \quad \forall i = 1, 2, \ldots, N$$

analog $\tilde{z}_i = z_i \xi_i^1$. Interpretation: $\tilde{\xi}_i$ ist gleich $+1$ wenn $\xi_i = \xi_i^1$ d. h. gleiche Ausrichtung. Unter Vernachlässigung der Übersprechterme wird die Energiefunktion

$$H\left[\vec{\xi}\right] = H\left(\tilde{\xi}\right) = -\frac{1}{N} \sum_{i<j} \tilde{\xi}_i \tilde{\xi}_j \tag{6.44}$$

Interpretieren wir $\tilde{\xi}_i$ als Spins des Teilchens i, so ist 6.44 das Modell eines Ferromagneten mit konstanter Kopplung zwischen den Spins.

Mit 6.38 gilt

$$\langle \tilde{z}_i \rangle = m^{(1)} + \text{Rauschen} \tag{6.45}$$

In der neuen Eichung ist also bis auf das Rauschen das mittlere PSP für alle Neuronen gleich und entspricht dem mittleren Feld im Ferromagneten! Unter Vernachlässigung des Rauschens gilt also auch

$$\langle \tilde{z}_i \rangle = \frac{1}{N} \sum_{i'=1}^{N} \langle \tilde{z}_{i'} \rangle = \frac{1}{N} \left\langle \sum_{i'=1}^{N} \tilde{z}_{i'} \right\rangle = \frac{1}{N} \sum_{i'=1}^{N} \tilde{z}_{i'} + O\left(\frac{1}{\sqrt{N}}\right)$$

oder

$$\langle \tilde{z}_i \rangle = \tilde{z}_i$$

in Übereinstimmung mit 6.41. Im thermodynamischen Limes ($N \to \infty$) kann also das thermische Mittel durch ein Mittel über die Plätze ersetzt werden (Voraussetzung: Die thermischen Fluktuationen des Platzmittels sind vernachlässigbar - keine "konspirativen" Korrelationen).

6.3 Informationstheorie und Statistische Mechanik

Eine detaillierte Analyse der Eigenschaften von Attraktornetzwerken mit symmetrischen Kopplungen (wie das Hopfieldmodell) gelingt mit Methoden der statistischen Mechanik, die im folgenden kurz dargestellt werden sollen. Wir setzen dabei keine Kenntnisse der Statistischen Mechanik voraus sondern leiten die gewünschten Ausdrücke aus der Informationstheorie ab.

6.3.1 Informationstheorie

Betrachte ein Alphabet $\mathcal{A}$ aus N_A Zeichen α.

Beispiel: Jeder Zustandsvektor $\vec{\xi} = (1, 1, -1, ..., 1, ..., -1, ...)$ eines NNs aus N Neuronen kann als ein Zeichen $\alpha \in \mathcal{A}$ interpretiert werden, $\alpha \in \{-1, 1\}^N$. Das Alphabet besteht dann aus 2^N Zeichen, d.h. $N_A = 2^N$.

Betrachte Nachrichtenquelle, die Zeichenketten $\alpha_1, \alpha_2, ...$ aus Buchstaben des Alphabets ($\alpha_i \in \mathcal{A}$) generiert.

Beispiel: Neuronales Netz mit der Update–Regel 6.5 generiert die Zeichenkette $\vec{\xi}(0), \vec{\xi}(1), \vec{\xi}(2), ...$

Über die Zeichenfolge seien nur die Wahrscheinlichkeiten (relative Häufigkeit) P_α für das Auftreten jedes Zeichens $\alpha \in \mathcal{A}$ bekannt. **Frage**: Wie viele Bits sind im Mittel **mindestens** zur Übertragung dieser Zeichenkette erforderlich. Um die Mindestzahl von Bits zu bestimmen überlege man sich eine geeignete Verschlüsselung der Kette (Datenkompression), z. B.

Übertrage nur den Abstand bis zum Wiederauftreten des Zeichens. Beispiel: sei $\mathcal{A} = \{a, b, c\}$. Die Zeichenkette $(a, b, b, b, a, c, a, a, a, c, ...)$ wird verschlüsselt als Tabelle

a	0	3	1	0	...
b	1	0	0	...	
c	5	3	...		

übertragen (der Abstand ist die Zahl der dazwischenliegenden Zeichen). Mit dieser Information kann die Zeichenkette eindeutig rekonstruiert werden.

Der mittlere Abstand zwischen zwei gleichen Zeichen α ist

$$d_\alpha = \frac{1}{P_\alpha} \tag{6.46}$$

da P_α die relative Häufigkeit (N_α / N_{Kette}) des Zeichens α ist. Der Abstand in Bits ist

$$d_\alpha^{[Bits]} = \log_2 \frac{1}{P_\alpha} = -\log_2 P_\alpha \tag{6.47}$$

Die im Mittel nötige Anzahl von Bits pro Zeichen ist demzufolge[2]

$$I = -\sum_{\alpha} P_\alpha \log_2 P_\alpha \tag{6.48}$$

Das ist die Shannonsche Formel für die Information einer Nachrichtenquelle.

Beispiele: Treten alle Zeichen gleich häufig auf, d. h. $P_\alpha = 1/N_A$ so ist

$$I = \log_2 N_A$$

d. h. die Information ist gleich der Anzahl der Zeichen des Alphabetes gemessen in Bits. Beispiele:

- Ist speziell $\mathcal{A} = \{-1, 1\}^N$, so ist in diesem Fall gleichwahrscheinlicher Zeichen $I = N$. Das ist klar, denn für zufällig gewählte Bitketten der Länge N gibt es keine Datenkompression die die Übertragung mit weniger denn N Bits ermöglichen würde.
- Treten nur K von den N_A Zeichen des Alphabetes überhaupt auf, so gilt bei Gleichwahrscheinlichkeit dieser Zeichen

$$I = \log_2 K \tag{6.49}$$

6.3.2 Statistische Mechanik

Die Boltzmann-Gibbssche Verteilung

Wir deuten die Mikrozustände eines physikalischen Systems als die Zeichen des Alphabets $\mathcal{A}$. Für unser NN sind das die Zustandsvektoren $\vec{\xi} \in \{-1, 1\}^N$. Angenommen wir hätten praktisch keine Information über die realisierten Mikrozustände im System uaßer der Tatsache, daß etwa gewisse makroskopische Größen einen vorgegebenen Wert annehmen sollen. Im Zustand des thermischen Gleichgewichtes ist das speziell die mittlere Energie des Systems

$$\langle H \rangle^{th} = \sum_{\alpha} H_\alpha P_\alpha \tag{6.50}$$

Für das NN ist $\alpha = \vec{\xi}$, so daß $H_\alpha = H\left[\vec{\xi}\right]$ und $\sum_\alpha$ der Summation über alle $\vec{\xi}$ entspricht, d. h.

$$\sum_{\alpha} g_\alpha = \sum_{\xi_1 \in \{-1,1\}} \cdots \sum_{\xi_N \in \{-1,1\}} g\left[\xi_1, \ldots \xi_N\right]$$

Gesucht ist die willkürfreie (d.h. ohne willkürliche Information einzuführen) Verteilung P_α der Mikrozustände des Systems, die diesen Mittelwert via **??** liefert.

[2]Wir betrachten sehr lange Zeichenketten, so daß die nur einmal zu übermittelnde Information über die Zeilennummer der Tabelle nicht ins Gewicht fällt.

Lösung: Variation der Information 6.48 unter Berücksichtigung der Nebenbedingung **??**. Die Methode der Lagrangeschen Parameter liefert (wie in der SM üblich messen wir die Information nicht in Bits, d. h. $\log_2 \Rightarrow \log$).

$$0 = \frac{\partial}{\partial P_\alpha}\left(-\sum_\gamma P_\gamma \log P_\gamma + \beta \sum_\gamma H_\gamma P_\gamma\right) \qquad \forall\alpha \tag{6.51}$$

so daß

$$0 = -\log P_\alpha - 1 + \beta H_\alpha \tag{6.52}$$

mit Lösung

$$P_\alpha \sim e^{-\beta H_\alpha}$$

Unter Berücksichtigung der Normierung der Wahrscheinlichkeitsverteilung folgt

$$P_\alpha = \frac{1}{Z} e^{-\beta H_\alpha} \tag{6.53}$$

mit der sog. Zustandssumme Z als Normierungsfaktor

$$Z = \sum_\alpha P_\alpha \tag{6.54}$$

Die Verteilung 6.53 ist die Boltzmann-Gibbssche Verteilung der klassischen Statistischen Mechanik. Wir wählen die Einheiten so, daß die Boltzmannkonstante gleich 1 ist, dann ist

$$\beta = \frac{1}{T} \tag{6.55}$$

mit T der Temperatur des Systems (im Wärmebad).

Freie Energie und Entropie

Die Information 6.48 in "natürlichen" Einheiten gemessen heißt in der SM Entropie S

$$S = -\sum_\alpha P_\alpha \log P_\alpha \tag{6.56}$$

Analog zu 6.49 kann man folgern, daß

$$M = e^S \tag{6.57}$$

die mittlere Zahl von Zuständen ist, die im thermischen Gleichgewicht mit etwa gleicher Wahrscheinlichkeit realisiert werden. Entsprechend ist e^S bzw. S selbst ein Maß für die Information die man braucht, um einen speziellen der unter dem makroskopischen *constraint* $\langle H\rangle^{th} = E$ möglichen Mikrozustände festzulegen.

Die freie Energie F ist definiert als

$$F = \langle H\rangle^{th} - TS \tag{6.58}$$

so daß

$$F = -T \log Z \tag{6.59}$$

Damit kann die freie Energie als das erzeugende Funktional für die Bestimmung fast aller physikalischen Größen des Systems verwendet werden. So folgt z. B. für die Korrelationsfunktion

$$\langle \xi_i \xi_j \rangle^{th} = -\frac{\partial F}{\partial w_{ij}} \tag{6.60}$$

Führt man zum Hamiltonian H des Systems geeignete Hilfsfelder hinzu und bestimmt die F, so ergeben sich die gesuchten Größen durch Ableitung nach den Feldern. Beispiel:

$$H = -\frac{1}{2} \sum_{ij} \xi_i w_{ij} \xi_j \Rightarrow -\frac{1}{2} \sum_{ij} \xi_i w_{ij} \xi_j - \sum_i b_i \xi_i \tag{6.61}$$

dann ist

$$Z_b = \sum_{\xi_1 \in \{-1,1\}} \ldots \sum_{\xi_N \in \{-1,1\}} \exp\left(\frac{\beta}{2} \sum_{ij} \xi_i w_{ij} \xi_j + \beta \sum_i b_i \xi_i \right) \tag{6.62}$$

$$\tag{6.63}$$

$$= Tr_\xi \exp\left(\frac{\beta}{2} \sum_{ij} \xi_i w_{ij} \xi_j + \beta \sum_i b_i \xi_i \right) \tag{6.64}$$

(Tr_ξ bedeutet die Summation über alle Aktivitätsvariablen ξ_i). Mit 6.59 folgt

$$\langle \xi_i \rangle^{th} = \frac{1}{Z_b} \sum_{\xi_1 \in \{-1,1\}} \ldots \sum_{\xi_N \in \{-1,1\}} \xi_i \exp\left(\frac{\beta}{2} \sum_{ij} \xi_i w_{ij} \xi_j \right) \tag{6.65}$$

$$\tag{6.66}$$

$$= -\left. \frac{\partial F_b}{\partial b_i} \right|_{h=0} \tag{6.67}$$

wobei $F_b = -T \log Z_b$

Phasenübergänge und Brechung der Ergodizität

Wir haben oben die zeitlichen Mittel $\langle \ldots \rangle^{th}$ gleich dem Mittel über die Gesamtheit gesetzt. Das ist dann richtig, wenn für alle Zustände $\vec{\xi} \in \{-1, +1\}^N$ die Wahrscheinlichkeiten der Zustände gemäß der BG-Verteilung mit der relativen Häfigkeit des Auftretens dieses Zustandes im Laufe der Zeit übereinstimmt[3]. Da

[3]Wir nehmen an, daß das System sich in einem stationären Zustand befindet, d. h. daß hinreichend viel Zeit seit dem Initialisieren des Systems vergangen ist, so daß das zeitliche Mittel selbst nicht mehr zeitabhängig ist. Falls die Temperatur nicht zu groß ist fluktuiert der Systemzustand $\vec{\xi}(t)$ dann um einen Fixpunkt der $T = 0$ Dynamik.

$P_\alpha = P\left[\vec{\xi}\right] > 0$ heißt das, daß das System jeden dieser Zustände mit der relativen Häufigkeit $P\left[\vec{\xi}\right]$ auch wirklich anläuft. Damit wandert der Systemzustand $\vec{\xi}(t)$ im Laufe der Zeit durch den gesamten Zustandsraum. Ein solches System heißt ergodisch.

Das kann aber eigentlich nicht richtig sein. Wegen $H\left[\vec{\xi}\right] = H\left[-\vec{\xi}\right]$ gilt nämlich ganz allgemein für die Boltzmann–Gibbs–Verteilung des Netzwerkes

$$P\left[\vec{\xi}\right] = P\left[-\vec{\xi}\right] \tag{6.68}$$

Für das Mattismodell ist z. B.

$$H = -\frac{1}{2}\sum_{i,j}\xi_i w_{ij}\xi_j = -\beta\frac{N}{2}\left(m^1\right)^2 \tag{6.69}$$

und $m^1 \Rightarrow -m^1$ falls $\vec{\xi} \Rightarrow -\vec{\xi}$. Damit folgt für den Mittelwert

$$\langle\xi_i\rangle^{th} = \sum_{\xi_1\in\{-1,1\}} \ldots \sum_{\xi_N\in\{-1,1\}} \xi_i \exp\left(\frac{\beta}{2}\sum_{ij}\xi_i w_{ij}\xi_j\right)/Z \tag{6.70}$$

daß

$$\langle\xi_i\rangle^{th} = 0 \tag{6.71}$$

für alle Zustände $\vec{\xi}$. Das steht in eklatantem Widerspruch zu 6.43, d. h. (bei kleinem cross talk=geringe Zahl von Mustern)

$$\langle\xi_i\rangle^{th} = \xi_i^1 m^1 \tag{6.72}$$

mit $|m^1| > 0$ falls sich das Netzwerk in einem retrieval state befindet.

Der Widerspruch löst sich auf, wenn wir berücksichtigen, daß 6.72 unter der Bedingung gewonnen wurde, daß einerseits die Zeitmittelung mit Zeitfester t_0 zwar lang genug zu wählen ist, um die thermischen Fluktuationen sicher auszumitteln[4], $t_0 >> t_F$, daß aber andererseits $t_0 << t_E$ sein soll, wobei t_E die mittlere Zeit des Überganges (unter dem Einfluß der "thermischen" Störungen) von einem Einzugsgebiet in ein anderes ist. Die Wahrscheinlichkeit $P\left[\vec{\xi}\right]$ für die Realisierung eines bestimmten Zustandes $\vec{\xi}$ ist nur dann durch die Boltzmann–Gibbs–Verteilung gegeben, wenn wir das System (bei $N = endlich$) im Prinzip unendlich lange beobachten, so daß es tatsächlich **alle überhaupt erreichbaren** Zustände auch anläuft. Real sollte die Beobachtungszeit t_B aber zumindest $t_B >> t_E$ sein. t_B ist von der Zahl der Zustände abhängig und wächst mit diesen an. Man kann grob

$$t_E \sim e^{\gamma N} \tag{6.73}$$

[4]Die Mittelungszeit muß also so lang sein, daß alle Zustände, die bei der Fluktuation um einen Attraktorzustand des Netzes bei $T = 0$ möglich sind auch wirklich oft genug aufgetreten sind.

schätzen. Jedenfalls geht diese Zeit mit $N \to \infty$ so daß es dann keine endliche Beobachtungszeit mehr gibt, unter der das System aus einem Einzugsgebiet eines Attraktors in ein anderes fliehen kann. Dann ist die Boltzmann–Gibbs–Verteilung nicht mehr anwendbar, bzw. die Boltzmann–Gibbs–Verteilung ist nur in dem Sinne anwendbar, daß man eine auf die jeweilige Phase beschränkte Summation für die Mittelwertbildung ausführt, d. h. anstelle von 6.71 gilt für den Mittelwert einer beliebigen Zustandsgröße $f\left[\vec{\xi}\right]$

$$\left\langle f\left[\vec{\xi}\right]\right\rangle^{th} = \sum_{\vec{\xi}\in\text{Phase}} f\left[\vec{\xi}\right] \exp\left(\frac{\beta}{2}\sum_{ij}\xi_i w_{ij}\xi_j\right)/Z_b \tag{6.74}$$

Für den Fall des einfachen Mattismodells finden wir z. B.

$$\left\langle f\left[\vec{\xi}\right]\right\rangle^{th} = \sum_{\vec{\xi}\,\text{mit}\,m^1>0} f\left[\vec{\xi}\right] e^{-\frac{N}{2}m^1} \tag{6.75}$$

was für $f = \xi_i$ wieder auf 6.43

$$\langle\xi_i\rangle^{th} = \xi_i^1 m^1 + \text{"cross talk"} \tag{6.76}$$

führt.

Dieses Scenario zeigt, daß bei der Existenz von Phasenübergängen der Zustandsraum in verschiedene, zu den einzelnen Phasen assoziierte Gebiete unterteilt zu denken ist, die das System nicht verlassen kann. Energetisch bedeutet das, daß zwischen diesen Gebieten für $N \to \infty$ unendlich hohe Energiebarrieren bestehen. Das entspricht einer **Brechung der Ergodizität** des Systems. Im Hopfieldmodell bzw. im Spinglas ist das Besondere, daß wir es mit einer makroskopisch großen Vielzahl von nichtergodischen Phasen zu tun haben.