Mathematische Statistische

Mechanik

Inhaltsverzeichnis

1 Einleitung

V. I. Arnold schreibt in seinem Aufsatz “On teaching mathematics”: “Mathematics is a part of physics, it is the part where experiments are cheap.” (Den vollständigen Text dieses Aufsatzes finden Sie im Anhang.) Diese Meinung ist sicherlich pointiert, wahr ist aber sicher, dass man einen Gutteil der Mathematik schlechter oder gar nicht versteht, wenn man die physikalischen Wurzeln außer Acht lässt. Um den Arnoldschen Gedanken noch einmal aufzugreifen, besteht Physik ja im wesentlichen darin, einen bestimmten Teil der Natur (Physiker würden sicher sagen: den wesentlichen) zu erklären. Interessanterweise lassen sich viele der physikalisch relevanten Naturphänomene in der Sprache der Mathematik ausdrücken – für Richard Feynman definiert dies geradezu eine Wissenschaft. Die Methode der mathematischen Physik besteht nun darin, mit Hilfe der Mathematik Schlussfolgerungen innerhalb der physikalischen Modelle zu ziehen. Dass dies funktioniert ist zunächst nicht selbstverständlich, denn die Natur weiß ja a priori nichts von unseren mathematischen Spielregeln (philosophisch spricht man vom Anwendungsproblem der Mathematik; für die meisten Mathematiker und Physiker stellt sich die Frage, ob Mathematik überhaupt die geeignete Sprache zur Beschreibung der Welt ist, interesssanterweise meist erst, wenn sie lange genug erfolglos an einem Problem gearbeitet haben). Wenn aber die Methode der mathematischen Physik erfolgreich ist, so ist, um es mit David Ruelle zu sagen, keine Technik zu kompliziert und keine zu trivial, um zu Resultaten zu gelangen.

Statistische Mechanik ist nun der Zweig der mathematischen Physik, der der Wahrscheinlichkeitstheorie besonders nahe steht. Obwohl der Gedanke, ein System mittels einiger statistischer Grundannahmen zu beschreiben, spätestens seit Clausius zum Arsenal der theoretischen Physiker gehört, ist es doch den fundamentalen Arbeiten von Boltzmann, Maxwell (beide um 1870) und wenig später Gibbs zu verdanken, dass die statistische Mechanik als eine eigenständige Disziplin begründet wurde. Allgemeine Grundannahmen sind hierbei:

a) Das System ist eine Ansammlung identischer Subsysteme;

b) die Anzahl dieser Subsysteme ist groß;

c) die Wechselwirkung zwischen den Subsystemen ist dergestalt, dass das Gesamtsystem ein makroskopisches Verhalten aufweist.

Ziel ist es, das makroskopische Verhalten aus Annahmen über das mikroskopische Verhalten, d. h. die Wechselwirkung der Subsysteme, abzuleiten.

Obschon diese Ideen prinzipiell auf jedes System anwendbar sind, das den gestellten Anfordrungen genügt, lag es gerade zu Beginn der statistischen Mechanik nahe, eine rigorose Herleitung der Gesetzmäßigkeiten der statistischen Mechanik zu versuchen. Zu diesem Zeitpunkt war der Aufbau der Materie aus kleinsten Teilchen zumindest eine konkurrenzfähige Hypothese, die sogenannte “Atomhypothese”, und die Wechselwirkung dieser kleinsten Teilchen in einem Gas ist so schwach, dass das obige Programm am ehesten durchführbar erscheint.

Auf einige der Probleme, die eine solche kinetische Gastheorie aufwirft, insbesondere auf das Wachstum der Entropie, werden wir in Kapitel 2 kurz eingehen. Dabei werden wir Boltzmanns Resultat zitieren, die Einwände dagegen aber ausführlicher an einem Spielzeugmodell, das P. und T. Ehrenfest vorgeschlagen haben, diskutieren.

Im dritten bis achten Kapitel werden wir dann auf die "klassische" mathematische statistische Mechanik eingehen. Sie erfuhr zu Beginn der 70er Jahre des 20. Jahrhunderts Zuwendung von vielen Wahrscheinlichkeitstheoretikern. Die zentrale Fragestellung, die man auf viele verschiedenartige Fragestellungen anwendet, ist hierbei die Frage der Phasenübergänge.

Die letzten Kapitel sollen dann einem besonders reichhaltigen Thema gewidmet sein: der statistischen Mechanik ungeordneter Systeme. Hierunter versteht man Systeme, die mit einem zusätzlichen Wahrscheinlichkeitsmechanismus ausgestattet sind, der die Interaktion zwischen zwei Teilchen nicht nur von deren Eigenschaften und deren Position abhängen lässt, sondern auch vom Zufall. Solche ungeordneten Systeme treten in der Physik bei der Modellierung so unterschiedlicher Phänomene wie Glas und neuronaler Strukturen auf. Die mathematische Forschung auf diesem Gebiet ist noch relativ jung (seit ungefährt 1990) und noch lange nicht abgeschlossen.

Diese Vorlesung basiert auf einer Menge Vorlagen, von denen Notizen von Mark Kac, das Buch von David Ruelle, sowie die Monographien von R. S. Ellis und Sinai explizit genannt seien. In weiten Teilen aber wird die Vorlesung einem Buch von Anton Bovier "Statistical Mechanics of Disordered Systems" (erscheint 2006 bei Cambridge University Press) folgen. Dieses sei jedem Studierenden ans Herz gelegt.

2 Boltzmanns Ideen und Ehrenfests Urnenmodell

2.1 Boltzmann-Gleichung und H-Theorem

Wie in der Einleitung erwähnt, gehen die Anfänge der statistischen Mechanik auf Boltzmanns Analyse der thermodynamischen Gesetzmäßigkeiten zurück. Wir wollen zunächst seine Ergebnisse betrachten: Boltzmanns Grundannahme ist, dass Materie, in diesem Fall ein Gas, aus kleinsten Teilchen besteht, Atomen oder Molekülen, über deren individuelle Bewegungen man zwar nichts aussagen kann, die aber gewissen statistischen Annahmen genügen. Genauer nehmen wir an, dass es zu jedem Zeitpunkt t und an jedem Ort r unseres Gasbehälters sowie zu jedem Impuls p eine Dichte

$$f(r, p, t)$$

gibt, die die relative Anzahl der Teilchen zur Zeit t in r mit Impuls p beschreibt. Nun ist es möglich, mittels Impuls- und Energieerhaltung auszurechnen, welchen Impuls zwei Teilchen mit Anfangsimpuls p_1 und p_2 nach einem Stoß haben. Berechnet man zudem, wie viele solcher Stöße innerhalb eines kleinen Zeitintervalls dt bei gegebener Dichte $f(r, p, t)$ stattfinden, so kommt man zu einer Differentialgleichung für f, der sogenannten Boltzmann-Gleichung, die die Änderungen der Dichtefunktion über die Zeit beschreibt. Das Erstaunliche an dieser unschuldig aussehenden Differentialgleichung ist nun der folgende Sachverhalt: Ist $f(r, p, t) \equiv f(p, t)$, also die Dichte nicht mehr abhängig vom Ort, so impliziert die Boltzmann-Gleichung, dass

$$\frac{d}{dt}H(t) := \frac{d}{dt}\left(-\int f(p,t)\log f(p,t)dp\right) \geq 0$$

gilt. Die Größe $H(t)$ stimmt bis auf eine additive Konstante mit der thermodynamischen Entropie überein, wir werden auch sie in der Folge mit Entropie bezeichnen.

Die Boltzmann-Gleichung impliziert also ein Anwachsen der Entropie. Das sogenannte Boltzmannsche H-Theorem stellt eine Brücke zwischen der klassischen Mechanik und der Thermodynamik dar.

Trotz dieser wichtigen Rolle, die Boltzmanns Überlegungen in der Physik spielen, gab es von Beginn an Einwände gegen sie (und es gibt bis heute Spielraum für Interpretationen). Die beiden wichtigsten Einwände gegen das H-Theorem sind:

1. Das H-Theorem widerspricht der Mechanik. In der Tat sind alle Gesetze der Mechanik invariant unter der Zeitumkehr $t \to -t$: Beobachtet man die Bewegung eines einzelnen Teilchens, so lässt sich an ihr nicht ablesen, ob "der Film vorwärts oder rückwärts läuft". Dagegen zeichnet das H-Theorem eine Zeitrichtung aus. Loschmidt formulierte diesen Einwand schon 1876 so: Wenn ein Gas sich aus einem Anfangszustand S_0 nach t Zeiteinheiten in einen Zustand S_t bewegt hat, dann folgt nach Boltzmann
$$H_t \geq H_0.$$

Wenn man nun alle Geschwindigkeiten umkehrt, dann wird das System nach weiteren t Schritten wieder in S_0 sein, wohingegen das H-Theorem

$$H_0 \geq H_t$$

impliziert, also ändert sich H überhaupt nicht. Boltzmanns Einwand darauf ("Bitte, drehen Sie die Geschwindigkeit um.") ist zwar wichtig, hilft aber nicht so recht weiter.

2. Ein weiterer, vielleicht noch wesentlicherer Einwand stammt von Zermelo, der bemerkte, dass aufgrund des Poincaréschen Wiederkehrsatzes ein abgeschlossenes dynamischen System beliebig nahe an seinen Ausgangspunkt zurückkehrt. Wenn H also nur von der Dynamik abhängt, kann es nicht immer steigen. Boltzmann versuchte herauszuarbeiten, dass die Zeiten bis zu einer Rückkehr enorm lang sind (und entgegnete: "Bitte, warten Sie darauf.").

Wir wollen dies genauer betrachten.

2.2 Der Poincarésche Wiederkehrsatz

Mathematisch lässt sich die Situation wie folgt beschreiben: Das Gas mit N Teilchen ist beschrieben durch $2N$ Koordinaten (also typischerweise einem Element aus dem $\mathbb{R}^{6N}$)

$$q_1, \ldots, q_N, \; p_1, \ldots, p_N,$$

wobei die q_i die Ortskoordinaten und die p_i die Impulskoordinaten des i-ten Partikels beschreiben. Das Verhalten des Systems wird durch die Hamiltonfunktion

$$H(q_1, \ldots, q_N, p_1, \ldots, p_N)$$

(die t nicht explizit enthält) und die Bewegungsgleichungen

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}$$

festgelegt. Es folgt, dass H (das man nicht mit der Entropie H verwechseln sollte) eine Konstante der Bewegungen des Systems ist (Energieerhaltungssatz) und wir nehmen weiter an, dass die "Energiefläche"

$$\mathcal{E} = \{(q,p) : H(q_1, \ldots, q_N, p_1, \ldots, p_N) = E\}$$

eine beschränkte Menge ist.

Nun angenommen, wir starten unser System im Punkte

$$P_0 = (q_1^0, \ldots, q_N^0, p_1^0, \ldots, p_N^0).$$

Die Position des Teilchens i zur Zeit t ist eine Funktion (f_i, g_i), die man aus der Lösung der Bewegungsgleichung erhält:

$$\begin{aligned} q_i(t) &= f_i(q_1^0, \ldots, q_N^0, p_1^0, \ldots, p_N^0), \\ p_i(t) &= g_i(q_1^0, \ldots, q_N^0, p_1^0, \ldots, p_N^0). \end{aligned}$$

Die Gesamtheit all dieser Funktionen definiert eine einparametrige Familie $(T_t)_{t\in\mathbb{R}^+}$ von Transformationen des Zustandsraums des Systems – dieses heißt Γ – in sich selbst:

$$P_t = T_t P_0.$$

Der berühmte Satz von Liouville besagt nun, dass das Lebesguemaß $\lambda\!\!\lambda^{6N}$ unter T invariant ist, also

$$\lambda\!\!\lambda^{6N}(A) = \lambda\!\!\lambda^{6N}(T_t A)$$

für alle messbaren Mengen $A \subseteq \mathbb{R}^{6N}$ und alle $t \in \mathbb{R}^+$.

Wenn wir uns nun auf $\mathcal{E}$ beschränken, impliziert der Satz von Liouville das folgende: Falls wir auf $\mathcal{E}$ das Maß μ vermöge

$$\mu(A) = \int_A \| \bigtriangledown H\|^{-1} d\sigma$$

(wobei wir annehmen, dass $\| \bigtriangledown H\| > c > 0$ auf $\mathcal{E}$ gilt, so dass $\mu(\mathcal{E}) < +\infty$ ist), wobei σ ein Oberflächenelement ist und

$$\| \bigtriangledown H\|^2 = \sum_{i=1}^{N} \left(\frac{\partial H}{\partial p_i}\right)^2 + \left(\frac{\partial H}{\partial q_i}\right)^2,$$

dann gilt

$$\mu(T_t(A)) = \mu(A).$$

All dies nur zur Einleitung.

Abstrakt sind wir in der folgenden Situation: Wir haben auf einer Menge Ω eine σ-Algebra und ein endliches Maß μ gegeben, wir können o.B.d.A. annehmen, dass

$$\mu(\Omega) = 1$$

gilt. Darüber hinaus haben wir eine einparametrige Familie $(T_t)_{t\in\mathbb{R}^+}$ maßerhaltender, bijektiver Abbildungen von Ω in sich gegeben (die Tatsache, dass die Abbildungen in unserem Beispiel bijektiv sind, folgt trivial aus den Bewegungsgleichungen). Wir nehmen an, dass die $(T_t)_{t\in\mathbb{R}^+}$ eine Halbgruppe sind, d. h. dass $T_t \circ T_s = T_{t+s}$ gilt.

Der Poincarésche Wiederkehrsatz lautet nun:

Theorem 2.1 *Sei A eine meßbare Teilmenge von Ω mit $\mu(A) > 0$. Dann gibt es für fast alle $\omega \in A$ beliebig große Werte $t \in \mathbb{R}^+$, so dass*

$$T_t(\omega) \in A$$

gilt.

Beweis: Es genügt, diskrete Zeiten $T_1, T_2, \dots$ zu betrachten. Offensichtlich gilt

$$T_2 = T_1^2,\ T_3 = T_1^3,\ \text{etc.}$$

Sei

$$A_1 = \{\omega \in A : T_1\omega \in A\}$$

und iterativ

$$A_n := \{\omega \in A : T_1\omega \notin A, \dots, T_{n-1}\omega \notin A, T_n\omega \in A\}.$$

Sei für eine Menge B

$$\chi_B(\omega) = \left\{ \begin{array}{ll} 1 & \omega \in B \\ 0 & \omega \notin B \end{array} \right. .$$

Dann gilt

$$\chi_{A_n}(\omega) = \chi_A(\omega)(1 - \chi_A(T_1(\omega)) \dots (1 - \chi_A(T_1^{n-1}(\omega)))\chi_A(T_1^n(\omega))$$

und daher

$$\mu(A_n) = \int_\Omega \chi_A(\omega)(1 - \chi_A(T_1(\omega))) \dots (1 - \chi_A(T_1^{n-1}(\omega))\chi_A(T_1^n(\omega))d\mu.$$

Setzen wir nun

$$f_n = \int_\Omega (1 - \chi_A(\omega))(1 - \chi_A(T_1(\omega)) \dots (1 - \chi_A(T_1^{n-1}(\omega)))d\mu.$$

Dann lässt sich $\mu(A_n)$ ausdrücken als

$$\mu(A_n) = f_{n-1} - 2f_n + f_{n+1}$$

unter der Randbedingung $f_0 = 1$. Tatsächlich folgt diese Formel aus der Tatsache, dass T maßerhaltend ist, z. B. für $\mu(A_3)$:

$$\begin{aligned}
\mu(A_3) &= \int_\Omega \chi_A(\omega)(1 - \chi_A(T_1(\omega)))(1 - \chi_A(T_1^2(\omega)))\chi_A(T_1^3(\omega))d\mu \\
&= \int_\Omega (1 - \chi_A(T_1(\omega))(1 - \chi_A(T_1^2(\omega)))d\mu \\
&- \int_\Omega (1 - \chi_A(\omega))(1 - \chi_A(T_1(\omega)))(1 - \chi_A(T_1^2(\omega)))d\mu \\
&- \int_\Omega (1 - \chi_A(T_1(\omega))(1 - \chi_A(T_1^2(\omega)))(1 - \chi_A(T_1^3(\omega)))d\mu \\
&+ \int_\Omega (1 - \chi_A(\omega))(1 - \chi_A(T_1(\omega)))(1 - \chi_A(T_1^2(\omega)))(1 - \chi_A(T_1^3(\omega)))d\mu.
\end{aligned}$$

In der Tat ergeben sich die ersten beiden Integrale zu

$$\int_\Omega \chi_A(\omega)(1 - \chi_A(T_1(\omega)))(1 - \chi_A(T_1^2(\omega)))d\mu$$

und die beiden letzten zu

$$-\int_\Omega \chi_A(\omega)(1 - \chi_A(T_1(\omega)))(1 - \chi_A(T_1^2(\omega)))(1 - \chi_A(T_1^3(\omega)))d\mu,$$

also in der Summe $\mu(A_3)$. Da nun T maßerhaltend ist, sind die beiden mittleren Integrale einander gleich und zwar exakt gleich f_3. Also

$$\mu(A_3) = f_2 - 2f_3 + f_4.$$

Für allgemeines n geht dies ebenso. Der Clou ist nun, dass in der Summe $\sum_{k=1}^{n} \mu(A_k)$ die meisten Terme verschwinden:

$$\sum_{k=1}^{n} \mu(A_k) = 1 - f_1 - (f_n - f_{n+1}).$$

Nun ist nach Konstruktion (f_n) eine fallende und (durch 0) von unten beschränkte Folge. Also existiert $\lim f_n$, und somit gilt

$$\lim_{n\to\infty} (f_n - f_{n+1}) = 0.$$

Also ergibt sich

$$\sum_{k=1}^{\infty} \mu(A_k) = 1 - f_1 = \mu(A).$$

Da die (A_k) konstruktionsgemäß paarweise disjunkt sind, gilt für fast jedes $\omega \in A$, dass wenigstens eine der iterierten Abbildungen $T_1(\omega),\ T_2(\omega), \ldots$ in A ist. Dies impliziert sofort, dass unendlich viele Iterationen in A landen. In der Tat: Sei D_ℓ die Menge aller ω, so dass $\omega \in A$ und $T^n\omega \notin A$ für alle $n \geq \ell$. Wendet man dann das soeben gezeigte auf T_1^ℓ anstelle von T_1 an, so ergibt sich

$$\mu(D_\ell) = 0$$

und somit auch

$$\mu(\bigcup_{\ell\geq 1} D_\ell) = 0.$$

Somit kehrt beinahe jeder Punkt aus A unter iterativer Abbildung durch T unendlich oft nach A zurück. □

Eine sehr naheliegende Frage, die, wie oben erwähnt, auch durch Boltzmann aufgeworfen wurde, ist die nach der durchschnittlichen Zeit, die eine solche Rückkehr benötigt (diese Zeit heißt auch die Länge eines Poincaré-Zyklus).

Für diskrete Zeit $t = 1, 2, \ldots$ ist diese definiert als

$$\Theta_1^* = \frac{1}{\mu(A)} \sum_{k=1}^{\infty} k\mu(A_k).$$

Misst man allgemeiner alle τ Zeiteinheiten und ersetzt so T_1 durch T_τ, so ist die mittlere Rückkehrzeit definiert als

$$\Theta_\tau^* = \frac{\tau}{\mu(A)} \sum_{k=1}^{\infty} k \cdot \mu(A_k^\tau),$$

wobei

$$A_k^\tau = A \cap (T_\tau^{-1}(A))^c \cap \ldots \cap (T_\tau^{-k-1}(A))^c \cap T_\tau^{-k}(A)$$

ist. Nun gilt nach dem eben gezeigten:

$$\sum_{k=1}^{n} k\mu(A_k) = \sum_{k=1}^{n} k(f_{k-1} - 2f_k + f_{k+1}) = 1 - f_n - n(f_n - f_{n+1}).$$

Nun ist die Summe links offenbar wachsend in n, also ist $f_n + n(f_n - f_{n+1})$ fallend. Außerdem ist diese Folge offensichtlich durch 0 von unten beschränkt, denn $f_n \geq 0$ und $f_n \geq f_{n+1}$. Somit existiert der Grenzwert

$$\lim_{n\to\infty} f_n + n(f_n - f_{n+1}),$$

also auch der Grenzwert

$$\lim_{n\to\infty} n(f_n - f_{n+1}).$$

Um den Grenzwert zu bestimmen, beobachten wir, dass

$$\sum_{n=1}^{\infty} (f_n - f_{n+1}) < +\infty,$$

da (f_n) konvergiert. Dies kann aber nur sein, wenn die einzelnen Summanden schneller als $\frac{1}{n}$ gegen 0 konvergieren, also

$$\lim_{n\to\infty} n(f_n - f_{n+1}) = 0.$$

Also ergibt sich

$$\sum_{k=1}^{\infty} k\mu(A_k) = 1 - \lim_{n\to\infty} f_n.$$

Ähnliches gilt für A_k^τ und ein geeignet definiertes f_n^τ. Nun nehmen wir an, dass T zu allen geforderten Eigenschaften auch ergodisch ist.

Erinnerung: T heißt ergodisch, falls aus

$$T(A) = A$$

folgt, dass

$$\mu(A) \in \{0, 1\}$$

ist.

Wir wollen zeigen, dass in diesem Fall

$$\lim_{n\to\infty} f_n = 0$$

gilt. In der Tat ist ja

$$f_n = \mu(B_n) = \mu(\{\omega : \omega \notin A, T_\tau(\omega) \notin A, \ldots, T_\tau^{n-1}(\omega) \notin A\}).$$

Da B_n fällt existiert

$$B = \bigcap_{n \geq 1} B_n.$$

Nun betrachte $T_\tau B$. Gilt

$$\omega \in T_\tau B,$$

so ist

$$T_\tau^{-1}(\omega) \in B,$$

also

$$T_\tau^{-1}(\omega) \notin A,\ \omega \notin A, T_\tau \omega \notin A, \ldots,$$

also ist $\omega \in B$, d. h.

$$T_\tau B \subseteq B.$$

Ebenso zeigt man, dass auch

$$T_\tau^2 B \subseteq T_\tau B \subseteq B$$

gilt. Sei

$$C = \lim_{n \to \infty} T_\tau^n B.$$

Offenbar gilt

$$T_\tau C = C.$$

Aus der Ergodizität von T ergibt sich, dass

$$\mu(C) \in \{0, 1\}$$

gilt. Da nun

$$\mu(T_\tau^n B) = \mu(B) \quad \text{und} \quad T_\tau^{n+1} B \subseteq T_\tau^n B$$

gilt, folgt

$$\mu(B) = \mu(C) \in \{0, 1\}.$$

Nun impliziert aber $\mu(B) = 1$, dass auch $\mu(B_1) = 1$ ist, also $\mu(\omega \notin A) = 1$, also $\mu(A) = 0$ im Widerspruch zu unseren Annahmen. Somit ist $\mu(B) = 0$, aber das heißt auch, dass

$$\lim_{n \to \infty} f_n = \mu(B) = 0$$

ist. Somit haben wir den folgenden Satz bewiesen.

Satz 2.2 *(Kac): Für bijektive, maßerhaltende, ergodische Abbildungen T gilt*

$$\Theta_\tau^* = \frac{\tau}{\mu(A)}$$

für alle $\tau > 0$.

Der Nachteil dieser Formel wird offenkundig, wenn man den Limes $\tau \to 0$ betrachtet, d. h. zu kontinuierlichen Systemen übergeht, dann erhalten wir stets

$$\lim_{\tau \to 0} \Theta_\tau^* = 0.$$

Dies ist natürlich deshalb der Fall, weil

$$\{\omega \in A,\, T_\tau \omega \in A\}$$

als eine Rückkehr zur Zeit τ gewählt wird (obschon dies für sehr kleine τ vermutlich keine echte Rückkehr sondern ein Verharren in A ist) und die Wahrscheinlichkeit dieses Ereignises für $\tau \to 0$ nahe bei 1 liegt. Einen Ausweg bietet eine Formel von Smoluchowski, der vorschlug, die mittlere Rückkehrzeit als

$$\Theta_\tau = \frac{\tau \sum_{k=1}^{\infty} k\mu(A_{k+1})}{\sum_{k=1}^{\infty} \mu(A_{k+1})}$$

zu definieren. Dies soll hier jedoch nicht weiter betrachtet werden.

2.3 Das Ehrenfestsche Urnenmodell

Durch die in Abschnitt 2.1 erwähnten Einwände von Loschmidt und Zermelo ist offenkundig, dass die naive Formulierung des Boltzmannschen Gesetzes, dass die Entropie stets wächst, unhaltbar ist. Boltzmann selbst sah, dass die Herleitung dieses Satzes auf statistischen Aussagen über die mittlere Anzahl von Kollisionen gewisser Art beruht (dem sogenannten "Stoßzahlansatz") und schlug daher vor, das Wachstum der Entropie als eine rein statistische Aussage zu betrachten.

1911 unterzogen P. und T. Ehrenfest in einem sehr lesenswerten Artikel (Enc. Math. Wiss.) die Boltzmannsche Argumentation einer tiefgründigen Analyse. Grob gesprochen stellen sich zwei Fragen:

1. Ist es möglich, dass sich zeitliche Reversibilität und Rekurrenz auf der Mikroebene mit einem zeitlich irreversiblen Verhalten im Großen vereinbaren lassen?
2. Ist diese Vereinbarkeit im Rahmen der klassischen Mechanik möglich?

Problem 2 ist bis heute nicht wirklich gelöst, während Problem 1 zunächst rein logischer Natur ist. Es ist (im positiven Sinne) gelöst, wenn man ein Modell findet, dass die geforderten Eigenschaften aufweist. Ein solches Modell ist das Ehrenfestsche Urnenmodell. Dies lässt sich folgendermaßen beschreiben:

$2R$ Kugeln, die von 1 bis $2R$ nummeriert sind, werden in eine Schachtel A gelegt. Zu diskreten Zeitpunkten $t = 1, 2, \ldots$ tut man mit Wahrscheinlichkeit $\frac{1}{2}$ gar nichts (dies nur, um die resultierende Markovkette aperiodisch zu machen) und mit Wahrscheinlichkeit $\frac{1}{2}$ zieht man eine Kugel und legt sie in eine zweite Schachtel B. So fährt man fort, wobei man jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ nichts tut und mit Wahrscheinlichkeit $\frac{1}{2}$ eine Kugel von $1, \ldots, 2R$ zieht und in die jeweils andere Schachtel legt.

Intuitiv ist klar, was geschieht. So lange die Anzahl der Kugeln in A – sagen wir n_A – sehr viel größer ist als R, sollten wir einen "Fluss" von Schachtel A nach Schachtel B beobachten; die Wahrscheinlichkeit, dass eine Kugel von A nach B wandert ist überwältigend viel größer als die Wahrscheinlichkeit, dass eine Kugel den umgekehrten Weg von B nach A nimmt. Wir haben somit global eine "beinahe irreversible" Bewegung.

Betrachten wir das ganze mathematisch: Sei $n_A(t)$ die Anzahl der Kugeln in A zur Zeit t. Um die zeitliche Reversibilität des Systems zu diskutieren, berechnen wir

$$\begin{aligned}&\mathbb{P}(n_A(t+1) = n | n_A(t) = m) \quad \text{und}\\ &\mathbb{P}(n_A(t-1) = n | n_A(t) = m).\end{aligned}$$

Die erste Wahrscheinlichkeit berechnet sich als

$$\mathbb{P}(n_A(t+1) = n | n_A(t) = m) = \frac{1}{4R} m \delta_{\{n=m-1\}} + \frac{1}{4R}(2R - m)\delta_{\{n=m+1\}} + \frac{1}{2}\delta_{\{n=m\}}$$

(wobei δ das Kronecker-Symbol ist). Die zweite ist etwas umständlicher herzuleiten.

$$\begin{aligned}&\mathbb{P}(n_A(t-1) = n | n_A(t) = m) = \frac{\mathbb{P}(n_A(t-1) = n, n_A(t) = m)}{\mathbb{P}(n_A(t) = m)}\\ = \;& \mathbb{P}(n_A(t) = m | n_A(t-1) = n) \cdot \frac{\mathbb{P}(n_A(t-1) = n)}{\mathbb{P}(n_A(t) = m)}\\ = \;& \left(\frac{n}{4R}\delta_{\{m=n-1\}} + \frac{2R-n}{4R}\delta_{\{m=n+1\}} + \frac{1}{2}\delta_{\{n=m\}}\right) \cdot \frac{\mathbb{P}(n_A(t-1) = n)}{\mathbb{P}(n_A(t) = m)}.\end{aligned}$$

Nun hängen die Wahrscheinlichkeiten $\mathbb{P}(n_A(t-1) = n)$ und $\mathbb{P}(n_A(t) = m)$ sowohl von t als auch von der anfänglichen Anzahl n_0 von Kugeln in A ab. Es ist jedoch plausibel und eine Konsequenz aus dem Ergodensatz für Markovketten, den wir gleich in Erinnerung rufen werden, dass diese Abhängigkeit für große t verschwindet. Hierzu rufe man sich vor Augen, dass $n_A(\mathrm{t})$ in t eine Markovkette ist, also ein stochastischer Prozess in diskreter Zeit mit Übergangswahrscheinlichkeiten

$$\begin{aligned}&\mathbb{P}(n_A(t) = n | n_A(t-1) = m, n_A(t-2) = n_{t-2}, \dots, n_A(0) = n_0)\\ = \;& \mathbb{P}(n_A(t) = n | n_A(t-1) = m) = \frac{m}{4R}\delta_{\{n=m-1\}} + \frac{2R-m}{4R}\delta_{\{n=m+1\}} + \frac{1}{2}\delta_{\{n=m\}}\\ := \;& Q(m,n).\end{aligned}$$

$(Q(m,n))_{m,n=0}^{2R}$ heißt die zu der Markovkette gehörige stochastische Matrix. Wie wir (einige von uns) schon in der Stochastikvorlesung festgestellt haben, genügen Markovketten dem folgenden Satz:

Satz 2.3 *(Ergodensatz für Markovketten): Es sei Q eine stochastische Matrix über einem endlichen Zustandsraum I und ν eine beliebige Anfangsverteilung auf I. Weiter existiere ein $N \in \mathbb{N}$, so dass Q^N nur strikt positive Einträge hat. Dann existiert eine Wahrscheinlichkeit ρ auf I, so dass*

$$\rho Q = \rho$$

gilt und es gilt

$$\nu Q^n \underset{n\to\infty}{\longrightarrow} \rho.$$

Desweiteren gilt für die Entropie

$$H(\nu|\rho) = \sum_{i\in I} \nu(i) \log(\frac{\nu(i)}{\rho(i)}),$$

dass für alle Verteilungen ν auf I die Entropie unter Anwendung von Q fällt:

$$H(\nu Q|\rho) \leq H(\nu|\rho).$$

Obschon wir den Beweis schon in der Vorlesung über Stochastik geführt haben, wollen wir hier an die wesentlichen Schritte erinnern.

Beweisskizze: Die Existenz eines ρ mit den geforderten Eigenschaften haben wir in der Stochastik aus Kompaktheitseigenschaften hergeleitet. Sie folgt aber auch aus Sätzen der linearen Algebra. Da Q stochastisch ist 1 ein Eigenwert von Q, denn die Konstanten sind rechte Eigenvektoren zum Eigenwert 1. 1 ist sogar der größte Eigenwert (betragsmäßig), denn es gilt für den betragsmäßig größten Eigenwert $\lambda_{\max}$ einer (nicht-negativen) Matrix $Q = (Q_{ij})_{j=1}^n$,

$$\frac{1}{n} \sum_{i,j} Q_{ij} \leq \lambda_{\max} \leq \max_j \sum_i Q_{ij}.$$

Aus dem Satz von Perron-Frobenius folgt somit, dass es einen linken Eigenvektor $\tilde{\rho}$ zum Eigenwert 1 gibt, der lauter strikt positive Einträge besitzt. Also ist

$$\rho := \frac{\tilde{\rho}}{\|\tilde{\rho}\|}$$

die gesuchte Wahrscheinlichkeit.

Außerdem ist ρ eindeutig, denn ist ρ' ein weiterer linker Eigenvektor zum Eigenwert 1, so auch $\rho - a\rho'$ mit

$$a = \min_i \frac{\rho(i)}{\rho'(i)} =: \frac{\rho(i_0)}{\rho'(i_0)}.$$

Damit ist aber

$$0 = (\rho - a\rho')(i_0) = \sum_{j \in I} (\rho - a\rho')(j) Q^N(j, i_0).$$

Hieraus folgt $\rho(j) = a\rho'(j)$ für alle $j \in I$, also $a = 1$, also $\rho = \rho'$.

Für die restlichen Beweisschritte schreiben wir die Entropie

$$H(\nu|\rho) = \sum_i \rho(i) \psi(\frac{\nu(i)}{\rho(i)}),$$

wobei ψ die strikt konvexe Funktion

$$\psi(t) = t \log t - t + 1$$

ist, $\psi(1) = 0$ und somit minimal genau dann, wenn $t = 1$ ist. Also ist

$$H(\nu|\rho) = 0 \Leftrightarrow \nu = \rho.$$

Daher ist

$$\begin{aligned} H(\nu Q|\rho) &= \sum_i \rho(i)\psi\left(\frac{\sum_j \nu(j)Q(j,i)}{\rho(i)}\right) \\ &\leq \sum_i \sum_j \rho(j)Q(j,i)\psi(\frac{\nu(j)}{\rho(j)}) \\ &= \sum_j \rho(j)\psi(\frac{\nu(j)}{\rho(j)}) = H(\nu|\rho). \end{aligned}$$

Dies zeigt, dass die Entropie unter Anwendung von Q fällt. Dies aber impliziert schon die Konvergenz der Folge νQ^n gegen ρ. In der Tat hat diese Folge als Folge in einer kompakten Teilmenge des $\mathbb{R}^{|I|}$ ja einen Häufungspunkt ρ' und eine Teilfolge (νQ^{n_ℓ}), die gegen ρ' konvergiert. Nun ist zum einen

$$H(\rho' Q|\rho) \leq H(\rho'|\rho),$$

andererseits

$$\begin{aligned} H(\rho' Q|\rho) &= \lim_{\ell\to\infty} \sum_j \rho(j)\psi\left(\frac{(\nu Q^{n_\ell})Q(j)}{\rho(j)}\right) \\ &= \lim_{\ell\to\infty} \sum_j \rho(j)\psi\left(\frac{\nu Q^{n_\ell+1}(j)}{\rho(j)}\right) \\ &\geq \lim_{\ell\to\infty} \sum_j \rho(j)\psi\left(\frac{\nu Q^{n_{\ell+1}}(j)}{\rho(j)}\right) = H(\rho'|\rho), \end{aligned}$$

also $H(\rho' Q|\rho) = H(\rho'|\rho)$. Aus der strikten Konvexität von ψ folgt nun, dass

$$\rho' = \rho$$

gilt. □

Wir wollen nun feststellen, dass die Markovkette $n_A(t)$ aus dem Ehrenfestschen Urnenmodell den Anforderungen des Ergodensatzes genügt.

Satz 2.4 *$n_A(t)$ genügt dem Ergodensatz.*

Beweis: Zunächst sei bemerkt, dass die Markovkette irreduzibel ist, d. h. für

$$n, m \in \{0, \ldots, 2R\}$$

existiert t, so dass

$$\mathbb{P}(n_A(t) = n | n_A(0) = m) > 0$$

gilt (in der Tat kann man $t \leq 2R$ wählen). Wegen der positiven Wahrscheinlichkeit, dass die Kette stehen bleibt, gilt sogar

$$\mathbb{P}(n_A(t) = n | n_A(0) = m) > 0$$

für alle $t \geq 2R$ und alle $n, m \in \{0, \dots, 2R\}$. Dies aber ist gleichbedeutend mit

$$Q^N \gg 0$$

für alle $N \geq 2R$. □

Das im Ergodensatz auftretende Limesmaß berechnet sich wie folgt:

Satz 2.5 *Sei ρ das invariante Maß der Markovkette $n_A(\cdot)$. Dann gilt*

$$\rho(m) = \binom{2R}{m} 2^{-2R}.$$

Beweis: Das rechnet man nach. □

Das Interessante daran ist, dass sich für dieses ρ die Identität

$$Q(n,m)\rho(n) = Q(m,n)\rho(m)$$

ableiten lässt. Vergegenwärtigen wir uns noch einmal, dass

$$Q(n,m) = \mathbb{P}(n_A(t) = m | n_A(t-1) = n)$$

(und eine analoge Formel für $Q(m,n)$) gilt und für große Zeiten t

$$\rho(n) = \mathbb{P}(n_A(t) = n) = \mathbb{P}(n_A(t-1) = n)$$

(und eine analoge Identität für $\rho(m)$), so lässt sich aus dieser Identität

$$\begin{aligned} \mathbb{P}(n_A(t-1) = n | n_A(t) = m) &= \frac{\rho(n)}{\rho(m)} \left(\frac{n}{4R} \delta_{\{n=m+1\}} + \frac{2R-n}{4R} \delta_{\{n=m-1\}} + \frac{1}{2} \delta_{\{n=n\}} \right) \\ = \; & \mathbb{P}(n_A(t+1) = n | n_A(t) = m) \end{aligned}$$

folgern.

In der Tat ist das Modell also (auf lange Sicht) zeitlich reversibel. Andererseits besagt der Ergodensatz für $n_A(\cdot)$, dass die Entropie als RELATIVE ENTROPIE bezüglich ρ stets fällt. Schließlich folgt aus der Irreduzibilität der Kette, dass sie jeden Zustand unendlich oft besucht. Wir haben also ein System gefunden, das den Anforderungen der ersten der aufgeworfenen Fragen genügt.

Bemerkung: In der Tat birgt das vorgestellte Modell allerdings eine kleine Falle: Die Berechnung der zeitlichen Reversibilität benutzte den Ergodensatz und vor allem die Approximation

$$\mathbb{P}(n_A(t) = m) \approx \rho(m) \quad \forall\, m.$$

Dies aber stimmt exakt nur, wenn sich das Modell im Gleichgewicht befindet, die Approximation also eine Gleichheit ist. Dann jedoch ist die relative Entropie konstant.

Wir wollen das Ehrenfestsche Urnenmodell hier verlassen und uns anderen Fragestellungen zuwenden, obschon es auch zu diesem Modell noch viel zu sagen gäbe.

Bemerkung: Eine letzte Bemerkung zum Boltzmannschen Argument, dass eine Rückkehr in einen unwahrscheinlichen Anfangszustand wesentlich länger dauert als die Beobachtungszeiträume, in denen ein Anwachsen der Entropie beobachtet wird. In der Tat kann man die folgende Überschlagsrechnung anstellen: Zöge man statt immer einer Kugel immer R Kugeln und ließe diese die Urnen tauschen, so wäre die Wahrscheinlichkeit, in einem Sprung von $n_A(t) = R$ zu $n_A(t+1) = 2R$ zu kommen,

$$\frac{R!}{(2R)^R} \approx \frac{\sqrt{R}}{(2e)^R} \leq \frac{1}{2^{2R}}.$$

Die Rückkehrzeit von einer mit der Hälfte der Kugeln gefüllten Urne zu einer vollen Urne sollte also mindestens 2^{2R} betragen. Bedenkt man, dass R für realistische Systeme von der Ordnung

$$R \approx 10^{23}$$

ist, so ist das gigantisch viel länger als jeder für natürliche Prozesse relevante Zeitraum.

3 Gittergase und Spinsysteme

Die ersten Objekte, die man Studien der statistischen Mechanik unterzog, waren sogenannte freie Gase, also Gase, bei denen man annimmt, dass die einzelnen Moleküle nicht miteinander wechselwirken. Dies entspricht auf der wahrscheinlichkeitstheoretischen Seite Systemen unabhängiger Zufallsvariablen. Ihr Studium verläuft recht ähnlich den Prinzipien großer Abweichungen, die wir schon in der Stochastik kennengelernt haben sowie einigen Schritten aus dem letzten Kapitel.

Wir werden uns daher auf etwas realistischere Systeme stürzen, die auch Wechselwirkungen zwischen den Teilchen ermöglichen. Um umgekehrt die Systeme nicht zu kompliziert werden zu lassen, beschränken wir die Anzahl ihrer Freiheitsgrade und lassen die Teilchen nur auf den Knoten eines Gitters leben. Dieses wird für die Zwecke dieser Vorlesung stets $\mathbb{Z}^d,\ d \geq 1$, sein.

3.1 Gittergase

Sei $\Lambda \subseteq \mathbb{Z}^d$ eine endliche Teilmenge. Wir identifizieren Λ mit seinen Knoten und stellen uns vor, dass in den Knoten von Λ Teilchen in verschiedenen Zuständen leben. Diese Zustände können z. B. "anwesend" $= 1$ und "abwesend" $= 0$ sein oder z. B. verschiedenen Magnetisierungen "spin up" $= +1$ und "spin down"$=-1$ entsprechen. Wir werden annehmen, dass nur Paare von Teilchen miteinander wechselwirken. Die Wechselwirkungsenergie zwischen Teilchen x_i an der Stelle $i \in \Lambda$ und Teilchen x_j an der Stelle $j \in \Lambda$ sei durch die Funktion

$$\varphi_{i,j}(x_i, x_j)$$

beschrieben. Die Gesamtenergie des Systems bestimmt sich dann als

$$H_\Lambda(x_1, \ldots, x_N) = \sum_{i \neq j \in \Lambda} \varphi_{i,j}(x_i, x_j).$$

Ein physikalisches Prinzip besagt, dass Systeme stets Zustände niedriger Energie aufsuchen und dass Zustände hoher Energie exponentiell unwahrscheinlich sind. Dem entspricht der Ansatz eines Gibbs-Maßes für die Partikelkonfiguration $(x_1, \ldots, x_N),\ N = |\Lambda|$, als

$$\mu_{\Lambda,\beta}(x_1, \ldots, x_N) = \frac{e^{-\beta H_V(x_1,\ldots,x_N)}}{Z_{\Lambda,\beta}}.$$

Hierbei ist $\beta > 0$ die inverse Temperatur, also $\beta = \frac{1}{T}$ und die bei den Physikern auftretende Boltzmann-Konstante k haben wir $k \equiv 1$ gesetzt. $Z_{\Lambda,\beta}$ heißt die Zustandssumme und muss sich als

$$Z_{\Lambda,\beta} = \sum_{x_1,\ldots,x_N \in \Lambda} e^{-\beta H_N(x_1,\ldots,x_N)}$$

schreiben lassen, um $\mu_{\Lambda,\beta}$ zu einem Wahrscheinlichkeitsmaß zu machen. Dies ist (im wesentlichen) die Standardform eines Gittergases. Allerdings sei hier erwähnt, dass man zumeist von Gittergasen im engeren Sinne nur dann spricht, wenn die Wechselwirkung von der Form

$$\varphi_{i,j}(x_i, x_j) = f(i,j)\delta_{x_i}\delta_{x_j}$$

ist. Hierbei ist $f(i,j)$ eine Funktion der Gitterplätze i und j und

$$\delta_{x_i} = \begin{cases} 1 & x_i = 1 \\ 0 & x_i = 0 \end{cases}$$

und die x_i nehmen Werte 0 und 1 ein. Man stellt sich also die jeweiligen Gitterplätze als "belegt" oder "frei" vor. Oft benötigt man noch einen "Korrekturfaktor", die chemische Energie, damit nicht die günstigsten Zustände diejenigen sind, in denen entweder $x_i = 1 \quad \forall\, i \in \Lambda$ oder $x_i = 0 \quad \forall\, i \in \Lambda$ gilt. Die Gesamtenergie nimmt dann die Form

$$H_V(x_1, \ldots, x_N) = \sum_{i \neq j \in \Lambda} f(i,j) x_i x_j - \frac{\mu}{\beta} \sum_{i \in \Lambda} x_i, \quad \mu \in \mathbb{R},$$

an.

3.2 Spin-Systeme

Im Gegensatz zu Gittergasen stellen wir uns bei Spin-Systemen alle Plätze von Λ als "belegt" vor, allerdings können die Teilchen dort in verschiedenen Zuständen sein.

Der Prototyp dieses Modells wurde 1924 von E. Ising in seiner Dissertation unter Anleitung von Lenz untersucht. Es ist ein Spielzeugmodell zur Erklärung magnetischer Phänomene. Eine fundamentale Erklärung solcher Phänomene müsste eine quantenmechanische Ableitung liefern und entsprechend Systeme von 10^{23} gekoppelten Schrödingergleichungen untersuchen.

Das sogenannte Ising-Modell wählt einen einfachen Ansatz: Es sei $\Lambda \subseteq \mathbb{Z}^d$ eine endliche Menge und $S = \{-1, +1\}$. Eine Spinkonfiguration σ ist ein Element

$$\sigma \in \{-1, +1\}^\Lambda = S^\Lambda.$$

Im Ising-Modell wird einem solchen $\sigma \in S^\Lambda$ die Wechselwirkungsenergie

$$H_{\Lambda,h}(\sigma) = -\sum_{\substack{i \neq j \in \Lambda \\ <i,j>}} \sigma_i \sigma_j - h \sum_{i \in \Lambda} \sigma_i$$

zugeordnet. Hierbei ist $h > 0$ die sogenannte Stärke des magnetischen Feldes und $\sigma \in S^\Lambda$ hat die Komponenten $\sigma_i,\ i \in \Lambda$ und $< i,j >$ bedeutet i und j sind Nachbarn. Es ist offensichtlich, dass sich das Ising-Modell in ein Gittergas übersetzen lässt vermöge der Transformation

$$x_i = \frac{\sigma_i + 1}{2}.$$

Wählt man statt $S = \{-1, +1\}$ eine andere Menge oder eine andere Interaktionsenergie, so erhält man ein allgemeines Spinsystem.

Eine wesentliche Größe bei all diesen Spinsystemen ist die sogenannte Magnetisierung

$$m_\Lambda = \frac{1}{|\Lambda|} \sum_{i \in \Lambda} \sigma_i.$$

Da sie, insbesondere im Fall des Ising-Modells, ein Maß für die Ordnung des Systems ist, spricht man auch von einem Ordnungsparameter. Analog zum Gittergas definiert man das Gibbs-Maß

$$\mu_{\Lambda,\beta,h}(\sigma) = \frac{e^{-\beta H_{\Lambda,h}(\sigma)}}{Z_{\Lambda,\beta,h}}$$

mit

$$Z_{\Lambda,\beta,h} = \sum_{\sigma' \in S^\Lambda} e^{-\beta H_{\Lambda,h}(\sigma')}.$$

Eine wesentliche Größe bei der Untersuchung solcher Systeme stellt die FREIE ENERGIE dar:

$$F_{\Lambda,\beta,h} = -\frac{1}{\beta} \log Z_{\Lambda,\beta,h}.$$

Wie A. Bovier in seinem Buch beschreibt, stellt das Ising-Modell einen Wendepunkt in der statistischen Mechanik dar, da es zum einen erstmals versucht, andere Phänomene zu erklären als die der Thermodynamik (nämlich den Ferromagnetismus) und zum anderen bereit ist, eine sehr komplizierte Ausgangssituation (elektromagnetische und quantenmechanische Wechselwirkung) durch einfachere Annahmen zu ersetzen. Beides kann man auch heute noch beobachten. Die Prinzipien der statistischen Mechanik finden heute Anwendung bei der Erklärung so verschiedenartiger Phänomene wie Gehirnleistung oder soziologischem Verhalten. Dabei setzt man oft auf sehr vereinfachte Modelle, in der Hoffnung, dass diese archetypisch sind für komplexere Modelle. Diese Annahmen nennt man UNIVERSALITÄTSHYPOTHESE.

Wir werden in dieser Vorlesung zunächst verschiedene magnetische Modelle diskutieren. Diese Diskussion beginnen wir mit einer kurzen Untersuchung der freien Energie.

3.3 Die Existenz der freien Energie

Wir wollen in diesem kurzen Abschnitt zeigen, dass im Ising-Modell (und vielen anderen Modellen) die freie Energie pro Punkt

$$-\frac{1}{\beta|\Lambda|} \log Z_{\Lambda,\beta,h}$$

für $|\Lambda| \to \infty$ in einem geeigneten Sinne existiert. Genauer beweisen wir

Satz 3.1 *Falls die freie Energie $F_{\Lambda,\beta,h}$ eines Systems subadditiv ist, d. h. falls*

$$-F_{\Lambda,\beta,h} \leq -F_{\Lambda_1,\beta,h} - F_{\Lambda_2,\beta,h}$$

für $\Lambda = \Lambda_1 \cup \Lambda_2$ und $\Lambda_1 \cap \Lambda_2 = \emptyset$, $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^d$ endlich gilt, so existiert für jede Folge von Rechtecken $\Lambda_1 \subseteq \Lambda_2 \subseteq \Lambda_n \subseteq \ldots$ mit $\Lambda_n \uparrow \mathbb{Z}^d$, d. h. für jede beschränkte Menge $A \subseteq \mathbb{Z}^d$ existiert ein n mit $A \subseteq \Lambda_n$, der Limes

$$\lim_{n\to\infty} \frac{1}{|\Lambda_n|} F_{\Lambda_n,\beta,h} = f_{\beta,h}.$$

Ist ferner

$$\frac{\sup_\sigma H_\Lambda(\sigma)}{|\Lambda|} \geq C > -\infty$$

für alle endlichen $\Lambda \subseteq \mathbb{Z}^d$, so ist $f_{\beta,h}$ endlich.

Beweis: Dies folgt aus dem folgenden Lemma (bzw. einer einfachen mehrdimensionalen Fassung), das wir schon aus der Wahrscheinlichkeitstheorie I kennen. □

Lemma 3.2 *Ist (a_n) eine Folge mit*

$$a_{n+m} \leq a_n + a_m,$$

dann existiert $\lim \frac{1}{n} a_n$ und es gilt

$$\lim \frac{1}{n} a_n = \inf \frac{1}{n} a_n.$$

Korollar 3.3 *Die freie Energie pro Punkt $f_{\beta,h}$ existiert im Ising-Modell. Sie ist endlich.*

Beweis: Wir behandeln nur den Fall periodischer Randbedingungen, d. h.

$$H_\Lambda(\sigma) = -\sum_{<i,j>} \sigma_i \sigma_j - h \sum_{i\in\Lambda} \sigma_i.$$

wobei $< i,j >$ jetzt bedeutet, dass i und j in dem periodisch fortgesetzten Gitter Λ Nachbarn sind. Definiere

$$\bar{H}(\sigma) = \bar{H}_\Lambda(\sigma) = \frac{1}{2} \sum_{\substack{i,j\in\Lambda \\ <i,j>}} (\sigma_i - \sigma_j)^2 - h \sum_{i\in\Lambda} \sigma_i.$$

Dann gilt

$$H(\sigma) - \bar{H}(\sigma) = -2d|\Lambda|,$$

da zu jedem Punkt ein $4d\sigma_i^2$ hinzukommt, wobei d die Dimension des Systems ist. Also gilt für die zugehörigen freien Energien F und $\bar{F}$

$$F - \bar{F} = -2d|\Lambda|.$$

Ist also $\bar{F}$ subadditiv, so auch F. $\bar{F}$ aber ist subadditiv, denn für seine Zustandssumme $\bar{Z}$ gilt:

$$\bar{Z}_{\Lambda,\beta,h} = \sum_{\sigma_i, i\in\Lambda_1} \sum_{\tau_j, j\in\Lambda_2} \exp\left(-\beta(\bar{H}_{\Lambda_1}(\sigma) + \bar{H}_{\Lambda_2}(\tau))\right) \times \exp(-\beta \sum_{\substack{i\in\Lambda_1, j\in\Lambda_2 \\ <i,j>}} (\sigma_i - \tau_j)^2),$$

wenn $\Lambda = \Lambda_1 \cup \Lambda_2$ und $\Lambda_1 \cap \Lambda_2 = \emptyset$ ist. Also ist

$$\bar{Z}_{\Lambda,\beta,h} \leq \bar{Z}_{\Lambda_1,\beta,h} \bar{Z}_{\Lambda_2,\beta,h}$$

und somit

$$-\bar{F}_{\Lambda,\beta,h} \leq -\bar{F}_{\Lambda_1,\beta,h} - \bar{F}_{\Lambda_2,\beta,h},$$

was zu zeigen war. Schließlich ist $f_{\beta,n}$ endlich, weil

$$H_\Lambda(\sigma) \geq -4|\Lambda|$$

für alle σ. □

Bemerkung: Man kann sogar auf die strenge Subadditivität verzichten. Dies wird z. B. in Simons Buch gut erklärt.

3.4 Das eindimensionale Ising-Modell

Wir kommen nun zur Diskussion des Ising-Modells in einer Dimension. Dies war der zentrale Untersuchungsgegenstand von Isings Dissertation. Er fand (richtigerweise), dass dieses Modell in $d = 1$ keine magnetische Phase aufweist und übertrug dieses Argument (fälschlicherweise) auf höhere Dimensionen. Es ist gewissermaßen eine Ironie der Geschichte, dass Ising nicht nur für diese nur halbrichtige Arbeit promoviert wurde, sondern dass dieses Modell auch heute noch seinen Namen trägt.

Betrachten wir noch einmal das Ising-Modell bzw. das zugehörige Gibbs-Maß

$$\mu_{\Lambda,\beta,h}(\sigma) = \frac{e^{\beta \sum\limits_{<i,j>\in\Lambda} \sigma_i\sigma_j + \beta h \sum\limits_{i\in\Lambda} \sigma_i}}{Z_{\Lambda,\beta,h}}.$$

Mit ein wenig Nachdenken sieht man, warum dieses Modell überhaupt ein interessantes Modell für Magnetismus sein kann: Es hat die Chance sich für große β (kleine Temperaturen) grundsätzlich anders zu verhalten als für kleine β (hohe Temperaturen). In der Tat bevorzugt der Energieterm im Gibbs-Maß Zustände mit möglichst viel gleichgerichteten Spins. Andererseits gibt es davon auch nur exponentiell wenig (aus Entropiegründen). Wir haben es also mit einer Art Wettkampf zwischen Energie und Entropie zu tun. Da erstere einen Faktor β mit sich führt, letztere aber nicht, könnte es sein, dass manchmal die Energie gewinnt (Zustände großer Ordnung haben große Wahrscheinlichkeit) und manchmal die Entropie (Zustände hoher Unordnung sind am wahrscheinlichsten). Wir würden daher für genügend große β einen Sprung der Magnetisierung für $h = 0$ und $h \neq 0$ erwarten. Einen solchen Sprung nennt man Phasenübergang. Leider tritt er im eindimensionalen Ising-Modell nicht auf. Wir schreiben in einer Dimension

$$H_N(\sigma) = -\sum_{i=1}^{N} \sigma_i\sigma_{i+1} - h\sum_{i=1}^{N} \sigma_i,$$

wobei es ein kleines Problem mit den Punkten $i = 1$ und $i = N$ gibt, da diese als einzige nur einen Nachbarn besitzen. Wir werden in höher dimensionalen Modellen noch sehen, dass dies in der Tat einen wesentlichen Punkt berührt. Hier werden wir die einfachste

Lösung wählen und die Punkte $1, \ldots, N$ auf einem Kreis anordnen, so dass die Addition in der Berechnung von H_N "modulo N" zu nehmen ist. Dann berechnet sich

$$Z_{N,\beta,h} = \sum_{\substack{\sigma_i=\pm 1\\ i=1,\ldots,N}} e^{\beta \sum \sigma_i \sigma_{i+1} + \beta h \sum \sigma_i} = \sum_{\substack{\sigma_i=\pm 1\\ i=1,\ldots,N}} \prod_{i=1}^{N} e^{\beta \sigma_i \sigma_{i+1} + \beta h \sigma_i}.$$

Schreiben wir für $s, s' \in \{-1, +1\}$

$$L(s, s') = e^{\beta s s' + \beta h s}$$

und fassen dies als eine 2×2 Matrix auf, die wir L nennen, so ist

$$Z_{N,\beta,h} = \sum_{\substack{\sigma_i=\pm 1\\ i=1,\ldots,N}} L(\sigma_1, \sigma_2), L(\sigma_2, \sigma_3) \ldots L(\sigma_N, \sigma_1) = tr L^N.$$

Aber die Spur von L^N lässt sich berechnen: Hat L die Eigenwerte λ_1 und λ_2, so ist

$$tr L^N = \lambda_1^N + \lambda_2^N.$$

Nun berechnen sich λ_1 und λ_2 als

$$\begin{aligned} \lambda_1 &= e^{\beta} \cosh(\beta h) + \sqrt{e^{2\beta} \sinh^2(\beta h) + e^{-2\beta}} \\ \lambda_2 &= e^{\beta} \cosh(\beta h) - \sqrt{e^{2\beta} \sinh^2(\beta h) + e^{-2\beta}}. \end{aligned}$$

Da $\lambda_1 > \lambda_2$, folgt

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N} \log Z_{N,\beta,h} = \log \lambda_1 &= \log \left(e^{\beta} \cosh(\beta h) + \sqrt{e^{2\beta} \sinh^2(\beta h) + e^{-2\beta}} \right) \\ &= \beta + \log \left(\cosh(\beta h) + \sqrt{\sinh^2(\beta h) + e^{-4\beta}} \right). \end{aligned}$$

Also

$$f_{\beta,h} = -1 - \frac{1}{\beta} \log \left(\cosh(\beta h) + \sqrt{\sinh^2(\beta h) + e^{-4\beta}} \right).$$

Nun ist nach dem Satz von der majorisierten Konvergenz

$$\begin{aligned} -\frac{\partial}{\partial h} f_{\beta,h} &= \lim_{N\to\infty} \frac{1}{\beta N} \frac{\sum_\sigma \frac{\partial}{\partial h} e^{\beta \sum \sigma_i, \sigma_{i+1} + \beta h \sum \sigma_i}}{Z_{N,\beta,h}} \\ &= \lim_{N\to\infty} \mathbb{E}_\mu \frac{1}{N} \sum \sigma_i =: m, \end{aligned}$$

wobei $\mathbb{E}_\mu$ den Erwartungswert bzgl. $\mu_{\beta,N,h}$ bezeichnet. Wir wollen diese Größe die asymptotische Magnetisierung nennen. Nun ist aber im eindimensionalen Ising-Modell

$$m = -\frac{\partial f_{\beta,h}}{\partial h} = \frac{\sinh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta}}}.$$

Diese Größe ist für alle $0 \leq \beta < \infty$ differenzierbar und monoton in h. Ein magnetisches Verhalten für $h \to 0$ ist daher nicht erkennbar. Daraus folgerte Ising (und folgern auch wir), dass das Ising-Modell in einer Dimension nicht die gewünschten Effekte zeigt. Bevor wir nun höherdimensionale Versionen des Ising-Modells studieren, wollen wir ein Modell vorführen, das die gewünschten Effekte sehr wohl aufweist.

3.5 Das Curie-Weiss-Modell

1896 entdeckte Pierre Curie, dass Ferromagnetismus ein Tieftemperaturphänomen ist: oberhalb einer gewissen Temperatur, der sogenannten Curie-Temperatur, verschwindet das Phänomen. Auf Basis dieser Entdeckung entwickelte Weiss 1907 eine Theorie des Ferromagnetismus. Vom Gesichtspunkt des Ising-Modells lässt es sich folgendermaßen beschreiben: Ein "natürliches" Modell des Ferromagnetismus würde nicht nur eine nächste Nachbarschaftswechselwirkung wie im reinen Ising-Modell in Betracht ziehen, sondern eine Wechselwirkung von σ_i mit allen anderen σ_j, wobei die Stärke der Wechselwirkung von $|i-j|$ abhinge und beispielsweise quadratisch mit $|i-j|$ abnähme. Dies ist jedoch ziemlich komplex – bislang haben wir ja noch nicht einmal das Ising-Modell in $d \geq 2$ gelöst. Daher könnte man auf die Idee kommen, die Wechselwirkung von σ_i mit allen anderen Spins durch die Wechselwirkung von σ_i mit einem Spin der Magnetisierung $\frac{1}{N}\sum \sigma_j$ zu ersetzen. Dies führt zu der Energie- bzw. Hamiltonfunktion

$$H_N(\sigma) = -\frac{1}{2N} \sum_{1 \leq i,j \leq N} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i.$$

Hierbei ist $\sigma \in \{-1,+1\}^N$. Zu beachten ist, dass man dieses Modell zwar auf $\mathbb{Z}^d$ definieren kann, dass aber die Geometrie von $\mathbb{Z}^d$ hier keine Rolle mehr spielt: Die Nachbarschaftsstruktur von i und $j \in \mathbb{Z}^d$ ist im Curie-Weiss-Modell bedeutungslos. Wir kommen zu einem besonderen Vorteil des Curie-Weiss-Modells. Schon im vorigen Kapitel haben wir die Bedeutung der Magnetisierung gesehen. Wir definieren:

$$m_N(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i.$$

Der große Vorteil des Curie-Weiss-Modells ist, dass die Energiefunktion eine Funktion des Ordnungsparameters m_N ist

$$H_N(\sigma) = -\frac{N}{2}[m_N(\sigma)]^2 - hN\, m_N(\sigma).$$

Diese Hamiltonfunktion hängt also von der mittleren Magnetisierung ab, man spricht auch von einem meanfield-Modell, einem Mittelwertmodell. Das zugehörige Gibbs-Mass ist von der Form

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{\frac{\beta N}{2} m_N^2(\sigma) + N\beta h m_N(\sigma)}}{Z_{N,\beta,h}}$$

mit

$$Z_{N,\beta,h} = \sum_{\sigma} e^{\frac{\beta N}{2} m_N^2(\sigma) + N\beta h m_N(\sigma)}.$$

Um das Gibbs-Maß und die Zustandssumme berechnen zu können, erinnern wir an die Theorie großer Abweichungen.

Definition 3.4 *Eine Folge $\mathbb{R}^d$-wertiger Zufallsvariablen $X_1, X_2, \ldots$ genügt einem Prinzip der großen Abweichungen mit Geschwindigkeit N und Ratenfunktion I, falls*

- *I von unten halbstetig ist und $0 \leq I \leq +\infty$, $I \not\equiv 0$.*
- *I kompakte Niveaumengen hat, d. h.*

$$N_L = \{x : I(x) \leq L\}$$

  ist kompakt für alle $L > 0$.
- *Für alle $A \in B^1$ gilt:*

$$-\inf_{x\in \mathring{A}} I(x) \leq \liminf_{x\in \mathring{A}} \frac{1}{N} \log \mathbb{P}(X_N \in \mathring{A}) \leq \limsup_{X\in \bar{A}} \frac{1}{N} \log \mathbb{P}(X_N \in \bar{A}) \leq -\inf_{X\in \bar{A}} I(x).$$

  Hierbei bezeichnen $\mathring{A}$ resp. $\bar{A}$ das topologisch Innere bzw. den topologischen Abschluss der Menge A.

Beispiel: Ein erstes Beispiel für ein Prinzip der großen Abweichungen (LDP) haben wir schon in der Stochastikvorlesung kennengelernt. Ist $X_1, X_2, \ldots$ eine Folge von i.i.d. Zufallsvariablen mit

$$\mathbb{P}(X_1 = 0) = 1 - p = 1 - \mathbb{P}(X_1 = 1),$$

so genügt die Folge der Mittelwerte

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

einem LDP mit Geschwindigkeit n und Ratenfunktion

$$H(x|p) = x \log \frac{x}{p} + (1 - x) \log \frac{1 - x}{1 - p}.$$

Der Beweis dieses Sachverhalts beruht i. w. auf der Anwendung der Stirling-Formel auf die Fakultäten der Binomialverteilung. Konzentriert man sich auf den Fall $p = \frac{1}{2}$, so hat die Rate die Form

$$H(x|\frac{1}{2}) = \log \frac{p}{1/2} + (1 - p) \log \frac{1 - p}{1/2} = p \log p + (1 - p) \log(1 - p) + \log 2.$$

Betrachtet man anstelle der (X_i) eine Folge (Y_i) mit

$$\mathbb{P}(Y_i = 1) = \mathbb{P}(Y_i = -1) = \frac{1}{2},$$

so entspricht dies einer Transformation

$$Y_i = 2X_i - 1.$$

Somit erhalten wir

$$\mathbb{P}(\frac{\sum_{i=1}^{n} Y_i}{n} = y) = \mathbb{P}(2S_n - 1 = y) = \mathbb{P}(S_n = \frac{y + 1}{2}).$$

Wir sehen also, dass mit S_n auch $\frac{\sum_{i=1}^n Y_i}{n}$ einem LDP genügt und zwar mit Rate

$$\begin{aligned} I(y) &= H(\frac{y+1}{2}|\frac{1}{2}) = \frac{y+1}{2}\log\frac{y+1}{2} + \frac{1-y}{2}\log\frac{1-y}{2} + \log 2 \\ &= \frac{(1+y)}{2}\log\frac{1+y}{2} + \frac{(1-y)}{2}\log(1-y). \end{aligned}$$

Hinter dieser Überlegung steckt sogar ein allgemeines Prinzip, das ebenso elementar zu beweisen wie (oft) nutzlos ist.

Satz 3.5 *(Kontraktionsprinzip) Es sei (X_n) eine Folge von Zufallsvariablen im $\mathbb{R}^d$, die einem LDP mit Geschwindigkeit n und Rate $I(\cdot)$ genüge. Es sei*

$$f : \mathbb{R}^d \to \mathbb{R}^m$$

stetig. Dann genügt auch die Folge

$$Y_n = f(X_n)$$

einem LDP mit Geschwindigkeit n und Ratenfunktion

$$J(y) = \inf_{x: f(x)=y} I(x).$$

Beweis: Es sei $A = \bar{A} \subseteq \mathbb{R}^m$. Dann ist

$$\limsup \frac{1}{n}\log \mathbb{P}(Y_n \in \bar{A}) = \limsup \frac{1}{n}\log \mathbb{P}(X_n \in f^{-1}[\bar{A}]) \leq - \inf_{x\in f^{-1}[\bar{A}]} I(x).$$

Eine untere Abschätzung geht analog. □

Eine weit wichtigere Konsequenz aus einem LDP ist das sogenannte Varadhansche Lemma, das es erlaubt, gewisse Integrale asymptotisch zu berechnen.

Satz 3.6 *(Varadhan): Es sei $(X_n)_n$ eine Folge von Zufallsvariablen in $\mathbb{R}^d$, die einen LDP mit Geschwindigkeit n und Rate I genügt und*

$$f : \mathbb{R}^d \to \mathbb{R}$$

sei stetig und beschränkt. Dann gilt

$$\lim \frac{1}{n}\log \mathbb{E}e^{nf(X_n)} = \sup_x [f(x) - I(x)]$$

(wobei die Existenz des Limes gleich mitbehauptet wird).

Der Beweis dieses Satzes ist recht instruktiv:

Beweis: Da f stetig und beschränkt ist, lässt sich eine Überdeckung des $\mathbb{R}^d$ durch endlich viele abgeschlossene Mengen $A_1, \ldots, A_M$ finden, so dass f auf jede dieser Mengen höchstens um ein vorgegebenes $\delta > 0$ variiert, also

$$\sup_{x \in A_i} f(x) - \inf_{x \in A_i} f(x) \le \delta \quad \forall\, i = 1, \ldots, M.$$

Dann folgt

$$\begin{aligned}
\int_{\mathbb{R}^d} e^{nf(X_n)} d\mathbb{P} &\le \sum_{j=1}^M \int_{A_j} e^{nf(X_n)} d\mathbb{P} \\
&\le \sum_{j=1}^M \int_{A_j} e^{n \sup_{y \in A_j} f(y)} d\mathbb{P}_{X_n}(y) \\
&\le \sum_{j=1}^M \int_{A_j} e^{n(\inf_{y \in A_j} f(y) + \delta)} d\mathbb{P}_{X_n}(y) \\
&= \sum_{j=1}^M e^{n(\inf_{y \in A_j} f(y) + \delta)} \mathbb{P}(X_n \in A_j).
\end{aligned}$$

Also

$$\begin{aligned}
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} &\le \sup_{1 \le j \le M} [\inf_{y \in A_j} f(y) + \delta - \inf_{y \in A_j} I(y)] \\
&\le \sup_{1 \le j \le M} [\sup_{y \in A_j} (f(y) - I(y)) + \delta] \\
&= \sup_{y \in \mathbb{R}^d} (f(y) - I(y)) + \delta.
\end{aligned}$$

Lässt man δ gegen 0 gehen, ergibt sich die obere Schranke. Für die untere Schranke argumentiert man, wie oft in der Theorie der großen Abweichungen, lokal. Da f stetig ist, gibt es zu jedem $y_0 \in \mathbb{R}^d$ und $\varepsilon > 0$ eine offene Umgebung $U_\varepsilon(y_0)$, so dass für alle $y \in U_\varepsilon(y_0)$ gilt

$$f(y) \ge f(y_0) - \varepsilon.$$

Somit folgt

$$\mathbb{E} e^{nf(X_n)} \ge \int_{U_\varepsilon(y_0)} e^{nf(X_n)} d\mathbb{P} \ge e^{n(f(y_0) - \varepsilon)} \mathbb{P}(X_n \in U_\varepsilon(y_0)).$$

Mit Hilfe des LDP ergibt sich

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} \ge f(y_0) - \varepsilon - \inf_{y \in U_\varepsilon(y_0)} I(y) \ge f(y_0) - I(y_0) - \varepsilon.$$

Da dies für alle y_0 stimmt, folgt

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E} e^{nf(X_n)} \ge \sup_y [f(y) - I(y)].$$

$\square$

Ganz analog zu Varadhans Lemma lässt sich der folgende Satz über große Abweichungen beweisen:

Satz 3.7 *Die Folge von Zufallsvariablen* (X_n) *im* $\mathbb{R}^d$ *genüge einem LDP mit Geschwindigkeit* n *und Ratenfunktion* I. *Für eine stetige beschränkte Funktion* $F : \mathbb{R}^d \to \mathbb{R}$ *setze*

$$J_n(S) = \int_S e^{nF(x)} d\mathbb{P}_{X_n}(x), \quad S \subseteq \mathbb{R}^d \text{ Borelsch.}$$

Definiere weiter

$$\mathbb{P}_n^F(S) = \frac{J_n(S)}{J_n(\mathbb{R}^d)}; \ S \in \mathcal{B}^d$$

die $\mathbb{P}_n^F$ *sind Wahrscheinlichkeitsmaße. Dann genügt die Folge der* $(\mathbb{P}_n^F)_n$ *einem LDP mit Geschwindigkeit* n *und Ratenfunktion*

$$I^F(x) = -[F(x) - I(x)] + \sup_{y \in \mathbb{R}^d} [F(y) - I(y)].$$

Beweis: Es ist

$$\limsup \frac{1}{n} \log \mathbb{P}_n^F(S) = \limsup \frac{1}{n} \log J_n(S) - \limsup \frac{1}{n} \log J_n(\mathbb{R}^d)$$

und Analoges für liminf. Nun wissen wir aus dem vorhergehenden Satz schon, dass

$$\lim \frac{1}{n} \log J_n(\mathbb{R}^d) = \sup_y [F(y) - I(y)]$$

gilt. Somit bleibt zu zeigen, dass für alle offenen Mengen G und alle abgeschlossenen Mengen A gilt

$$\liminf \frac{1}{n} \log J_n(G) \geq - \inf_{x \in G} [F(x) - I(x)]$$

und

$$\limsup \frac{1}{n} \log J_n(A) \leq - \inf_{x \in A} [F(x) - I(x)].$$

Dies aber geht genau wie im Beweis des Varadhanschen Lemmas. □

Mit diesen beiden Sätzen gerüstet, können wir nun das Curie-Weiss-Modell eingehender studieren.

Satz 3.8 *Für jede inverse Temperatur* $\beta > 0$ *und jedes magnetische Feld* h *gilt*

$$\begin{aligned} f_{\beta,h} := \lim_{N \to \infty} -\frac{1}{\beta N} \log Z_{N,\beta,h} &= \inf_{m \in [-1,+1]} [-\frac{m^2}{2} - hm - \frac{1}{\beta}(\log 2 - I(m))] \\ &= - \sup_{m \in [-1,+1]} [\frac{m^2}{2} + hm + \frac{1}{\beta}(\log 2 - I(m))]. \end{aligned}$$

Hierbei ist

$$I(m) = \frac{1+m}{2} \log(1+m) + \frac{1-m}{2} \log(1-m).$$

Beweis: Es ist

$$Z_{N,\beta,h} = \sum_{\substack{\sigma_i \in \{\pm 1\} \\ \forall i=1,\dots,N}} e^{\frac{\beta N}{2}(\frac{\sum_{i=1}^N \sigma_i}{N})^2 + N\beta h(\frac{\sum_{i=1}^N \sigma_i}{N})} = 2^N \frac{1}{2^N} \sum_{\substack{\sigma_i \in \{\pm 1\} \\ \forall i=1,\dots,N}} e^{\frac{\beta N}{2}(\frac{\sum_{i=1}^N \sigma_i}{N})^2 + N\beta h(\frac{\sum_{i=1}^N \sigma_i}{N})}.$$

(Dies ist der Grund für das Auftreten des log 2-Terms im Ausdruck für $f_{\beta,h}$). Nun haben wir schon in einem Beispiel geklärt, dass $\frac{\sum_{i=1}^N \sigma_i}{N}$ unter dem Produktmaß einem LDP mit Geschwindigkeit N und Ratenfunktion $I(\cdot)$ genügt. Nun ist

$$F(x) = \frac{\beta x^2}{2} + \beta h x$$

natürlich im allgemeinen zwar stetig, aber nicht beschränkt. Nun lebt $\frac{\sum_{i=1}^N \sigma_i}{N}$ aber auf $[-1, +1]$. Dort ist $F(\cdot)$ sehr wohl beschränkt. Eine Anwendung des Varadhanschen Lemmas ergibt somit

$$\lim \frac{1}{N} \log(2^{-N} Z_{N,\beta,h}) = \sup_{m \in [-1,+1]} [\frac{\beta m^2}{2} + \beta h m - I(m)].$$

Dies ist äquivalent zu unserer Behauptung. □

Ebenso lässt sich zeigen, dass die Magnetisierung

$$m_N := \frac{1}{N} \sum_{i=1}^N \sigma_i$$

unter dem Gibbs-Maß

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{-\beta H_N(\sigma)}}{Z_{N,\beta,h}}$$

einem LDP genügt.

Satz 3.9 *$m_N(\cdot)$ genügt unter dem Gibbs-Maß $\mu_{N,\beta,h}$ einem LDP mit Geschwindigkeit N und Ratenfunktion*

$$J(x) = -\frac{\beta x^2}{2} - \beta h x + I(x) + \sup_{y \in [-1,1]} [\frac{\beta y^2}{2} + \beta h y - I(y)],$$

wobei wieder

$$I(x) = \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x)$$

ist.

Beweis: Dies folgt direkt aus den Sätzen über große Abweichungen, die wir zuvor bewiesen haben, da $\mu_{N,\beta,h}$ genau die dortige exponentielle Struktur besitzt. □

Das Schöne an Prinzipien der großen Abweichungen ist, dass sie auch Gesetze der großen Zahlen implizieren. Dieses Faktum haben wir schon in Wahrscheinlichkeitstheorie II kennengelernt. Es soll hier kurz wiederholt werden.

Satz 3.10 *Eine Folge von Zufallsvariablen* $(X_n)_n$ *in* $\mathbb{R}^d$ *genüge einem LDP mit Rate* $I(\cdot)$ *und Geschwindigkeit* n*. Dann gilt für die Menge der Nullstellen*

$$\mathcal{N} = \{x \in \mathbb{R}^d : I(x) = 0\}$$

der Ratenfunktion und jedes $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) < +\infty.$$

Hierbei ist

$$\mathcal{N}_\varepsilon = \{x : \|x - \mathcal{N}\| < \varepsilon\}.$$

Ist insbesondere $\mathcal{N}$ *einelementig, so folgt für* $\nu \in \mathcal{N}$

$$\mathbb{P}(X_n \to \nu) = 1.$$

Beweis: Da $\mathcal{N}_\varepsilon^c$ abgeschlossen ist, I von unten halbstetig und $\mathcal{N}$ die Menge der globalen Minima von I ist, folgt

$$a := \inf_{x \in \mathcal{N}_\varepsilon^c} I(x) > 0.$$

Aus der oberen Abschätzung der großen Abweichungen folgt, dass für hinreichend große n

$$\mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) \leq e^{-na/2}$$

gilt. Somit ist

$$\sum_n \mathbb{P}(X_n \in \mathcal{N}_\varepsilon^c) < +\infty.$$

Die fast sichere Konvergenz ist eine unmittelbare Konsequenz dieser Summierbarkeit und des Borel-Cantelli-Lemmas. □

Wir müssen uns somit, um die Minima von

$$J(x) = -\frac{\beta x^2}{2} - \beta h x + I(x) + \sup_{y \in [-1,+1]} [\frac{\beta y^2}{2} + \beta h y - I(y)]$$

mit

$$I(x) = \frac{1+x}{2} \log(1+x) + \frac{1-x}{2} \log(1-x)$$

kümmern. Diese Minima erfüllen also

$$I'(x) = \frac{1}{2} \log \frac{1+x}{1-x} = \beta(x+h),$$

also

$$e^{2\beta(x+h)} = \frac{1+x}{1-x}.$$

Durch langes Hinschauen erkennt man, dass dies äquivalent ist zu

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

Man unterscheidet nun verschiedene Fälle:

- Ist $h > 0$, hat diese Gleichung zwei Lösungen, von denen aber nur die positive ein Minimum von $J(\cdot)$ liefert (die andere ein Maximum).
- Für $h < 0$ gibt es ebenfalls zwei Lösungen, von denen aber nur die negative ein Minimum, die positive aber ein Maximum ist.
- Für $h = 0$ ist die Situation symmetrisch zum Ursprung. Für $\beta \leq 1$ ist $x = 0$ die einzige Lösung dieser Gleichung und liefert somit ein Minimum von $J(\cdot)$. Für $\beta > 1$ hat die Gleichung allerdings drei Lösungen, von denen die Lösung $x = 0$ diesmal ein Maximum von $J(\cdot)$ liefert, die Lösungen, die verschieden sind von 0 aber Minima.

Zusammen erhält man:

Satz 3.11 *Im Curie-Weiss-Modell gelten für die Magnetisierung m_N die folgenden Grenzwertsätze unter $\mu_{N,\beta,h}$.*

1. *Ist $h > 0$, so konvergiert m_N exponentiell schnell gegen die positive Lösung von*
$$x = \tanh(\beta(x + h)).$$
2. *Ist $h < 0$, so konvergiert m_N exponentiell schnell gegen die negative Lösung von*
$$x = \tanh(\beta(x + h)).$$
3. *Ist $h = 0$ und $\beta \leq 1$, so konvergiert m_N exponentiell schnell gegen 0.*
4. *Ist $h = 0$ und $\beta > 1$, so konvergiert m_N exponentiell schnell gegen die beiden von Null verschiedenen Lösungen von*
$$x = \tanh(\beta x).$$
5. *Insbesondere erhält man*
$$\lim_{h \downarrow 0} \lim_{N \to \infty} \mu_{N,\beta,h} \circ m_N^{-1} \Rightarrow \delta_{m^*(\beta)},$$
wobei $m^*(\beta)$ die größte Lösung von
$$x = \tanh(\beta x)$$
ist.

5. ist in der physikalischen Literatur auch als "spontane Magnetisierung" bekannt: Ein Material, das in ein Magnetfeld gehalten wird, merkt sich dies bei genügend tiefen Temperaturen und wird selbst magnetisch. Bei hohen Temperaturen bleibt dieses Phänomen aus.

Auf der Ebene der Gesetze der großen Zahlen ist hiermit eigentlich alles gesagt. Als Wahrscheinlichkeitstheoretiker kann man sich fragen, ob m_N auch einem Zentralen Grenzwertsatz genügt und ob sich das kritische Verhalten bei $\beta = 1$ auch hier widerspiegelt. Wir werden der Einfachheit halber hierfür nur den Fall $h = 0$ betrachten. Der erste Trick besteht nun darin, dass man gar nicht betrachtet, was einen eigentlich interessiert, nämlich $\mu_{N,\beta,0} \circ m_N^{-1}$, sondern eine "geglättete" Version hiervon.

Lemma 3.12 *Betrachte das Maß*

$$\chi_{N,\beta,a} = Q_N * \mathcal{N}(0, \frac{a^2}{\beta N}),$$

also die Konvolution des Maßes

$$Q_N := \mu_{N,\beta,0}(am_N)^{-1}$$

mit einer Normalverteilung mit Erwartungswert 0 und Varianz $\frac{a}{\beta N}$. Hierbei ist $a = a_N$ möglicherweise abhängig von N. Dann hat Q_N eine Dichte bezüglich $\lambda\!\!\lambda^1$ der Form

$$f_{N,\beta,a}(x) = \frac{e^{-N\beta\Phi_\beta(\frac{x}{a})}}{\int_{\mathbb{R}} e^{-N\beta\Phi_\beta(\frac{x}{a})}}.$$

Hierbei ist

$$\Phi_\beta(x) = \frac{x^2}{2} - \frac{1}{\beta}\log\cosh(\beta x).$$

Beweis: Sei $A \in \mathcal{B}^1$. Dann ist

$$\begin{aligned}
\chi_{N,\beta,a}(A) &= \frac{1}{2^N}\sum_{\sigma\in\{\pm1\}^N} \mathcal{N}(0,\frac{a^2}{\beta N})(A - m_N(\sigma))\mu_{N,\beta,0}(\sigma) \\
&= \frac{1}{2^N}\sum_{\sigma}\sqrt{\frac{\beta N}{2\pi a^2}}\int_A e^{-\frac{\beta N}{2a^2}(x-a^2m_N(\sigma))^2}dx\mu_{N,\beta,0}(\sigma) \\
&= \frac{1}{Z_{N,\beta,0}}\sqrt{\frac{\beta N}{2\pi a^2}}\int_A \frac{1}{2^N}\sum_{\sigma} e^{-\frac{\beta N}{2a^2}(x-am_N(\sigma))^2} \times e^{\frac{\beta N}{2}m_N(\sigma)^2}dx \\
&= \frac{1}{Z_{N,\beta,0}}\sqrt{\frac{\beta N}{2\pi a^2}}\int_A e^{-\frac{\beta N}{2a^2}x^2} + \frac{1}{2^N}\sum_{\sigma} e^{\frac{\beta}{a}x\sum_{i=1}^N\sigma_i}dx \\
&= \frac{1}{Z_{N,\beta,0}}\sqrt{\frac{\beta N}{2\pi a^2}}\int_A e^{-\frac{\beta N}{2a^2}x^2+N\log\cosh(\frac{\beta x}{a})}dx \\
&= \frac{1}{Z_{N,\beta,0}}\sqrt{\frac{\beta N}{2\pi a^2}}\int_A e^{-N\beta\Phi_\beta(\frac{x}{a})}dx.
\end{aligned}$$

Da $\chi_{N,\beta,a}$ als Faltung zweier Wahrscheinlichkeitsmaße wieder ein Wahrscheinlichkeitsmaß ist, folgt

$$Z_{N,\beta,0}\sqrt{\frac{\beta_N}{2\pi a^2}} = \int_{\mathbb{R}} e^{-N\beta\phi_\beta(\frac{x}{a})}dx.$$

□

Wir werden nun zunächst einen Zentralen Grenzwertsatz für $\chi_{N,\beta,a}$ beweisen, wobei wir a geeignet wählen. Tatsächlich fällt die Wahl von a nicht schwer: Die Vermutung, dass der Limes von einem geeignet skalierten m_N Gaußsch verteilt ist und man m_N mit $\sqrt{N}$ skalieren muss, liegt nahe. Betrachten wir also $\chi_{N,\beta,\sqrt{N}}$.

Satz 3.13 *Für $\beta < 1$ konvergiert $\chi_{N,\beta,\sqrt{N}}$ gegen eine $\mathcal{N}(0, \frac{1}{\beta(1-\beta)})$-Verteilung.*

Beweis: Die Dichte hat schon die gewünschte Exponentialstruktur, allerdings muss noch gezeigt werden, dass nur quadratische Terme im Exponenten überleben. Betrachten wir die Taylor-Entwicklung des $\log\cosh(\cdot)$, so erhalten wir

$$\log\cosh(x) = \frac{x^2}{2} + O(x^4).$$

Somit hat der Exponent von $f_{N,\beta,\sqrt{N}}(x)$ die Entwicklung:

$$-N\beta\,\Phi_\beta(\frac{x}{\sqrt{N}}) = -\frac{N\beta}{2}(\frac{x}{\sqrt{N}})^2 - N\beta^2\frac{x^2}{N} + N\beta O(\frac{x^4}{N^2}) = -\frac{x^2}{2}(\beta(1-\beta)) + O(\frac{x^4}{N^3}).$$

Mit Hilfe des Satzes von der majorisierten Konvergenz folgt somit für $A \in \mathcal{B}^1$

$$\lim_{N\to\infty} \chi_{N,\beta,\sqrt{N}}(A) = \frac{\int_A e^{-\frac{x^2}{2}\beta(1-\beta)}dx}{\int_{\mathbb{R}} e^{-\frac{x^2}{2}\beta(1-\beta)}dx}.$$

Da $\lim_{N\to\infty} \chi_{N,\beta,\sqrt{N}}$ wieder ein Wahrscheinlichkeitsmaß ist, muss

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}\beta(1-\beta)}dx = \sqrt{\frac{\beta(1-\beta)}{2\pi}}$$

gelten.

$$\sqrt{\frac{\beta(1-\beta)}{2\pi}}e^{-\frac{x^2}{2}\beta(1-\beta)}$$

ist aber die Dichte der $\mathcal{N}(0, \frac{1}{\beta(1-\beta)})$-Verteilung. □

Aus diesem Satz lässt sich nun auch die Grenzverteilung von $\sqrt{N}m_N$ unter $\mu_{N,\beta,0}$ bestimmen.

Satz 3.14 *Für alle $\beta < 1$ gilt*

$$\mu_{N,\beta,0} \circ (\sqrt{N}m_N)^1 \Rightarrow \mathcal{N}(0, \frac{1}{1-\beta}).$$

Beweis: Das Maß $\chi_{N,\beta,\sqrt{N}}$ muss nur noch "entfaltet" werden. Das geschieht so: Ist $Y \sim \mathcal{N}(0, \frac{1}{\beta})$-verteilt, so besagt der vorhergehende Satz gerade, dass $Y + \sqrt{N}m_N \Rightarrow \mathcal{N}(0, \frac{1}{\beta(1-\beta)})$. Ist für eine Zufallsvariable X, φ_X definiert als

$$\varphi_X(t) = \mathbb{E}e^{itX},$$

so folgt

$$\varphi_{X_n+Y} \to \varphi_{X+Y} = \varphi_X \cdot \varphi_Y$$

für eine geeignete Zufallsgröße X. Da aber X_n und Y unabhängig sind, und $\varphi_Y(t) \neq 0$ für alle t, erhalten wir

$$\varphi_{X_n} \to \varphi_X.$$

Der Rest ist eine einfache Varianzberechnung. □

Satz 3.14 kann für $\beta = 1$ offenbar nicht richtig bleiben, denn eine Normalverteilung mit unendlicher Varianz ist unsinnig. Da $\beta = 1$ auch die kritische Temperatur ist, bei der die Menge der Limespunkte im Gesetz der großen Zahlen von einer einelementigen in eine zweielementige Menge übergeht, sieht man, dass dieses kritische Verhalten auch auf der Ebene der Fluktuation widergespiegelt wird. Tatsächlich leben die Fluktuationen auf einer ganz anderen (kleineren) Skala. Das lässt sich am besten an einer Analyse des Maßes $\chi_{N,1,\cdot}$ erkennen. Da sich bei $\beta = 1$ die 2. Ordnungsterme gegenseitig auslöschen, wird der 4. Ordnungsterm in der Entwicklung des $\log\cosh(\cdot)$ den Hauptbeitrag liefern. Damit dieser nicht einfach verschwindet, muss statt $a_N = \sqrt{N}$, $a_N = \sqrt[4]{N}$ gewählt werden.

Satz 3.15 *$\chi_{N,1,\sqrt[4]{N}}$ konvergiert schwach gegen ein Maß $\bar{\chi}$. $\bar{\chi}$ hat eine Lebesguedichte der Form*

$$\bar{f}(x) = \frac{e^{-\frac{1}{12}x^4}}{\int_{\mathbb{R}} e^{-\frac{1}{12}x^4} dx}.$$

Beweis: Wieder entwickeln wir die $\log\cosh(\cdot)$:

$$\log\cosh(x) = \frac{x^2}{2} + \frac{x^4}{12} + O(x^6).$$

Damit ergibt sich für $-N\Phi_1(\frac{x}{\sqrt[4]{N}})$ die Entwicklung

$$-N\Phi_1(\frac{x}{\sqrt[4]{N}}) = -\frac{x^4}{12} + O(N\frac{x^6}{N^{3/2}}).$$

Die Behauptung folgt nun wieder mit Hilfe des Satzes von der majorisierten Konvergenz. □

Im Falle von $\beta = 1$ ist das Limesmaß $\bar{\chi}$ der $\chi_{N,1,\sqrt[4]{N}}$ auch schon das Limesmaß der Magnetisierung.

Satz 3.16 *Bei $\beta = 1$ konvergiert*

$$\mu_{N,1,0} \circ (\sqrt[4]{N} m_N)^{-1}$$

schwach gegen die Verteilung $\bar{\chi}$ aus dem letzten Satz.

Beweis: $\chi_{N,1,\sqrt[4]{N}}$ ist die Verteilung der Summe von $\sqrt[4]{N} m_N$ (unter dem Gibbs-Maß bei Temperatur 1) und einer $\mathcal{N}(0, \frac{1}{\sqrt{N}})$-verteilten Zufallsvariablen, die davon unabhängig ist. Die letzte geht im Limes $N \to \infty$ in Wahrscheinlichkeit gegen 0. Da 0 eine Konstante ist folgt daraus die Verteilungskonvergenz von $\mu_{N,1,0} \circ (\sqrt[4]{N} m_0)^{-1}$ gegen $\bar{\chi}$. $\square$

Wir haben somit ein Bild davon gewonnen, wie sich der Phasenübergang im Curie-Weiss-Modell auf der Ebene der zentralen Grenzwertsätze widerspiegelt. Tatsächlich lassen sich CLTs für $\sqrt{N} m_N$ auch für $h \neq 0$ und für $\beta > 1$ beweisen. Hierbei ist es notwendig, die Maße richtig zu zentrieren und im Falle $h = 0,\ \beta > 1$, darauf zu bedingen, dass man sich "in der Nähe" von $m^*(\beta)$ bzw. $-m^*(\beta)$ befindet. Wir werden dies nicht ausführlicher betrachten, sondern zum Studium des Ising-Modells übergehen.

4 Gibbs-Maße für Gittersysteme

4.1 Grundsätzliches

Die Untersuchung des Ising-Modells in höheren Dimensionen wird nicht schwieriger, wenn wir sie in den allgemeinen Rahmen der Spinsysteme auf dem Gitter $\mathbb{Z}^d$ einbetten.

Wie schon erwähnt ist das Ising-Modell das historisch erste Spinsystem: Es sollte das Phänomen des Ferromagnetismus (grob) erklären. Zu diesem Zeitpunkt war klar, dass Ferromagnetismus durch eine gleichsinnige Ausrichtung von magnetischen Momenten, sogenannten Spins, hervorgebracht werden sollte. Dieses wird durch ein externes magnetisches Feld hervorgerufen, überdauert jedoch, wenn das Magnetfeld ausgeschaltet wird. Dieses Phänomen verschwindet, wenn man das Material erhitzt.

Es war auch klar, dass dieses Verhalten durch eine Wechselwirkung zwischen den Spins hervorgerufen werden sollte, die kurzreichweitig und anziehend ist. Die Fragen, die sich dabei jedoch stellten, sind:

- Wieso kann eine kurzreichweitige Wechselwirkung langreichweitige Phänomene hervorrufen?
- Wiese sollte ferromagnetisches Verhalten von der Temperatur abhängen?

Um dies zu erklären wurde das Ising-Modell eingeführt. Wir erinnern uns, dass dieses durch eine Hamiltonfunktion $\bar{H}_\Lambda$, $\Lambda \subseteq \mathbb{Z}^d$ endlich, $H_\Lambda : \{-1,+1\}^{\mathbb{Z}^d} \to \mathbb{R}$

$$\bar{H}_\Lambda(\sigma) = -\sum_{\substack{i,j\in\Lambda\\ <i,j>}} \sigma_i\sigma_j - h\sum_{i\in\Lambda}\sigma_i$$

beschrieben wurde. Λ haben wir uns bislang als großes aber endliches Teilgebiet von $\mathbb{Z}^d$ vorgestellt. Da allerdings schon im Falle des Curie-Weiss-Modells die interessanten Beobachtungen im thermodynamischen Limes $N \to \infty$ bzw. $|\Lambda| \to \infty$ gemacht werden konnten, liegt es nahe zu versuchen, das Modell gleich auf $\mathbb{Z}^d$ zu definieren, also beispielsweise

$$H(\sigma) = -\sum_{\substack{i,j\in\mathbb{Z}^d\\ <i,j>}} \sigma_i\sigma_j - h\sum_{i\in\mathbb{Z}^d}\sigma_i$$

und

$$\mu_\beta(\sigma) = \frac{1}{Z_\beta} e^{-\beta H(\sigma)}\rho(\sigma),$$

wobei ρ ein a priori-Maß auf $S^{\mathbb{Z}^d}$ ist, das Produktgestalt hat (da es das “freie”, nichtinteragierende System beschreibt), und Z_β wieder ein Normierungsfaktor ist

$$Z_\beta = \int e^{-\beta H(\sigma)}\rho(d\sigma).$$

Das Problem hierbei ist, dass die beiden obigen Gleichungen so keinen Sinn ergeben, denn für fast alle σ konvergieren die Reihen in der Definition von H nicht. Dieses Problem tritt bei $\bar{H}_\Lambda(\sigma)$ und dem zugehörigen Gibbs-Maß

$$\mu_{\Lambda,\beta} = \frac{e^{-\beta \bar{H}_\Lambda(\sigma)}}{\bar{Z}_{\Lambda,\beta}} \rho_\Lambda(d\sigma)$$

nicht auf. Es liegt daher nahe μ_β dadurch zu spezifizieren, dass nun $\mu_{\Lambda,\beta}$ als seine endlich dimensionalen Marginalien beschreibt (ähnlich wie man das für die Existenz des unendlichen Produktmaßes ρ tut) und den Satz von Kolmogorov zu verwenden. Das Problem bei diesem Ansatz ist, dass der Satz von Kolmogorov gewisse Verträglichkeitseigenschaften stellt, nämlich

$$\mu_{\Lambda,\beta}(A_\Lambda) = \mu_{\Lambda',\beta}(\pi^{-1}_{\Lambda',\Lambda}(A_\Lambda)) \tag{4.1}$$

für alle $A_\Lambda \in \mathcal{F}_\Lambda$ für $\Lambda \subseteq \Lambda'$. Hierbei ist

$$\mathcal{F}_\Lambda = \sigma(\{\sigma_V \in A\} : A \in \mathcal{B}(s^\Lambda), V \leq \Lambda,$$

$\sigma_V = (\sigma_i)_{i \in V}$, und $\pi_{\Lambda',\Lambda}$ ist die Projektion von $S^{\Lambda'}$ nach S^Λ. Diese sind vom Produktmaß, aber nicht von den $(\mu_{\Lambda,\beta})_\Lambda$ erfüllt. Dies liegt daran, dass bei der Berechnung von $\mu_{\Lambda',\beta}(A_\Lambda)$ auch die Spins außerhalb von Λ einen Einfluss ausüben – zumindest, wenn sie nicht weiter als Distanz 1 von Λ entfernt sind. Es liegt daher nahe, nicht nur die Elemente aus Λ, sondern auch die aus dem Rand von Λ in Betracht zu ziehen. [Dabei sollte betont werden, dass “es liegt nahe” natürlich eine Formulierung ist, die aus der Retrospektive leicht getroffen ist; dieser Ansatz galt in den Jahren 1968/69, als er von R. Dobrushin und wenig später von O. E. Lanford und D. Ruelle entwickelt wurde, durchaus als bahnbrechend.] Entsprechend definiert man eine Version von H als $H : S^{\mathbb{Z}^d} \to \mathbb{R}$:

$$H_\Lambda(\sigma) = - \sum_{\substack{i \in \Lambda \\ \vee\, j \in \Lambda \\ <i,j>}} \sigma_i \sigma_j - h \sum_{i \in \Lambda} \sigma_i.$$

Hierbei spielen also auch die Spins im Rand von Λ eine Rolle. Diese Hamiltonfunktion hat den Vorteil, mit der Inklusion verträglich zu sein: Für $\Lambda \subseteq \Lambda'$ gilt

$$(H_{\Lambda'})_\Lambda(\sigma) = H_\Lambda(\sigma).$$

Um dies besser zu verstehen, sollte man den Index Λ als eine Operation auf eine vorher defineirte Funktion (in dem Fall unser obiges H) betrachten. Hierfür ist natürlich notwendig, dass H die Struktur $H = \sum \Phi_A(x)$ hat. Entsprechend definiert man sich eine Version des lokalen Gibbsmaßes auf $(S^{\mathbb{Z}^d}, \mathcal{F})$

$$\mu^\eta_{\Lambda,\beta}(\sigma_\Lambda) = \frac{1}{Z^\eta_{\Lambda,\beta}} e^{-\beta H_\Lambda((\sigma_\Lambda, \eta_{\Lambda^c}))} \rho_\Lambda(\sigma) \tag{4.2}$$

mit $\sigma \in S^{\mathbb{Z}^d}$ und $\eta \in S^{\mathbb{Z}^d}$. Hierbei ist $\sigma_\Lambda \in \{\pm 1\}^\Lambda$, $\eta_{\Lambda^c} \in \{\pm 1\}^{\mathbb{Z}^d \setminus \Lambda}$ und $(\sigma_\Lambda, \eta_{\Lambda^c})$ die Konfiguration, die auf Λ mit σ_Λ und auf $\mathbb{Z}^d \setminus \Lambda$ mit η_{Λ^c} übereinstimmt. ρ_Λ ist ein Produktmaß auf $\{\pm 1\}^\Lambda$. Das Nette daran ist, dass die μ^η_Λ automatisch die Verträglichkeitsbedingungen (4.1) für bedingte Wahrscheinlichkeiten erfüllen. Die Idee ist, dass diese $\mu^\eta_{\Lambda,\beta}$ die bedingten Verteilungen des Gibbsmaßes im endlichen Volumen beschreiben sollen:

Ein Wahrscheinlichkeitsmaß μ_β auf $(\{\pm 1\}^{\mathbb{Z}^d}, \mathcal{F})$, wobei $\mathcal{F}$ die σ-Algebra ist, die die endlichen Projektionen messbar macht, heißt Gibbs-Maß für den Hamiltonian H und zur inversen Temperatur β, dann und nur dann, wenn seine bedingten Verteilungen (gegeben $\eta_{\Lambda^c} \in \{\pm 1\}^{\Lambda^c}$ auf Λ^c, $\Lambda \subseteq \mathbb{Z}^d$ endlich) durch (4.2) gegeben sind. Das machen wir im Laufe des Kapitels präziser.

Es ergeben sich sofort zwei Fragen:

1. Gibt es solche Maße?
2. Wenn es solche Maße gibt, sind sie eindeutig?

Wir werden sehen, dass Frage 1 für eine große Klasse von Wechselwirkungen, u. a. für das Ising-Modell, positiv beantwortet werden kann. Die Antwort auf Frage 2 kann und wird in den interessantesten Fällen von β abhängen. Für gewisse Bereiche von β (in unseren Fälles großes β) wird es mehrere Gibbs-Maße geben, die beispielsweise eine positive und eine negative Phase des Ferromagneten beschreiben. Für kleine β hingegen ist das Gibbs-Maß eindeutig. Im Falle mehrerer Gibbs-Maße spricht man auch von Phasenübergang (dies ist allerdings nur *eine* Art, den Phasenübergang zu definieren).

Bevor wir uns im Detail dem Ising-Modell und Phasenübergang dort zuwenden, wollen wir zunächst einen allgemeineren Rahmen für Gibbs-Maße anbieten.

4.2 Lokale Spezifikation und Gibbs-Maße

Hier soll nun die oben skizzierte Methode zur Konstruktion globaler Gibbs-Maße für eine große Klasse von Potentialen durchgeführt werden. Sei hierfür S ein vollständiger, separabler metrischer Raum, z. B. $S = \{-1, +1\}$.

Definition 4.1 *Eine Interaktion ist eine Familie von Funktionen $(\Phi_A)_{A \subset \mathbb{Z}^d} =: \Phi$, wobei $\Phi_A : S^{\mathbb{Z}^d} \to \mathbb{R}$ jeweils $\mathcal{F}_A$-messbar sei. Φ heißt stetig, falls alle Φ_A stetig sind. Φ heißt* REGULÄR, *falls für alle $x \in \mathbb{R}^d$ ein c existiert, so dass*

$$\sum_{A \ni x} \|\Phi_A\|_\infty \leq c < \infty$$

gilt.

Aus Interaktionen lässt sich vermöge

$$H_\Lambda(\sigma) = -\sum_{A \cap \Lambda \neq \emptyset} \Phi_A(\sigma)$$

für endliche $\Lambda \subseteq \mathbb{Z}^d$ eine Hamiltonfunktion konstruieren. Wählt man z. B. $S = \{-1, +1\}$ und

$$\Phi_A(\sigma) = \begin{cases} \sigma_i \sigma_j & |A| = 2,\ A = \{i, j\} \quad \text{und} \\ 0 & \text{sonst} \end{cases},$$

so erhält man den Hamiltonian des Ising-Modells aus dem vorhergehenden Abschnitt für $h = 0$.

Mit Hilfe dieser Hamiltonfunktionen lässt sich nun ein globales Gibbs-Maß konstruieren. Hierzu benötigen wir:

Definition 4.2 *Eine lokale Spezifikation ist eine Familie von Wahrscheinlichkeitskernen* $\{\mu_{\Lambda,\beta}^{(\cdot)}\}_{\Lambda \subseteq \mathbb{Z}^d}$*, so dass*

1. *Für alle* Λ *und alle* $A \in \mathcal{F}_\Lambda$ *ist* $\mu_{\Lambda,\beta}^{(\cdot)}(A)$ *eine* $\mathcal{F}_{\Lambda^c}$*-messbare Funktion.*
2. *Für jedes* $\eta \in S^{\mathbb{Z}^d}$ *ist* $\mu_{\Lambda,\beta}^{\eta}$ *ein Wahrscheinlichkeitsmaß auf* $(S^{\mathbb{Z}^d}, \mathcal{F})$*.*
3. *Für alle* Λ, Λ' *mit* $\Lambda \subseteq \Lambda'$ *und jedes messbare* f *gilt*
$$\begin{aligned} &\int \int f(\sigma_\Lambda, \sigma'_{\Lambda' \setminus \Lambda}, \eta_{\Lambda'^c}) \mu_{\Lambda,\beta}^{(\eta_{\Lambda'^c}, \sigma'_{\Lambda'})}(d\sigma) \mu_{\Lambda',\beta}^{\eta}(d\sigma') \\ = &\int f(\sigma'_{\Lambda'}, \eta_{\Lambda'^c}) \mu_{\Lambda',\beta}^{\eta}(d\sigma'), \end{aligned} \tag{4.3}$$
wobei die Notationen $\sigma_\Lambda, \sigma'_{\Lambda'}, \eta_{\Lambda'^c}$ *etc. selbsterklärend sind.*

1. und 2. oben sind nichts weiter als die Definition eines Wahrscheinlichkeitskerns (sie sind notwendig, um 3. Sinn zu geben), während 3. die besagte Kompatibilitätsbedingung und somit das Herzstück der Definition ist.

Mit Hilfe von Hamiltonfunktionen lassen sich nun in der Tat lokale Spezifikationen konstruieren:

Lemma 4.3 *Ist* Φ *eine reguläre Interaktion, dann ist vermöge* $\mu_{\Lambda,\beta}^{\eta}(d\sigma)$ *mit*
$$\int f(\sigma) \mu_{\Lambda,\beta}^{\eta}(d\sigma) := \int \frac{e^{-\beta H_\Lambda(\sigma_\Lambda, \eta_{\Lambda^c})}}{Z_{\Lambda,\beta}^{\eta}} f(\sigma_\Lambda, \eta_{\Lambda^c}) \rho_\Lambda(d\sigma_\Lambda)$$
(für alle integrierbaren f*) eine lokale Spezifikation definiert. Diese heißt Gibbs-Spezifikation zur Interaktion* Φ *und zu inverser Temperatur* β*. Hierbei ist wieder* ρ_Λ *ein Produktmaß, also* $\rho_\Lambda = \prod_{i \in \Lambda} \rho$*.*

Beweis: Übung. □

Wir werden die wichtige Beziehung (4.2) auch kurz mit
$$\mu_{\Lambda',\beta}^{(\cdot)} \mu_{\Lambda,\beta}^{(\cdot)} = \mu_{\Lambda',\beta}^{(\cdot)}$$
abkürzen.

In der Folge werden wir den Begriff der bedingten Erwartung des öfteren benutzen. Dieser ist hinlänglich aus der Wahrscheinlichkeitstheorie II bekannt. Zur bedingten Erwartung gehört der Begriff der regulären bedingten Verteilung.

Definition 4.4 *Seien $\mathcal{G} \subseteq \mathcal{F}$ zwei σ-Algebren über einer Menge Ω. Eine Funktion $\mu^{\eta}_{\mathcal{G}}$ heißt* REGULÄRE BEDINGTE VERTEILUNG, *falls*

1. *Für alle $\eta \in \Omega$ ist $\mu^{\eta}_{\mathcal{G}}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{F}$.*
2. *Für alle $A \in \mathcal{F}$ ist $\mu^{\eta}_{\mathcal{G}}(A)$ eine $\mathcal{G}$-messbare Funktion, dergestalt, dass für fast alle η*

$$\mu^{\eta}_{\mathcal{G}}(A) = \mu(\mathbb{1}_A|\mathcal{G})(\eta)$$

gilt. Hierbei ist $\mu(\cdot|\mathcal{G})$ die bedingte Erwartung gegeben $\mathcal{G}$.

Reguläre bedingte Verteilungen existieren, wann immer der zugrunde liegende Wahrscheinlichkeitsraum Polnisch, d. h. ein vollständiger separabler metrischer Raum ist; insbesondere ist dies in all unseren Situationen der Fall.

Wie A. Bovier schreibt sind nun lokale Spezifikationen "bedingte Verteilungen auf der Suche nach einem Maß". Dieses Maß ist das Gibbs-Maß in unendlichen Volumen.

Wir fassen nun den vorher skizzierten Gedanken zu einer Definition zusammen.

Definition 4.5 *Sei $\{\mu^{(\cdot)}_{\Lambda,\beta}\}$ eine lokale Spezifikation. Ein Maß μ_β heißt verträglich mit dieser Spezifikation dann und nur dann, wenn für alle $\Lambda \subseteq \mathbb{Z}^d$ und alle beschränkten $\mathcal{F}$-messbaren Funktionen f gilt*

$$\mu_\beta(f|\mathcal{F}_{\Lambda^c}) = \mu^{(\cdot)}_{\Lambda,\beta}(f) \quad \mu_\beta\text{-}f.s.$$

Wird die lokale Spezifikation von einer Interaktion Φ erzeugt, so heißt ein mit dieser lokalen Spezifikation verträgliches Maß μ_β Gibbs-Maß zu (Φ, ρ) bei inverser Temperatur β.

Satz 4.6 *Ein Wahrscheinlichkeitsmaß μ_β ist ein Gibbs-Maß für Φ, ρ, β genau dann, wenn für alle $\Lambda \subseteq \mathbb{Z}^d$ gilt*

$$\mu_\beta \mu^{(\cdot)}_{\Lambda,\beta} = \mu_\beta. \tag{4.4}$$

Beweis: Falls

$$\mu^{(\cdot)}_{\Lambda,\beta}(f) = \mu_\beta(f|\mathcal{F}_{\Lambda^c}) \quad \mu_\beta\text{-f.s.}$$

ist, so folgt die Behauptung definitionsgemäß. Also muss nur die umgekehrte Richtung bewiesen werden. Gelte also (4.3) und wir müssen zeigen, dass

$$\mu_\beta(f|\mathcal{F}_{\Lambda^c}) = \mu^{(\cdot)}_{\Lambda,\beta}(f) \quad \mu_\beta\text{-f.s.}$$

gilt. Nun sind lokale Spezifikationen $\mu^{(\cdot)}_{\Lambda,\beta}$ definitionsgemäß $\mathcal{F}_{\Lambda^c}$-messbar. Wendet man nun (4.4) auf

$$f'(\eta) = f(\eta)h(\eta_{\Lambda^c})$$

mit einer $\mathcal{F}_{\Lambda^c}$-messbaren Funktion h an, so ergibt sich

$$\mu_\beta(h\mu^{(\cdot)}_{\Lambda,\beta}(f)) = \mu_\beta(f \cdot h).$$

Dies ist eine Art, bedingte Erwartung zu definieren. □

Die Gleichungen (4.4) sind als DLR-Gleichung (nach Dobrushin, Lanford und Ruelle) bekannt geworden. Was nun noch zu unserem Glück fehlt, ist ein Resultat, das besagt, dass es in typischen Situationen auch Gibbs-Maße gibt.

Satz 4.7 *Sei Φ eine stetige, reguläre Interaktion mit zugehöriger Spezifikation $(\mu_{\Lambda,\beta}^{(\cdot)})$. Sei (Λ_n) die Folge endlicher Volumen, die gegen $\mathbb{Z}^d$ aufsteigt. Falls für ein η die Folge $(\mu_{\Lambda_n,\beta}^{\eta})$ schwach gegen ein Wahrscheinlichkeitsmaß ν konvergiert, dann ist dieses ν ein Gibbs-Maß für Φ, ρ, β.*

Beweis: Sei f eine stetige beschränkte Funktion. Dann gilt

$$\mu_{\Lambda_n,\beta}^{\eta}(f) \to \nu(f) \quad \text{mit} \quad n \to \infty. \tag{4.5}$$

Andererseits gilt für alle $\Lambda_n \supseteq \Lambda$

$$\mu_{\Lambda_n,\beta}^{\eta}\mu_{\Lambda,\beta}^{(\cdot)}(f) = \mu_{\Lambda_n,\beta}^{\eta}(f). \tag{4.6}$$

Wir wollen nun gern behaupten, dass $\mu_{\Lambda_n,\beta}^{\eta}\mu_{\Lambda,\beta}^{(\cdot)}(f)$ gegen $\nu\mu_{\Lambda,\beta}^{(\cdot)}(f)$ konvergiert. Dazu aber muss $\mu_{\Lambda,\beta}^{(\cdot)}(f)$, d. h. $\eta \mapsto \mu_{\Lambda,\beta}^{\eta}(f)$, eine stetige Funktion sein. Dann können wir (4.5) anwenden, um das Gewünschte zu erhalten. Dies ist die sogenannte FELLER-EIGENSCHAFT:

Lemma 4.8 *Die lokalen Spezifikationen einer stetigen und regulären Interaktion haben die Feller-Eigenschaft, d. h. für jedes stetige f ist auch $\mu_{\Lambda,\beta}^{(\cdot)}(f)$ stetig.*

Beweis: Zu zeigen ist, dass für $\eta_n \to \eta$ auch gilt

$$\mu_{\Lambda,\beta}^{\eta_n}(f) \to \mu_{\Lambda,\beta}^{\eta}(f).$$

Da f stetig ist, folgt dies, falls

$$H_\Lambda(\sigma_\Lambda, \eta_{n,\Lambda^c}) \to H_\Lambda(\sigma_\Lambda, \eta_{\Lambda^c})$$

gilt. H_Λ ist aber nach Voraussetzung eine gleichmäßig konvergente Summe stetiger Funktionen, also auch stetig. □

Nun können wir den Beweis des Satzes beenden. Wir wenden auf (4.6) den Limes $n \to \infty$ an: Nach der Feller-Eigenschaft erhalten wir links

$$\nu\mu_{\Lambda,\beta}^{(\cdot)}(f) \quad \text{und rechts} \quad \nu(f),$$

also

$$\nu\mu_{\Lambda,\beta}^{(\cdot)}(f) = \nu(f).$$

Doch dies sagt gerade, dass ν ein Gibbs-Maß ist. □

Korollar 4.9 *Ist S kompakt und Φ stetig und regulär, so gibt es für jedes $0 \leq \beta < \infty$ mindestens ein Gibbs-Maß.*

Beweis: Nach dem Satz von Tychonov ist auch $S^{\mathbb{Z}^d}$ kompakt. Dann ist auch die Menge der Wahrscheinlichkeitsmaße auf $S^{\mathbb{Z}^d}$ schwach kompakt. Also hat jede Folge $\mu^{\eta}_{\Lambda_n,\beta}$ konvergente Teilfolgen. Die Häufungspunkte dieser Folgen sind Gibbs-Maße. $\square$

Bemerkung: Es gibt Situationen, in denen S nicht kompakt ist und auch keine Gibbs-Maße existieren.

5 Dobrushins Kriterium und Peierls Argument – Phasenübergänge

Wir werden uns nun der Frage zuwenden, wieviele Limes-Gibbs-Maße mit einer gegebenen Spezifikation verträglich sind, d. h. für welche Parameterbereiche es einen Phasenübergang gibt (und ob das überhaupt je der Fall ist). Intuitiv ist klar: Das Maß ρ selbst ist als Produktmaß eindeutig. Die Möglichkeit, mehr als eine Limesverteilung zu haben, kann also nur auf den Einfluss von H zurückzuführen sein, genauer darauf, dass H mehreren Spinkonfigurationen, die fundamental verschiedenartig sind, die gleiche minimale Energie zuweist. Dieser Einfluss von H steigt mit der Größe von β. Für kleine β, also hohe Temperaturen, sollte es möglich sein, μ_β als Permutationen des Produktmaßes H aufzufassen; dann sollte es eindeutig sein.

Um dies genauer zu fassen, definieren wir für zwei Maße μ, ν auf einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A})$ den Abstand der totalen Variation als

$$d_{TV}(\mu, \nu) := \|\mu - \nu\| := \sup_{A \in \mathcal{A}} \|\mu(A) - \nu(A)\|.$$

Dann gilt:

Satz 5.1 *Sei $\mu_{\Lambda,\beta}^{(\cdot)}$ eine lokale Spezifikation, die der Feller-Bedingung genügt. Für $i, j \in \mathbb{Z}^d$ setze*

$$\rho_{ij} := \frac{1}{4} \sup_{\substack{\eta,\eta' \\ \forall\, k \neq i : \eta_i = \eta_i'}} \|\mu_{j,\beta}^{\eta} - \mu_{j,\beta}^{\eta'}\|.$$

Ist

$$\sup_{j \in \mathbb{Z}^d} \sum_{i \in \mathbb{Z}^d} \rho_{ij} < 1,$$

so ist die lokale Spezifikation mit höchstens einem Gibbs-Maß verträglich.

Beweis: Sei $f : S^{\mathbb{Z}^d} \to \mathbb{R}$ eine stetige Funktion. Wir definieren ihre Variation in $j \in \mathbb{Z}^d$ durch

$$\delta_j(f) = \sup_{\substack{\eta,\eta' \\ \forall\, k \neq j : \eta_k = \eta_k'}} |f(\eta) - f(\eta')|$$

und ihre totale Variation als

$$\Delta(f) = \sum_{i \in \mathbb{Z}^d} \delta_i(f).$$

sei

$$\mathcal{T} = \{f \in C(S^{\mathbb{Z}^d}) : \Delta(f) < +\infty\}.$$

□

Übung: Man überprüfe, dass

$$\overline{\mathcal{T}} = C(S^{\mathbb{Z}^d}).$$

Wir gehen nun in 2 Schritten vor:

1. Zeige, dass Δ eine Halbnorm auf $\mathcal{T}$ ist und
$$\Delta(f) = 0 \Rightarrow f \equiv \text{const.}$$
2. Konstruiere eine Kontraktion T, so dass Lösungen der DLR-Gleichung darunter invariant sind.

Dann gilt für jede Lösung μ der DLR-Gleichungen
$$\mu(f) = \mu(T^n f) \to c(f).$$

Da aber der Wert auf stetige Funktionen ein Maß eindeutig festlegt, sind alle Lösungen der DLR-Gleichungen identisch.

Wir beginnen mit dem zweiten Punkt: Sei $x_1, x_2, \ldots$ eine Aufzählung der $\mathbb{Z}^d$. Setze
$$Tf := \lim_{n\to\infty} \mu^{(\cdot)}_{x_1,\beta}\mu^{(\cdot)}_{x_2,\beta} \ldots \mu^{(\cdot)}_{x_n,\beta}(f).$$

Übung: Man zeige, dass dieser Limes in der Sup-Norm existiert.

Dies impliziert, dass T stetige Funktionen auf stetige Funktionen abbildet.

Wenn μ_β die DLR-Gleichungen erfüllt, so gilt für alle lokalen Spezifikationen
$$\mu_\beta \mu^{(\cdot)}_{\Lambda,\beta} = \mu_\beta,$$
also auch
$$\mu(Tf) = \mu(f).$$
Wir zeigen noch, dass T tatsächlich unter Δ eine Kontraktion ist, falls
$$\sup_{j\in\mathbb{Z}^d} \sum_{i\in\mathbb{Z}^d} \rho_{i,j} \le \alpha < 1.$$

Dazu zeigen wir

Lemma 5.2 *Sei $f \in \mathcal{T}$. Dann gilt*

(i) $\delta_i(\mu_{i,\beta}(f)) = 0$ *für* $i \in \mathbb{Z}^d$.

(ii) Für $j \neq i$ *gilt*
$$\delta_i(\mu_j(f)) \le \delta_i(f) + \rho_{i,j}\delta_j(f).$$

Beweis: Für (i) beachtet man, dass δ_i gerade die maximale Veränderung einer Funktion beschreibt, wenn man den Wert des Spins in (i) ändert. Über diesen wird in $\mu_{i,\beta}(f)$ aber schon integriert, also hängt $\mu_{i,\beta}(f)$ gar nicht mehr von η_i ab und $\delta_i(\mu_i(f)) = 0$.

Nun sei $i \neq j$: Es ist

$$
\begin{aligned}
\delta_i(\mu_{j,\beta}(f)) : &= \sup_{\substack{\eta,\eta' \\ \eta_k=\eta'_k\,\forall k\neq i}} |\mu^{\eta}_{j,\beta}(f) - \mu^{\eta'}_{j,\beta}(f)| \\
&= \sup_{\substack{\eta,\eta' \\ \eta_k=\eta'_k\forall k\neq i}} |\int f(\sigma_j,\eta_{j^c})\mu^{\eta}_j(d\sigma_j) - \int f(\sigma_j,\eta'_{j^c})\mu^{\eta}_j(d\sigma_j) \\
&\quad + \int f(\sigma_j,\eta'_{j^c})(\mu^{\eta}_j(d\sigma_j) - \mu^{\eta'}_j(d\sigma_j))| \\
&\leq \sup_{\substack{\eta,\eta' \\ \eta_k=\eta'_k\,\forall k\neq i}} \int |f(\sigma_j,\eta_{j^c}) - f(\sigma_j,\eta'_{j^c})|\mu^{\eta}_j(d\sigma_j) \\
&\quad + \sup_{\substack{\eta,\eta' \\ \forall k\neq i,\eta_k=\eta'_k}} |\int f(\sigma_j,\eta'_{j^c})(\mu^{\eta}_j(d\sigma_j) - \mu^{\eta'}_j(d\sigma_j))|.
\end{aligned}
$$

Nun gilt für den ersten Summanden

$$
\sup_{\substack{\eta,\eta' \\ \eta_k=\eta'_k\,\forall k\neq i}} \int |f(\sigma_j,\eta_{j^c}) - f(\sigma_j,\eta'_{j^c})|\mu^{\eta}_j(d\sigma_j) \leq \delta_i(f).
$$

Für den zweiten Term können wir einfach eine Konstante unter den Integral addieren, da diese bei der Integration gegen die Differenz zweier Wahrscheinlichkeitsmaße verschwindet. Also:

$$
\begin{aligned}
&= \sup_{\substack{\eta,\eta' \\ \forall k\neq i:\eta_k\eta'_k}} |\int f(\sigma_j,\eta'_{j^c})(\mu^{\eta}_j(d\sigma_j) - \mu^{\eta'}_j(d\sigma_j))| \\
&\leq \sup_{\substack{\eta,\eta' \\ \forall k\neq i:\eta_k\eta'_k}} \int |(f(\sigma_j,\eta'_{j^c}) - \int_{\tau_j} f(\tau_j,\eta_{j^c}))(\mu^{\eta}_j(d\sigma_j) - \mu^{\eta'}_j(d\sigma_j)| \\
&\leq \sup_{\substack{\eta,\eta' \\ \forall k\neq i:\eta_k\eta'_k}} |f(\eta) - f(\eta')| \sup_{\substack{\eta,\eta' \\ \forall k\neq i:\eta_k\eta'_k}} \sup_A |\mu^{\eta}_j(A) - \mu^{\eta'}_j(A)| \\
&= \|\mu^{\eta}_j - \mu^{\eta'}_j\|\delta_j(f).
\end{aligned}
$$

□

Lemma 5.3 *Unter der Annahme, dass*

$$
\sup_j \sum_i \rho_{i,j} \leq \alpha
$$

folgt für alle $n \in \mathbb{N}_0$:

$$
\Delta(\mu^{(\cdot)}_{x_i} \dots \mu^{(\cdot)}_{x_n}(f)) \leq \alpha \sum_{i=1}^{n} \delta_{x_i}(f) + \sum_{j\geq n+1} \delta_{x_j}(f).
$$

Beweis: Wir beweisen dies per Induktion. Für $n = 0$ ist nichts zu zeigen, denn dort gilt per Definition von Δ Gleichheit.

Gilt die Hypothese nun für $n \in \mathbb{N}$, so folgt

$$\begin{aligned}
\Delta(\mu_{x_1}^{(\cdot)} \dots \mu_{x_{n+1}}^{(\cdot)}(f) &\leq \alpha \sum_{i=1}^{n} \delta_{x_i}(\mu_{x_{n+1}}^{(\cdot)}(f)) + \sum_{j \geq n+1} \delta_{x_j}(\mu_{x_{n+1}}^{(\cdot)}(f)) \\
&\leq \alpha \sum_{i=1}^{n} [\delta_{x_i}(f) + \rho_{x_i,x_{n+1}} \delta_{x_{n+1}}(f)] + \sum_{j \geq n+2} [\delta_{x_j}(f) + \rho_{x_j,x_{n+1}} \delta_{x_{n+1}}(f)],
\end{aligned}$$

da $\delta_{x_{n+1}}(\mu_{x_{n+1}}^{(\cdot)}(f)) = 0$. Dies ist nun

$$\begin{aligned}
&= \alpha \sum_{i=1}^{n} \delta_{x_i}(f) + \sum_{i=1}^{\infty} \rho_{x_i,x_{n+1}} \delta_{x_{n+1}}(f) + \sum_{j \geq n+2} \delta_{x_j}(f) \\
&\leq \alpha \sum_{i=1}^{n+1} \delta_{x_i}(f) + \sum_{j \geq n+2} \delta_{x_j}(f).
\end{aligned}$$

Und das Lemma folgt. □

Nimmt man im letzten Lemma nun den Limes $n \to \infty$, so erhält man

$$\Delta(Tf) \leq \alpha \Delta(f).$$

Es bleibt für Punkt 1. zu zeigen, dass $\Delta(f) = 0$ impliziert, dass $f \equiv$const. ist. Tatsächlich gilt

$$\Delta(f) \geq \sup(f) - \inf(f).$$

In der Tat, da f stetig ist, gibt es für alle $\varepsilon > 0$ einen endlichen Würfel Λ und Konfigurationen ω^+ und ω^- mit $\omega^+_{\Lambda^c} = \omega^-_{\Lambda^c}$, so dass

$$\sup(f) \leq f(\omega^+) + \varepsilon \quad \text{und} \quad \inf(f) \geq f(\omega^-) - \varepsilon.$$

Andererseits ist

$$f(\omega^+) - f(\omega^-) \leq \sum_{x \in \Lambda} \delta_x(f) \leq \Delta(f).$$

Dies ergibt

$$\sup(f) - \inf(f) \leq \Delta(f) + 2\varepsilon.$$

Da ε beliebig war, folgt die Behauptung. Damit ist auch der Satz bewiesen. □

Wir wollen dieses Kriterium für das Ising-Modell in $\mathbb{Z}^d$ betrachten:

Beispiel 5.4 *Das Ising-Modell in $\mathbb{Z}^d$. Abzuschätzen ist also*

$$\sup_j \sum_{i \in \mathbb{Z}^d} \sup_{\substack{\eta, \eta' \\ \forall k \neq i: \eta_k = \eta'_k}} \|\mu_{j,\beta}^{\eta} - \mu_{j,\beta}^{\eta'}\|.$$

Zunächst kann man dabei das äußere Supremum weglassen, da das Potential

$$H(\sigma) = -\sum_{<i,j>} \sigma_i \sigma_j - h \sum_{i \in \mathbb{Z}^d} \sigma_i$$

translationsinvariant ist und somit die in Frage stehenden Ausdrücke für alle $j \in \mathbb{Z}^d$ identisch sind. Wir wollen uns auf den Fall $h = 0$ konzentrieren. Weiter sind die Maße $\mu^{\eta}_{j,\beta}$ und $\mu^{\eta'}_{j,\beta}$ zu vergleichen. Diese können nur für zwei Ereignisse verschieden sein, $\sigma_j = +1$ oder $\sigma_j = -1$. Aus Symmetriegründen muss die Differenz für beide Ereignisse gleich groß sein. Betrachten wir also das Ereignis $A = \{\sigma_i = +1\}$. Die Summe $\sum_{i \in \mathbb{Z}^d}$, die formal über alle Elemente des $\mathbb{Z}^d$ läuft, hat eigentich nur $2d$ viele Elemente. In der Tat sind $\mu^{\eta}_{j,\beta}$ und $\mu^{\eta'}_{j,\beta}$ identisch, wenn i und j nicht nächste Nachbarn sind (wobei i der einzige Index ist, in dem sich η und η' unterscheiden. Schließlich liefern alle i mit $\|i - j\| = 1$ die gleichen Werte für $\mu^{\eta}_{j,\beta}(A)$ und $\mu^{\eta'}_{j,\beta}(A)$. Somit gilt

$$\sup_j \sum_{i \in \mathbb{Z}^d} \sup_{\substack{\eta,\eta' \\ \forall k \neq i: \eta_k = \eta'_k}} \|\mu^{\eta}_{j,\beta} - \mu^{\eta'}_{j,\beta}\| = 2d \sup_{\substack{\eta,\eta' \\ \forall k \neq i: \eta_k = \eta'_k}} |\mu^{\eta}_{j,\beta}(A) - \mu^{\eta'}_{j,\beta}(A)|.$$

Hierbei ist i irgendein Nachbar von j.

Nun ist für ein (η, η'), das an der Supremumsbildung teilnimmt (und es gibt nur endlich viele solcher Paare)

$$|\mu^{\eta}_{j,\beta}(A) - \mu^{\eta'}_{j,\beta}(A)| = \left| \frac{e^{\beta \sum_{<k,j>} \eta_k}}{e^{\beta \sum_{<k,j>} \eta_k} + e^{-\beta \sum_{<k,j>} \eta_k}} - \frac{e^{+\beta \sum_{<k,j>} \eta'_k}}{e^{\beta \sum_{<k,j>} \eta'_k} + e^{-\beta \sum_{<k,j>} \eta'_k}} \right|,$$

wobei k ein Nachbar von j ist und $\eta_k = \eta'_k$ für alle $k \neq i$ gilt. Diese Differenz ist eine stetige Funktion in β, die in β den Wert 0 annimmt. Also gilt für β hinreichend klein

$$\sup_{\substack{\eta,\eta' \\ \forall k \neq i: \eta_k = \eta'_k}} |\mu^{\eta}_{j,\beta}(A) - \mu^{\eta'}_{j,\beta}(A)| < \frac{1}{2d}.$$

Dies zeigt zusammen mit dem vorhergehenden Satz, dass es für hinreichend große Temperaturen, d. h. kleine β, nur nöchstens ein Gibbs-Maß gibt. Dies interpretiert man als Abwesenheit von Phasenübergängen im d-dim. Ising-Modell bei hohen Temperaturen.

Bemerkungen:

1. Ähnlich lässt sich zeigen, dass im d-dim Ising-Modell mit einem genügend großen externen Magnetfeld $|h|$ keine Phasenübergänge stattfinden.

2. Für Gibbs'sche Spezifikationen bezüglich einer regulären Interaktion $(\Phi_A)_{A \subseteq \mathbb{Z}^d}$ lässt sich zeigen, dass keine Phasenübergänge stattfinden, falls

$$\sup_{x \in \mathbb{Z}^d} \sum_{A \ni x} (|A| - 1) \|\Phi_A\|_\infty < \frac{1}{\beta}$$

gilt. Dies findet man z. B. bei Georgii [3], Prop. 8.8. Für das Curie-Weiss-Modell, als ein Modell auf $\mathbb{Z}^d$ betrachtet, liefert dies lustigerweise genau die richtige kritische Temperatur.

3. Ohne das Dobrushinsche Kriterium zu verwenden lässt sich zeigen, dass für eine ziemlich allgemeine Klasse von Potentialen in *einer* Dimension höchstens ein Gibbs-Maß existiert (insbesondere für das Λ-dimensionale Ising-Modell). Dies findet man bei Georgii [3], Kapitel 8.3.

Da wir nun im Ising-Modell festgestellt haben, dass es für genügend hohe Temperaturen keinen Phasenübergang gibt, stellt sich auf natürliche Weise die Frage, ob man vielleicht für tiefere Temperaturen einen Phasenübergang nachweisen kann. Es sollte darauf hingewiesen werden, dass dies lange Zeit nicht für möglich gehalten wurde. Darüber hinaus sei betont, dass der Nachweis eines Phasenübergangs sehr viel schwieriger ist als der Nachweis des Gegenteils und zumeist sehr auf den Einzelfall zugeschnittene Methoden erfordert. Wir konzentrieren uns somit auf das d-dim. Ising-Modell und stellen ein Argument vor, das auf Peierls zurückgeht und das von Dobrushin und Griffith mathematisch rigoros gemacht wurde. Die Kernidee ist, dass es im Ising-Modell mit $h = 0$ zwei Zustände minimaler Energie gibt, nämlich

$$\sigma^+ : \sigma_i^+ = 1 \quad \forall i \quad \text{und} \quad \sigma^- : \sigma_i^- = -1 \quad \forall i.$$

Es ist naheliegend, dass im Falle, dass es mehr als ein Gibbs-Maß im unendlichen Volumen gibt, zwei verschiedene Gibbs-Maße als Limes zweier Folgen erhalten werden: Eine, bei der man sämtliche Spins am Rande von Λ_n (Λ_n eine Folge von Rechtecken, die gegen ∞ konvergieren) auf -1 setzt, und eine andere, bei der sämtliche dieser Spins auf $+1$ gesetzt sind. Um zu sehen, ob diese Folgen tatsächlich verschiedene Limiten besetzen, misst man die Wahrscheinlichkeit, dass der Spin im Ursprung $\sigma_0 = +1$ ist. Unter dem Gibbs-Maß $\mu^-_{\Lambda_N}$ auf Λ_N mit Randbedingungen -1, muss dieser positive Spin in 0 von den negativen Spins am Rande durch eine Schicht von Plus-Minus-Verbindungen getrennt sein (dies ist für das Gibbs-Maß $\mu^+_{\Lambda_n}$ auf Λ_n mit Randbedingungen $+1$ nicht der Fall). Dies ist für niedrige Temperaturen T, also große β, exponentiell teurer unter $\mu^-_{\Lambda_N}$. Es liegt daher nahe, dass es für hinreichend großes β zwei Gibbs-Maße im unendlichen Volumen geben sollte. Versuchen wir dies rigoros zu machen. Wir benötigen die folgende Definition:

Definition 5.5 *a) Sei* $< i,j >$, $i,j \in \mathbb{Z}^d$ *eine Kante des Gitters* $\mathbb{Z}^d$. *Mit* $< i,j >^*$ *bezeichnen wir die dazu duale Facette, also die (eindeutige)* $(d-1)$*-dim. Fläche, die die Kante* $< i,j >$ *in der Mitte schneidet.*

b) Für eine Spinkonfiguration $\sigma \in \{-1,+1\}^{\mathbb{Z}^d}$ *sei*

$$\Gamma(\sigma) := \{< i,j >^* : \sigma_i\sigma_j = -1\}.$$

$\Gamma(\sigma)$ *ist eine Fläche im* $\mathbb{R}^{d-1}$.

Direkt aus der Definition von $\Gamma(\sigma)$ erhält man

Lemma 5.6 *Für die Fläche* $\Gamma(\sigma)$ *sei* $\partial\Gamma(\sigma)$ *ihr* $(d-2)$*-dimensionaler Rand. Dann gilt*

a) $\partial\Gamma(\sigma) = \emptyset$ *für alle* $\sigma \in \{-1,+1\}^{\mathbb{Z}^d}$ *(obschon* $\Gamma(\sigma)$ *unbeschränkte Zusammenhangskomponenten besitzen kann).*

b) Es sei Γ eine Fläche im dualen Graphen des $\mathbb{Z}^d$, so dass $\partial\Gamma = \emptyset$. Dann gibt es genau zwei Konfigurationen σ und $-\sigma$, so dass

$$\Gamma(\sigma) = \Gamma(-\sigma) = \Gamma.$$

Definition 5.7 *Die Zusammenhangskomponenten γ einer Fläche Γ mit $\partial\Gamma = \emptyset$ nennen wir Konturen und schreiben $\gamma \in \Gamma$. Jedes $\gamma \in \Gamma$ erfüllt $\partial\gamma = \emptyset$. Jede Kontur ist daher entweder eine endliche geschlossene Fläche oder eine unendliche, unbeschränkte Fläche. Mit S_γ bezeichnen wir das Volumen, das von γ eingeschlossen wird, $|\gamma|$ bezeichnet die "Länge" von γ, also die Anzahl der Facetten in γ.*

Der folgende Satz macht unsere obigen Bemerkungen rigoros.

Satz 5.8 *(Peierls, Dobrushin, Griffith) Es sei μ_β ein Gibbs-Maß für das Ising-Modell in $d \geq 2$ Dimensionen mit äußerem Magnetfeld $h = 0$. Dann gibt es ein $\beta_d = \beta(d) < +\infty$, so dass für alle $\beta > \beta_d$ gilt*

$$\mu_\beta[\sigma : \exists\, \gamma \in \Gamma(\sigma) : 0 \in int\, \gamma] < \frac{1}{2}.$$

Hierbei bezeichnet int γ das Innere der Kontur γ.

Der Zusammenhang zu den Phasenübergangsüberlegungen eingangs ist dieser

Korollar 5.9 *Im d-dim. Ising-Modell gibt es für $\beta > \beta_d$ zwei extremale Gibbs-Maße μ_β^+ und μ_β^-. Diese erfüllen*

$$\mathbb{E}_{\mu_\beta^+}(\sigma_0) = -\mathbb{E}_{\mu_\beta^-}(\sigma_0) > 0.$$

Beweis des Korollars: Es sei $\Lambda_n \uparrow \mathbb{Z}^d$ eine Folge von Volumina, so dass μ_{β,Λ_n}^+ gegen ein Gibbs-Maß μ_β^+ konvergiert. Hierbei steht $+$ für die Randbedingung $\eta_j = 1\ \forall\, j$. Dann gilt, wie uns der Beweis des Satzes zeigen wird, gleichmäßig in n

$$\mu_{\beta,\Lambda_n}^+(\sigma_0 = -1) \leq \mu_{\beta,\Lambda_n}^+(\exists \gamma : 0\, int\, \gamma) \leq a < \frac{1}{2}.$$

Damit gilt auch

$$\mu_\beta^+(\sigma_0 = -1) \leq \lim_n \mu_{\Lambda,\beta_n}^+(\sigma_0 = -1) < \frac{1}{2}.$$

Da der Spin in 0 nur zwei mögliche Werte hat, folgt daraus die Behauptung. $\square$

Für den Beweis des Satzes genötigen wir noch das folgende vorbereitende Lemma:

Lemma 5.10 *Es sei μ_β ein Gibbs-Maß im Ising-Modell und γ eine endliche Kontur. Dann gilt*

$$\mu_\beta[\sigma : \gamma \in \Gamma(\sigma)] \leq 2e^{-2\beta|\gamma|}.$$

Beweis: Wir werden von der DLR-Charakterisierung von μ_β Gebrauch machen. Da γ geschlossen und endlich ist, kann man die Knoten entlang γ in zwei Mengen γ_{int} und γ_{out} unterteilen, wobei γ_{int} die inneren und γ_{out} die äußeren Knoten bezeichnet. Wenn $\gamma \in \Gamma(\sigma)$ ist, folgt, dass entweder $\sigma_{\gamma_{\text{int}}} \equiv -1$ und $\sigma_{\gamma_{\text{out}}} \equiv +1$ oder $\sigma_{\gamma_{\text{int}}} \equiv +1$ und $\sigma_{\gamma_{\text{out}}} \equiv -1$ sein muss.

Also:

$$\mu_\beta[\gamma \in \Gamma(\sigma)] = \mu_\beta[\sigma_{\gamma_{\text{out}}} \equiv +1, \sigma_{\gamma_{\text{int}}} \equiv -1] + \mu_\beta[\sigma_{\gamma_{\text{int}}} \equiv +1, \sigma_{\gamma_{\text{out}}} \equiv -1].$$

Da $\sigma_{\gamma_{\text{int}}} \equiv \pm 1$ ein Ereignis ist, das nur von den Spin in (einem genügend großen) Λ_n abhängt (nämlich Λ_n als das Innere von γ gewählt) und das Ising-Gibbs-Maß nur Randbedingungen der Reichweite 1 "wahrnimmt", folgt aus den DLR-Gleichungen:

$$\begin{aligned}\mu_\beta[\sigma_{\gamma_{\text{out}}} \equiv +1, \sigma_{\gamma_{\text{int}}} \equiv -1] &= \mu_\beta[\sigma_{\gamma_{\text{out}}} \equiv +1]\mu_\beta[\sigma_{\gamma_{\text{int}}} \equiv -1 | \sigma_{\gamma_{\text{out}}} \equiv +1] \\ &= \mu_\beta[\sigma_{\gamma_{\text{out}}} \equiv +1]\mu^{+1}_{\text{int}\gamma,\beta}[\sigma_{\gamma_{\text{int}}} \equiv -1].\end{aligned}$$

Der letzte Faktor aber lässt sich abschätzen. In der Tat: bezeichnet ρ das Produktmaß auf $\{-1,+1\}^{\mathbb{Z}^d}$, das jeder endlichen Spinkonfiguration die gleiche Wahrscheinlichkeit gibt, so gilt

$$\mu^{+1}_{\text{int}\,\gamma,\beta}[\sigma_{\gamma_{\text{int}}} \equiv -1] = \frac{\mathbb{E}_{\sigma_{\text{int}\gamma|\gamma_{\text{int}}}} \rho(\sigma_{\gamma_{\text{int}}} \equiv -1) e^{-\beta H_{\text{int}(\gamma)}(\sigma_{\text{int}\gamma|\gamma_{\text{int}}}, -1_{\gamma_{\text{int}}}, +1_{\gamma_{\text{out}}})}}{\mathbb{E}_{\sigma_{\gamma_{\text{int}}}} \mathbb{E}_{\sigma_{\text{int}\gamma|\gamma_{\text{int}}}} e^{-\beta H_{\text{int}\gamma}(\sigma_{\text{int}\gamma|\gamma_{\text{int}}}, \sigma_{\gamma_{\text{int}}}, +1_{\gamma_{\text{out}}})}}$$

(wobei wir die Definition des Gibbs'schen Maßes benutzen und $(\sigma_{\text{int}\gamma|\gamma_{\text{int}}}, -1_{\gamma_{\text{int}}}, +1_{go}$ bzw. $(\sigma_{\text{int}\gamma|\gamma_{\text{int}}}, \sigma_{\gamma_{\text{int}}}, +1_{\gamma_{\text{int}}})$ die Konfigurationen sind, bei denen die Spins in int $|\gamma_{\text{int}}$ auf $\sigma_{\gamma_{\text{int}}|\gamma_{\text{int}}}$, auf γ_{int} auf $\sigma_{\gamma_{\text{int}}}$ (bzw. -1) und auf γ_{int} auf $+1$ gesetzt sind.)

Schreibt man

$$\begin{aligned}& H_{\text{int}(\gamma)}(\sigma_{\text{int}\gamma\setminus\gamma_{\text{int}}}, -h_{\text{int}}, 1_{\gamma_{\text{int}}}) \\ = \; & H_{\text{int}(\gamma)\setminus\gamma_{\text{int}}}, (\sigma_{\text{int}\gamma\setminus\gamma_{\text{int}}}, -1_{\gamma_{\text{int}}}) \\ & + H_{\gamma_{\text{out}}\cup\gamma_{\text{int}}}(-1_{\gamma_{\text{int}}}, +1_{\gamma_{\text{out}}})\end{aligned}$$

und Analoges für $H_{\text{int}\gamma}(\sigma_{\text{int}\gamma\setminus\gamma_{\text{int}}}, \sigma_{\gamma_{\text{int}}}, +1_{\gamma_{\text{out}}})$ und erinnert die Definition von $Z_{\text{int}\gamma|\gamma_{\text{int}}}$, so erhält man

$$\mu^{+}_{\text{int}\gamma,\beta}[\sigma_{\gamma_{\text{int}}} \equiv -1] = \frac{e^{-\beta|\gamma| U^{(-1)}_{\text{int}\gamma|\gamma_{\text{int}}} \rho(\sigma_{\gamma_{\text{int}}} \equiv -1)}}{\mathbb{E}_{\sigma_{\gamma_{\text{int}}}} e^{\beta \sum_{i\in\gamma_{\text{int}}, j\in\gamma_{\text{out}}} \sigma_i} Z^{\sigma_{\gamma_{\text{int}}}}_{\text{int}\gamma|\gamma_{\text{int}}}}.$$

Nun gilt für den ersten Faktor im Neuner

$$\mathbb{E}_{\sigma_{\gamma_{\text{int}}}} e^{\beta \sum_{i\in\gamma_{\text{int}}, j\in\gamma_{\text{out}}} \sigma_i} \geq e^{\beta|\gamma|} 2^{-|\gamma_{\text{int}}|}$$

(das ist der Summand, den man erhält, wenn man alle σ_i, $i \in \gamma_{\text{int}}$, gleich plus 1 setzt). Da auch

$$\rho(\sigma_{\gamma_{\text{int}}} \equiv -1) = 2^{-|\gamma_{\text{int}}|}$$

ist, folgt

$$\mu^{+}_{\text{int}\gamma,\beta}[\sigma_{\gamma_{\text{int}}} \equiv -1] \leq \frac{e^{-2\beta|\gamma|} Z^{(-1)}_{\text{int}\gamma|\gamma_{\text{int}}}}{Z^{(+1)}_{\text{int}\gamma|\gamma_{\text{int}}}}.$$

Nun ist aber $Z^{(-1)}_{\mathrm{int}\gamma\setminus\gamma_{\mathrm{int}}} = Z^{(+1)}_{\mathrm{int}\gamma\setminus\gamma_{\mathrm{int}}}$, da die Hamiltonfunktion H das Ising-Modell mit $h = 0$ (dies ist hier wichtig) invariant ist unter dem Flip $\sigma_i \to -\sigma_i$ für alle $i \in \mathbb{Z}^d$. Also

$$\mu^+_{\mathrm{int}\gamma,\beta}[\sigma_{\gamma_{\mathrm{int}}} \equiv -1] \leq 4e^{-2\beta|\gamma|}$$

Behandelt man

$$\mu_\beta[\sigma_{\mathrm{out}} \equiv -1, \sigma_{\gamma_{\mathrm{int}}} \equiv +1]$$

ebenso, ergibt sich ein zweiter Summand $e^{-2\beta|\gamma|}$, also zusammen

$$\mu_\beta[\sigma : \gamma \in \Gamma(\sigma)] \leq 2e^{-2\beta|\gamma|}.$$

□

Beweis des Satzes: Offenbar gilt

$$\mu_\beta[\exists\gamma \in \Gamma(\sigma) : 0 \in \operatorname{int}\gamma] \leq \sum_{\gamma:0\in\operatorname{int}\gamma} \mu_\beta[\gamma \in \Gamma(\sigma)] \leq 2 \sum_{\gamma:0\in\operatorname{int}\gamma} e^{-2\beta|\gamma|}.$$

Bleibt also die Menge

$$\{\gamma : 0 \in \operatorname{int}\gamma,\ |\gamma| = h\}$$

abzuzählen. Da uns irgendeine Schranke für die kritische Temperatur genügt, können wir die Anzahl durch $k \cdot (2d)^{2d_k}$ beschränken. Dies sieht man ein, indem man irgendeine Facette von γ als "Start" wählt. Es gibt dafür k Möglichkeiten. Da eine Facette $2d$ "Flächen" besitzt, erhält man $2d$ Wahlmöglichkeiten für die Anschlussfacette. Von dort kann man insgesamt höchstens $2d$ Richtungen gehen. Dies ergibt, dass wir höchstens $k(2d)^{2d_k}$ Elemente in der fraglichen Menge haben. In der Tat sieht man mit der gleichen Technik schnell ein, dass es in $d = 2$ höchstens 3^k Elemente in dieser Menge gibt. Diese Schranke fand Ruelle auch für allgemeines d. Wir erhalten

$$\mu_\beta[\sigma : \exists\gamma, 0 \in \mathrm{int}\gamma] \leq \sum_{k=4}^{\infty} k e^{k(-2\beta+\log 2d)}.$$

Wählt man β hinreichend groß, ist diese Reihe echt kleiner als $\frac{1}{2}$. □

Korollar 5.11 *Uniform in Λ gilt*

$$\mu^+_{\Lambda,\beta}(\sigma(0) = -1) \underset{\beta\to\infty}{\longrightarrow} 0.$$

Beweis: Folgt direkt aus dem vorhergehenden Beweis. □

6 FKG-Ungleichungen, Monotonie und Phasenübergänge

Das im vorigen Kapitel vorgestellte Peierlsche Argument erlaubt es zu zeigen, dass es im Unendlichen mehr als ein Gibbs-Maß gibt – es sagt jedoch nicht, wie wir ein solches erhalten. Wie schon in der Motivation dieses Arguments beschrieben liegt es nahe zu vermuten, dass man solche verschiedenen Gibbs-Maße erhält, indem man die lokalen Gibbs-Maße auf einer Folge von Würfeln $\Lambda_n \to \infty$ betrachtet und diese einmal mit Plus- und einmal mit Minus-Randbedingungen versieht. Dies ist lustigerweise nicht so einfach zu beweisen und erfordert den Einsatz sogenannter Korrelationsungleichungen, von denen die FKG-Ungleichungen zu den erfolgreichsten zählen.

Definition 6.1 *Es sei S der Zustandsraum für die Einzelspins. S sei linear geordnet. Für eine endliche Teilmenge Λ des $\mathbb{Z}^d$ bezeichne μ ein Wahrscheinlichkeitsmaß auf S^Λ. Wir sagen, dass μ* POSITIV KORRELIERT IST *(bzw. μ die Spins positiv korreliert) oder dass μ die* FKG-UNGLEICHUNGEN ERFÜLLT, *falls für alle beschränkten, $\mathcal{F}_\Lambda$ messbaren Funktionen*

$$f, g : S^\Lambda \to \mathbb{R},$$

die bezüglich der partiellen Ordnung auf S^Λ, die durch die Ordnung auf S induziert wird, steigend sind, gilt

$$\mu(f \cdot g) := \mathbb{E}\mu(fg) \geq \mu(f) \cdot \mu(g).$$

Bemerkungen:

1. Die Bezeichnung FKG-Ungleichungen geht auf die theoretischen Physiker Fortuin, Kasteleyn und Ginibre zurück.
2. Ist $\Lambda = \{x_0\}$ und $S \subseteq \mathbb{R}$, so sind die FGK-Ungleichungen für jedes Wahrscheinlichkeitsmaß μ auf S erfüllt. In der Tat gilt ja stets
$$\mu(fg) - \mu(f)\mu(g) = \frac{1}{2} \int \int (f(\sigma) - f(\tau))(g(\sigma) - g(\tau))(\mu(\sigma) d\mu(\tau)).$$
Nun sind für $S^{\{x_0\}}$ (und in jedem vollständig geordneten Raum), je zwei Elemente miteinander vergleichbar, und da sowohl f als auch g steigend sind, haben $f(\sigma) - f(\tau)$ und $g(\sigma) - g(\tau)$ stets das gleiche Vorzeichen. Also ist
$$\mu(fg) - \mu(f)\mu(g) \geq 0.$$

Theorem 6.2 *Sei $|S| = 2,\ S = \{-1, +1\}$. Dann gelten für das Ising-Gibbs-Maß $\mu_{\Lambda,\beta}$ zu beliebigen Volumina Λ und beliebigen Temperaturen $\frac{1}{\beta}$ die FKG-Ungleichungen.*

Bemerkung: Dies gilt sogar für beliebige ferromagnetische Paarwechselwirkungen, soll hier aber nur für das Ising-Modell gezeigt werden. Wir werden für die lokalen Spezifikationen $\mu^{(\eta)}_{\Lambda,\beta}$ nicht nur $\eta \in \{-1, +1\}^{\mathbb{Z}^d | \Lambda}$, sondern sogar $\eta \in \mathbb{R}^{\mathbb{Z}^d | \Lambda}$ zulassen.

Beweis: Via Induktion über $|\Lambda|$. Für $|\Lambda| = 1$ gilt die Behauptung trivialerweise, wie oben bemerkt. Angenommen, die Behauptung gilt für $\Lambda \subseteq \mathbb{Z}^d$. Sei $y \in \Lambda^c$. Wir wollen die Gültigkeit der FKG-Ungleichungen für $\Lambda' = \Lambda \cup \{y\}$ nachweisen. Hierbei seien

$$f, g : S^{\Lambda'} \to \mathbb{R}$$

beschränkte, $\mathcal{F}_{\Lambda'}$-messbare, wachsende Funktionen. Da die lokale Spezifikation $\mu^{\eta}_{\Lambda',\beta}$ mit $\mu^{\eta}_{\Lambda,\beta}$ verträglich ist, können wir den Spin "ausintegrieren", indem wir ihn als Teil der Randbedingung auffassen:

$$\begin{aligned}\mu^{\eta}_{\Lambda',\beta}(f \cdot g) &= \sum_{\eta_y = \pm 1} \mu^{\eta}_{\Lambda',\beta}(\sigma_y = \eta_y)(\mu^{\eta}_{\Lambda,\beta}(fg)) \\ &\geq \sum_{\eta_y = \pm 1} \mu^{\eta}_{\Lambda',\beta}(\sigma_y = \eta_y)(\mu^{\eta}_{\Lambda,\beta}(f))(\mu^{\eta}_{\Lambda,\beta}(g))\end{aligned}$$

nach Induktionsvoraussetzung. Wir fassen nun

$$\eta_y \mapsto \mu^{\eta}_{\Lambda,\beta}(f) \quad \text{und} \quad \eta_y \mapsto \mu^{\eta}_{\Lambda,\beta}(g) \tag{6.1}$$

als Funktionen in η_y auf und die obige Summe auf der rechten Seite der letzten Gleichung als den Erwartungswert des Produkts. Kann man zeigen, dass die Abbildungen in (6.1) wachsend sind, so kann man mit der einelementigen Menge $\{y\} \subseteq \mathbb{Z}^d$ die FKG-Ungleichungen anwenden und ist fertig. Bleibt der Monotonienachweis: Zu zeigen ist also

$$\eta_y \mapsto \mu^{(\eta_{\Lambda'^c}, \eta_y)}_{\Lambda,\beta}(f(\sigma_\Lambda, \eta_y))$$

ist steigend, also

$$\mu^{(\eta_{\Lambda'^c}, +1)}_{\Lambda,\beta}(f(\sigma_\Lambda, +1)) \geq \mu^{(\eta_{\Lambda'^c}, -1)}_{\Lambda,\beta}(f(\sigma_\Lambda, -1)).$$

Da $f(\sigma_\Lambda, +1) \geq f(\sigma_\Lambda, -1)$, genügt es auch

$$\mu^{(\eta_{\Lambda'^c}, +1)}_{\Lambda,\beta}(f(\sigma_\Lambda, +1)) \geq \mu^{(\eta_{\Lambda^c}, -1)}_{\Lambda,\beta}(f(\sigma_\Lambda, +1))$$

zu beweisen. Fassen wir nun, wie angekündigt, η als aus reellen Variablen bestehend auf, so können wir bezüglich η_y differenzieren. Tun wir dies, so erhalten wir

$$\begin{aligned}\frac{d}{d\eta_y}\mu^{(\eta_{\Lambda'^c}, \eta_y)}_{\Lambda,\beta}(f(\sigma_\Lambda, +1)) &= \frac{d}{d\eta_y}\sum e^{+\beta\sum_{j,i\in\Lambda}\sigma_i\sigma_j + \beta\sum_{\substack{i\in\Lambda \\ <i,y>}}\sigma_i\eta_y + h\sum_{i\in\Lambda}\sigma_i} \\ &\quad \times f(\sigma_\Lambda, +1)/Z^{(\eta_{\Lambda'^c}, \eta_y)}_{\Lambda,\beta} \\ &= \mu^{(\eta_{\Lambda'^c}, \eta_y)}_{\Lambda,\beta}(\beta f(\sigma_\Lambda, +1)\sum_{<i,y>}\sigma_i) \\ &\quad -\mu^{(\eta_{\Lambda'^c}, \eta_y)}_{\Lambda,\beta}(\beta\sum_{<i,y>}\sigma_i)\mu^{(\eta_{\Lambda'^c}, \eta_y)}_{\beta,\Lambda}(f(\sigma_\Lambda, \eta_y)) \geq 0.\end{aligned}$$

Die Ungleichung folgt dabei, da auch $\sum \sigma_i$ eine nicht-fallende Funktion ist und man noch einmal die FKG-Ungleichungen der Induktionsvoraussetzungen anwenden kann. Dies beweist die Behauptung. □

Nun wollen wir zeigen, dass unsere Mühe nicht umsonst war und man aus den FKG-Ungleichungen Nützliches ableiten kann.

Lemma 6.3 *Es seien* $(\mu^{\eta}_{\Lambda,\beta})$ *die lokalen Spezifikationen eines Gibbs-Maßes, das die FKG-Ungleichungen erfüllt. Es sei* $+$ *die Konfiguration mit* $\eta_x \equiv 1$ *für alle* $x \in \mathbb{Z}^d$*. Dann gilt*

1. *Für jedes* $\Lambda \subseteq \mathbb{Z}^d$ *und jedes* $\eta \in S^{\mathbb{Z}^d}$ *und jede steigende Funktion*
$$f : S^{\Lambda} \to \mathbb{R}$$
gilt
$$\mu^{+}_{\Lambda,\beta}(f) \geq \mu^{\eta}_{\Lambda,\beta}(f).$$

2. *Für jedes Paar von Volumina* $\Lambda_2 \supseteq \Lambda_1$ *und jede steigende Funktion* $f : S^{\Lambda_1} \to \mathbb{R}$ *gilt*
$$\mu^{+}_{\Lambda_2,\beta}(f) \leq \mu^{+}_{\Lambda_1,\beta}(f).$$

Beweis: Betrachte nur den Fall $S = \{-1, +1\}$.

1. Sei $x \in \Lambda^c$ und $\eta_x \in [-1, +1]$ für alle x. Wir werden zeigen, dass für alle x
$$\eta_x \mapsto \mu^{\eta}_{\Lambda,\beta}(f)$$
steigend ist in η_x, dann folgt 1. sofort. Nun gilt
$$\frac{\partial}{\partial \eta_x}\mu^{\eta}_{\Lambda,\beta}(f) = \sum_{y\in\Lambda} \beta J_{xy}(\mu^{\eta}_{\Lambda,\beta}(\sigma_y f) - \mu^{\eta}_{\Lambda,\beta}(\sigma_y)\mu^{\eta}_{\Lambda,\beta}(f)),$$
wobei wir annehmen, dass das Gibbs-Maß bezüglich eines ferromagnetischen Paar-Potentials mit Hamiltonfunktion
$$H(\sigma) = -\sum_{x,y} J_{xy}\sigma_x\sigma_y,$$
$J_{xy} \geq 0$, gebildet wird. Nun ist sowohl f als auch
$$g(\sigma) = \sigma_y$$
wachsend. Da die J_{xy} allesamt nicht-negativ sind, folgt aus den FKG-Ungleichungen
$$\frac{\partial}{\partial \eta_x}\mu^{\eta}_{\Lambda,\beta}(f) \geq 0.$$

2. Einerseits gilt aufgrund der FKG-Ungleichungen
$$\mu^{+}_{\Lambda_2,\beta}(\mathbb{1}_{+1_{\Lambda_2\setminus\Lambda_1}} \cdot f) \geq \mu_{\Lambda_2,\beta}(\mathbb{1}_{+1_{\Lambda_2\setminus\Lambda_1}})\mu_{\Lambda_2,\beta}(f),$$
andererseits folgt wegen der DLR-Gleichungen:
$$\mu^{+}_{\Lambda_2,\beta}(\mathbb{1}_{+1_{\Lambda_2\setminus\Lambda_1}} \cdot f) = \mu^{+}_{\Lambda_2,\beta}(\mathbb{1}_{+1_{\Lambda_2\setminus\Lambda_1}}) \cdot \mu_{\Lambda_1,\beta}(f).$$
Daraus folgt die Behauptung.

□

Korollar 6.4 1. *Für jede Folge $\Lambda_n \uparrow \infty$ existieren die Limiten*

$$\lim_{n\to\infty} \mu^+_{\Lambda_n,\beta} =: \mu^+_\beta$$

und

$$\lim_{n\to\infty} \mu^-_{\Lambda_n,\beta} =: \mu^-_\beta$$

und sind unabhängig von der Folge (Λ_n).

2. *Für alle Gibbs-Maße μ_β bezüglich derselben Wechselwirkung und alle steigenden, beschränkten und stetigen Funktionen f gilt*

$$\mu^-_\beta(f) \le \mu_\beta(f) \le \mu^+_\beta(f).$$

Beweis:

1. Für alle Maße ν gilt $\nu(f) \in [-\|f\|_\infty, \|f\|_\infty]$. Außerdem ist für steigende Funktionen f

$$\Lambda_n \mapsto \mu^+_{\Lambda_n,\beta}(f)$$

nach Lemma 6.3 monoton, somit existiert

$$\lim_{n\to\infty} \mu^+_{\Lambda_n,\beta}(f).$$

Gleichermaßen zeigt man die Existenz von

$$\lim_{n\to\infty} \mu^-_{\Lambda_n,\beta}(f).$$

Nun seien (Λ_n), (Λ'_n) zwei Folgen, die gegen ∞ aufsteigen. Dann gibt es Teilfolgen (Λ_{n_k}), $(\Lambda'_{n'_k})$ mit

$$\Lambda_{n_k} \subset \Lambda'_{n'_k} \subset \Lambda_{n_{k+1}}.$$

Die Monotonieüberlegungen aus Lemma 6.3 ergeben nun, dass einerseits

$$\lim_{k\to\infty} \mu^+_{\Lambda_{n_k},\beta}(f) \ge \lim_{k\to\infty} \mu^+_{\Lambda'_{n'_k}}(f)$$

und andererseits

$$\lim_{k\to\infty} \mu^+_{\Lambda_{n_{k+1}},\beta}(f) \le \lim_{k\to\infty} \mu^+_{\Lambda'_{n'_k},\beta}(f),$$

also sind die Limiten gleich. Für μ^-_β argumentiert man analog. Schließlich approximiert man stetige beschränkte f durch Linearkombinationen monotoner f.

2. Dies folgt unmittelbar daraus, dass man die gewünschten f durch $\mathcal{F}_\Lambda$-messbare Funktionen approximieren kann ($\Lambda \to \infty$) und für Λ-messbare Funktionen die gewünschten Ungleichungen sofort aus Teil 1 des vorhergehenden Lemmas folgen.

□

Eine interessante und wichtige Folge der FKG-Ungleichungen ist es, dass beispielsweise im Ising-Modell (und etlichen anderen Spinsystemen) die Frage der Existenz von mehreren Gibbs-Maßen im unendlichen Volumen, d. h. die Frage nach einem Phasenübergang, mit einer makroskopischen Observablen, einem sogenannten ORDNUNGSPARAMETER, verbunden werden kann.

Definition 6.5 *Für ein Gibbs-Maß μ setze*

$$m^{\mu} := \lim_{\Lambda\uparrow\infty} \frac{1}{|\Lambda|} \sum_{x\in\Lambda} \mu(\sigma_x),$$

falls dieser Limes existiert. Wir schreiben auch

$$m_\beta^{\pm} := m^{\mu_\beta^{\pm}}.$$

Satz 6.6 *Im Ising-Modell gilt $\mu_\beta^+ = \mu_\beta^-$ genau dann, wenn $m_\beta^+ = m_\beta^-$ gilt.*

Bemerkungen:

1. Dieser Satz gilt auch allgemeiner für translationsinvariante Systeme, in denen die FKG-Ungleichungen gelten.
2. Da stets $\mu_\beta^+(f) \geq \mu_\beta(f) \geq \mu_\beta^-(f)$ für alle Gibbs-Maße μ_β und alle steigenden, monotonen f gilt, sind μ^+ und μ^- extremal, d. h. mit $\mu_\beta^+ = \mu_\beta^-$ sind alle Gibbs-Maße identisch.

Lemma 6.7 *Ein Modell mit Ising – d. h. ± 1-wertigen Spins – und FKG-Ungleichungen erfüllt für alle endlichen Mengen $A, B \subseteq \Lambda$ die Ungleichungen*

$$\begin{aligned} & \mu_\beta^+(\sigma_{A\cup B} = +1) - \mu_\beta^-(\sigma_{A\cup B} = +1) \\ \leq\ & \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1) + \mu_\beta^+(\sigma_B = +1) - \mu_\beta^-(\sigma_B = +1). \end{aligned}$$

Beweis: Es gilt:

$$\mathbb{1}_{\sigma_A=+1\wedge\sigma_B=+1} = \mathbb{1}_{\sigma_A=+1} + \mathbb{1}_{\sigma_B=+1} - \mathbb{1}_{\sigma_A=+1\vee\sigma_B=+1}.$$

Hieraus folgt per Integration bzgl. μ_β^+ und μ_β^-

$$\begin{aligned} & \mu_\beta^+(\sigma_{A\cup B} = +1) - \mu_\beta^-(\sigma_{A\cup B} = -1) \\ =\ & \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1) + \mu_\beta^+(\sigma_B = +1) - \mu_\beta^-(\sigma_B = +1) \\ +\ & \mu_\beta^-(\sigma_A = +1) \vee \sigma_B = +1) - \mu_\beta^+(\sigma_A = +1) \vee \sigma_B = +1). \end{aligned}$$

Nun ist aber $\mathbb{1}_{\sigma_A=+1\vee\sigma_B=+1}$ eine wachsende Funktion, so dass nach Teil 2 des vorhergehenden Lemmas gilt

$$\mu_\beta^-(\sigma_A = +1 \vee \sigma_B = +1) - \mu_\beta^+(\sigma_A = +1 \vee \sigma_B = +1) \leq 0.$$

Hieraus folgt die Behauptung des Lemmas. □

Beweis des Satzes: Ein Vorteil des Ising-Modells besteht darin, dass für jedes endliche Λ die $\mathcal{F}_\Lambda$-messbaren (lokalen) Funktionen durch die Indikatoren $\mathbb{1}_{\sigma_A=+1}$ (und skalare Vielfache hiervon) ausgedrückt werden können. (Übung: Man überlege sich das.) Aus Lemma 6.7 erhalten wir

$$0 \leq \mu_\beta^+(\sigma_A = +1) - \mu_\beta^-(\sigma_A = +1) \leq \sum_{i \in A} \mu_\beta^+(\sigma_i = +1) - \mu_\beta^-(\sigma_i = +1).$$

Falls nun $\mu_\beta^+ = \mu_\beta^-$, folgt selbstverständlich $m_\beta^+ = m_\beta^-$. Umgekehrt ist in dem (translationsinvarianten) Ising-Modell

$$\mu_\beta^\pm(\sigma_x) = \mu_\beta^\pm(\sigma_y) \quad \text{für alle} \quad x, y \in \mathbb{Z}^d.$$

Somit impliziert $m_\beta^+ = m_\beta^-$ auch $\mu_\beta^+(\sigma_i = +1) = \mu_\beta^-(\sigma_i = +1)$ für alle $i \in \mathbb{Z}^d$. Nach den obigen Überlegungen ist dies aber gleichbedeutend mit

$$\mu_\beta^+(f) = \mu_\beta^-(f)$$

für alle $\mathcal{F}_\Lambda$-messbaren, beschränkten Funktionen f und alle Λ. Daraus folgt $\mu_\beta^+ = \mu_\beta^-$. □

Bemerkungen:

1. Das Konzept der Ordnungsparameter wird uns noch in das Gebiet der ungeordneten Systeme begleiten. Hier sei nur angemerkt, dass
$$m = \frac{1}{N} \sum \sigma_i$$
auch im Curie-Weiss-Modell ein Ordnungsparameter ist.

2. Wir haben gesehen, dass μ_β^+ und μ_β^- in gewissem Sinne extreme Gibbs-Maße sind. Dies lässt sich formalisieren, indem man zeigt, dass gilt, falls
$$\mu_\beta^\pm = \alpha\nu_1 + (1 - \alpha)\nu_2$$
für zwei Gibbs-Maße ν_1 und ν_2, so ist $\alpha \in \{0, 1\}$ und entweder ν_1 oder ν_2 gleich $\mu_\beta^\pm$.

3. Es liegt nahe zu vermuten, dass μ_β^+ und μ_β^- die einzigen extremalen Gibbs-Maße sind. Lustigerweise stimmt dies nur für TRANSLATIONSINVARIANTE Gibbs-Maße oder für $d = 2$. In $d \geq 3$ gibt es für niedrige Temperaturen nicht-translationsinvariante, extremale Gibbs-Maße, die verschieden sind von μ_β^+ und μ_β^-.

7 Cluster-Entwicklung

Ein generelles mathematisches Konzept ist es, Situationen, die man nicht versteht, durch solche zu approximieren, die man besser versteht. Im Ising-Modell oder allgemein in vielen Spin-Modellen, gibt es zwei Situationen, die man recht gut versteht: Den extremen Hochtemperaturfall $\beta = 0$, bei dem alle Konfigurationen gleich wahrscheinlich sind, und den Fall von Temperatur 0, d. h. $\beta = +\infty$, bei dem das System in den Grundzuständen, d. h. den Zuständen minimaler Energie, "hängenbleibt". Wir wollen in diesem Kapitel Entwicklungen um diese Zustände kurz anreißen.

7.1 Hochtemperaturentwicklungen

Hier werden wir annehmen, dass β sehr klein ist. Wir wollen uns auf Entwicklungen der Zustandssumme und den Fall des Ising-Modells mit $h = 0$ beschränken. Wir betrachten also

$$Z_{\Lambda,\beta} = \frac{1}{2^{|\Lambda|}} \sum_{\sigma \in \{\pm 1\}^\Lambda} e^{\beta \sum_{<i,j>} \sigma_i \sigma_j}.$$

Um nicht vollständig den Überblick zu verlieren, werden wir den Fall sogenannter "periodischer Randbedingungen" annehmen, d. h. wir stellen uns das Modell auf einem d-dimensionalen Torus vor. Wir betrachten den Fall $d = 2$. Nun gilt

$$e^x = \cosh(x)(1 + \tanh(x)).$$

Wendet man dies an, so ergibt sich

$$Z_{\Lambda,\beta} = \mathbb{E}_\sigma(\prod_{<i,j>} \cosh(\beta \sigma_i \sigma_j) \prod_{<i,j>} 1 + th(\sigma_i \sigma_j \beta)).$$

Nun ist $\cosh(\cdot)$ eine gerade Funktion und $th(\cdot)$ eine ungerade Funktion, also

$$Z_{\Lambda,\beta} = \cosh(\beta)^{2|\Lambda|} \mathbb{E}_\sigma[\prod_{<i,j>} (1 + \sigma_i \sigma_j th(\beta)],$$

da $\cosh(\sigma_i \sigma_j \beta)$ stets gleich $\cosh(\beta)$ ist und es $2|\Lambda|$ Verbindungen gibt. Es bleibt, den zweiten Term auszuwerten. Hierzu bietet es sich an, das Produkt zu entwickeln

$$\mathbb{E}_\sigma[\prod_{<i,j>} [1 + \sigma_i \sigma_j th(\beta)]] = 1 + \sum_{k=1}^\infty \mathbb{E}_\sigma \prod_{\ell=1}^k \sigma_{i_\ell} \sigma_{j_k} th(\beta).$$

Nun verschwinden beim Erwartungswert auf der rechten Seite alle Ausdrücke, bei denen es ein i gibt, das ungerade häufig vorkommt. In der Tat sind die einzelnen σ_i ja unabhängig und es gilt $\mathbb{E}(\sigma_i) = 0$. Dies reduziert den Erwartungswert zu

$$\mathbb{E}_\sigma[\prod_{<i,j>} (1 + \sigma_i \sigma_j th(\beta))] = 1 + \sum_{k=1}^\infty c_k th(\beta)^k.$$

Hierbei bezeichnet c_k die Anzahl der Graphen auf Λ mit k Kanten, bei denen jeder Knoten gerade häufig vorkommt (und natürlich nur die Kanten des $\mathbb{Z}^d$ verwendet werden dürfen). Offenbar gilt

$$c_1 = c_2 = c_3 = 0 \quad \text{und} \quad c_4 = |\Lambda|.$$

Letzteres, da ein Quadrat eindeutig bestimmt ist, wenn man seinen linken oberen Eckpunkt festlegt. Weiter gilt $c_5 = c_7 = 0$ und allgemein $c_{2n+1} = 0$ für alle n. Es ist

$$c_6 = 2|\Lambda|.$$

Da es zwei "Grundfiguren" eines geschlossenen Graphen mit 6 Kanten gibt (eine "liegende" und eine "stehende"). Die Koeffizienten c_{2k} zu berechnen wird für größer werdendes k immer schwieriger. Kac und Ward gaben 1953 eine geschlossene Form für diese Koeffizienten in Form einer Determinante an. Tatsächlich führen die dort angegebenen Werte für c_{2k} auf die exakte Lösung des 2-dimensionalen Ising-Modells, die schon Onsager 1944 berechnet hatte. Demnach sind diese Werte nicht für alle k korrekt (abgesehen davon, dass Kac und Ward keinen mathematischen Beweis gaben). 1960 konnte Sherman basierend auf Ideen von Feynman die Werte von c_{2k} exakt bestimmen. Leider ist die Methode zu kompliziert, um hier dargestellt zu werden (und außerdem ist das Paper von Sherman kaum lesbar). Festzuhalten bleibt, dass wir im 2-dimensionalen Ising-Modell die folgende Entwicklung für die Zustandssummen erhalten haben:

$$Z^{per}_{\Lambda,\beta} = (\cosh(\beta))^{2|\Lambda|}(1 + |\Lambda| th(\beta)^4 + O(\beta^6))$$

für $\beta \to 0$. Tatsächlich lässt sich hier noch viel mehr zeigen, wir wollen jedoch nicht weiter in dieses Kapitel eintauchen.

7.2 Tieftemperaturentwicklungen

Eine Grundidee der Tieftemperaturentwicklung haben wir schon im vorhergehenden Kapitel beim Peierlschen Argument kennengelernt, die Zerlegung in Konturen. Generell lässt sich sagen, dass das Tieftemperaturverhalten der meisten Modelle weit komplexer und schwieriger zu analysieren ist als ihr Hochtemperaturverhalten. Wir werden auch hier relativ an der Oberfläche bleiben. Wir betrachten wieder das Ising-Modell mit verschwindendem äußeren Magnetfeld $h = 0$, also die Hamiltonfunktion

$$H(\sigma) = -\sum_{<i,j>} \sigma_i \sigma_j, \quad \sigma \in \{-1, +1\}^{\mathbb{Z}^d}.$$

Es ist klar, dass eine additive Konstante in der Hamiltonfunktion zu den gleichen Gibbs-Maßen führt, da sich diese durch die Normierung in der Partitionsfunktion wieder herauskürzt. Eine Multiplikation der Hamiltonfunktion mit einer Konstanten (z. B. $\frac{1}{2}$) führt hingegen dazu, dass die Temperatur mit der inversen dieser Konstanten multipliziert werden muss (z. B. 2). Also ist die lokale Ising-Hamilton-Funktion im wesentlichen durch

$$H_\Lambda(\sigma) = +\sum_{<x,y>\cap\Lambda\neq\emptyset} \mathbb{1}_{\sigma_x \neq \sigma_y}.$$

gegeben. Wir betrachten der Einfachheit halber nur $\oplus$-Randbedingungen oder $\ominus$-Randbedingungen. Somit ist

$$H_\Lambda(\sigma) = vol(\Gamma(\sigma)) =: vol(\{< x, y > \cap \Lambda = \emptyset, \sigma_x \neq \sigma_y\}).$$

Somit lässt sich die Zustandssumme wie folgt schreiben:

$$Z_{\Lambda,\beta} = \sum_{\Gamma} \sum_{\sigma:\Gamma(\sigma)=\Gamma} e^{-\beta|\Gamma|},$$

wobei $|\Gamma|$ das Volumen von Γ bezeichnet. Γ kann, wie schon beim Peierlschen Argument aufgeführt, in Konturen zerlegt werden. Diese sind im Falle des 2d-Ising-Modells geschlossene Kurven, die die Gebiete mit positivem Spin und die Gebiete mit negativen Spin voneinander trennen. Somit gilt

$$Z_{\Lambda,\beta} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\gamma_1,\dots,\gamma_k} \sum_{\sigma:\Gamma(\sigma)=(\gamma_1,\dots,\gamma_k)} \prod_{\ell=1}^{k} e^{-\beta|\gamma_c|}.$$

Der Faktor $k!$ kommt dabei dadurch ins Spiel, dass die $\gamma_1, \dots, \gamma_k$ die Menge Γ nur bis auf Vertauschung der Reihenfolge eindeutig bestimmen. Bedenkt man nun, dass es zu jeder Menge von Konturen eine Spin-Konfiguration σ gibt, so dass $\Gamma(\sigma)$ gerade die Menge dieser Konturen ergibt und dass dieses σ sogar eindeutig ist, wenn man eine konstante Randbedingung vorgibt, so folgt

$$Z_{\Lambda,\beta} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\substack{\gamma_1,\dots,\gamma_k \\ \gamma_i \not\sim \gamma_j \\ \forall\, i \neq j}} \prod_{\ell=1}^{k} e^{-\beta|\gamma_\ell|}.$$

Hierbei bedeutet $\gamma_i \not\sim \gamma_j$, dass γ_i und γ_j nicht zusammenhängend sind. Dies ist die sogenannte Polymer-Darstellung von $Z_{\Lambda,\beta}$.

Es sollte darauf hingewiesen werden, dass die Tieftemperaturentwicklung des Ising-Modells leicht zu dem (verkehrten) Schluss verführt, die Tieftemperaturentwicklung sei immer so einfach. Tatsächlich macht es die Symmetrie der Ising-Hamilton-Funktion möglich, einen Großteil der grundsätzlichen Schwierigkeiten zu vertuschen. Schon der Fall $h \neq 0$ erlaubt eine solche Darstellung nicht mehr (allerdings gibt es dort auch keine Tieftemperaturphase). Insgesamt führt der hier geschilderte Zugang zur sogenannten Pirogov-Sinai-Theorie, einer der mächtigsten Hilfsmittel bei der Analyse der Tieftemperaturphase von Spin-Systemen. Diese Theorie ist so umfangreich, dass wir sie hier nicht behandeln können. Wir verweisen auf einen Übersichtsartikel von Borgs und Kotecky [2] für einen näheren Einblick.

8 Algebraische Lösung des 2D Ising-Modells

Die Matrixmethode, die wir schon zur Berechnung der freien Energie des eindimensionalen Ising-Modells verwendet hatten, bietet sich auch in Dimension 2 an. Hierzu stellen wir uns das Ising-Modell auf einem 2-dimensionalen Zylinder Λ aufgerollt vor und schreiben die Hamiltonfunktion als

$$H_\Lambda(\sigma) = -\sum_{<i,j>} \sigma_i\sigma_j = -\sum_{i=1}^{m-1}\sum_{j=1}^{m} \sigma_{i,j}\sigma_{i+1,j} - \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma_{i,j}\sigma_{i,j+1}.$$

Fassen wir alle Spins $(\sigma_{i,j})_{i=1}^m$ zu einem Spin $S_j = (\sigma_{1,j}, \sigma_{2,j}, \dots, \sigma_{m,j})$ zusammen und schreiben

$$V_1(S_j) = -\sum_{i=1}^{m-1} \sigma_{i,j}\sigma_{i+1,j}$$

als Energie der j-ten Spalte und

$$V_2(S_j, S_{j+1}) = -\sum_{i=1}^{m} \sigma_{i,j}\sigma_{i,j+1}$$

als Wechselwirkung zwischen den Spalten, so ergibt sich

$$H_\Lambda(\sigma) = \sum_{j=1}^{m} V_1(S_j) + V_2(S_j, S_{j+1}).$$

Dies ergibt für die Zustandssumme

$$Z_{\Lambda,\beta} = Z_{\Lambda,\beta,0} = \sum_{S_1,\dots,S_m} e^{-\beta\sum_{j=1}^m V_1(S_j)+V_2(S_j,S_{j+1})} = \sum_{S_1,\dots,S_m} L(S_1,S_2)L(S_2,S_3)\dots L(S_m,S_1).$$

Hierbei haben wir für $S = (\sigma_1,\dots,\sigma_m)$ und $S' = (\sigma'_1,\dots,\sigma'_m)$

$$L(S,S') = e^{-\beta V_1(S)}e^{-\beta V_2(S,S')} = e^{\beta\sum_{i=1}^{m-1}\sigma_i\sigma_{i+1}}e^{\beta\sum_{i=1}^{m}\sigma_i\sigma'_i}$$

gesetzt. Dies alles sieht der eindimensionalen Situation sehr ähnlich und ebenso wie dort ist

$$Z_{\Lambda,\beta} = Tr(L^m) = \sum_{j=1}^{2^m} \lambda_j^m,$$

nur dass nun L $2^m \times 2^m$-dimensional ist. Man erhält

$$\lim \frac{1}{|\Lambda|}\log Z_{\Lambda,\beta} = \lim_{m\to\infty}\frac{1}{m}\log\lambda_1,$$

wobei aber diesmal λ_1 der größte Eigenwert einer $2^m \times 2^m$-Matrix ist. Diesen wollen wir nun berechnen. Hierfür erinnern wir uns an den eindimensionalen Fall mit $h = 0$. Die Matrix L hatte dort die Gestalt

$$L = \begin{pmatrix} e^\nu & e^{-\nu} \\ e^{-\nu} & c^\nu \end{pmatrix} = I_2 e^\nu + \tau^1 e^{-\nu},$$

wobei $\nu = 2\beta,\ I_2$ die 2×2 Einheitsmatrix und

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

ist. Die darstellende Gleichung für L lässt sich auch in der Form

$$L = (2\sinh 2\nu)^{1/2} \exp(\nu^* \tau^1) \tag{8.1}$$

schreiben, wobei

$$\tanh \nu^* = e^{-2\nu},$$

d. h.

$$\sinh 2\nu^* \sinh 2\nu = 1$$

gilt. Dies ergibt sich, weil man durch die Reihendarstellung der Exponentialfunktion und $(\tau^1)^2 = I_2$

$$\begin{aligned}
\exp(\nu^*\tau^1) &= I_2 + (\nu^*\tau^1) + \frac{(\nu^*\tau^1)^2}{2!} \\
&= I_2(1 + \frac{(\nu^*)^2}{2!} + \frac{(\nu^*)^4}{4!} + \ldots) + \tau^1(\nu^* + \frac{(\nu^*)^3}{3!} + \ldots) \\
&= I_2 \cosh(\nu^*) + \tau^1 \sinh \nu^*
\end{aligned}$$

erhält. (Übung)

Um nun die 2-dimensionale Transfermatrixmethode zu vereinfachen, schreiben wir

$$V_1'(S, S') = \exp(\beta \sum_{k=1}^m \sigma_k \sigma_k') = \prod_{k=1}^m \exp(\beta \sigma_k \sigma_k').$$

Diese Matrix ist ein DIREKTES PRODUKT der L-Matrizen aus Gleichung (8.1) (wenn man $\nu = \beta$ setzt), d. h.

$$V_1' = (2\sinh 2\nu)^{m/2} (\exp(\nu^*\tau^1) \otimes \ldots \otimes \exp(\nu^*\tau^1)).$$

Hier sei an einige Fakten über direkte Produkte erinnert. Es seien A, B $n \times n$ Matrizen $A = (A_{i,j})_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}},\ B = (B_{i,j})_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$. Das direkte Produkt $A \times B$ ist eine $n^2 \times n^2$-Matrix mit Matrixeinträgen

$$A \otimes B = ((A \otimes B)_{ii',jj'})_{\substack{i,i'=1,\ldots,n \\ j,j'=1,\ldots,n}},$$

wobei

$$(A \otimes B)_{ii',jj'} = A_{i,j} B_{i'j'}.$$

Ist z. B.

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},$$

so ist

$$A \otimes B = \left(\begin{array}{c|c} A_{11}B & A_{12}B \\ \hline A_{21}B & A_{22}B \end{array}\right) = \begin{pmatrix} A_{11}B_{11} & A_{11}B_{12} & A_{12}B_{11} & A_{12}B_{12} \\ A_{11}B_{21} & A_{11}B_{22} & A_{12}B_{21} & A_{12}B_{22} \\ A_{21}B_{11} & A_{21}B_{12} & A_{22}B_{11} & A_{22}B_{12} \\ A_{21}B_{21} & A_{21}B_{22} & A_{22}B_{21} & A_{22}B_{22} \end{pmatrix}.$$

Einige Eigenschaften des direkten Produkts sind:

1. Es gilt die folgende Regel für die gewöhnliche Matrix-Multiplikation:

$$(A \otimes B) \cdot (C \otimes D) = A \cdot C \otimes B \cdot D.$$

   Tatsächlich ist ja das (ii', jj')-Element der linken Seite

$$\sum_{k,k'} (A \otimes B)_{ii',kk'} (C \otimes D)_{kk',jj'} = \sum_k A_{i_k} C_{k_j} \sum_{k'} B_{i',k'} D_{k',j'} = (A \cdot C)_{ij} (BD)_{i',j'}.$$

2. Wird die Matrix A durch S diagonalisiert (d. h. $S^{-1}AS$ hat Diagonalgestalt) und die Matrix B durch T, dann wird $A \otimes B$ durch $S \otimes T$ diagonalisiert. In der Tat bekommt man dies aus der ersten Eigenschaft, denn

$$\begin{aligned} (S \otimes T)^{-1} &= S^{-1} \otimes T^{-1}, \quad \text{denn} \\ (S \otimes T)(S^{-1} \otimes T^{-1}) &= S \cdot S^{-1} \otimes T \cdot T^{-1} = Id, \quad \text{und} \\ (S^{-1} \otimes T^{-1})(A \otimes B)(S \otimes T) &= S^{-1}AS \otimes T^{-1}BT. \end{aligned}$$

3. Aus 2. ergibt sich, dass die Eigenwerte von $A \otimes B$ $a_i b_j$, $i, j = 1, \ldots, n$, sein müssen, wobei a_i die Eigenwerte von A und b_j die Eigenwerte von B sind.

4. Aus 3. ergibt sich schließlich, dass

$$Tr(A \otimes B) = (TrA)(TrB)$$

   gilt.

Aus der ersten Eigenschaft erhält man nun, dass sich das V_1' von oben schreiben lässt als

$$V_1' = (2 \sinh 2\nu)^{m/2} = \exp(\nu^* \sum_{k=1}^{m} \tau_k^1), \tag{8.2}$$

wobei wir τ_k^1 als

$$\tau_k^1 = I_2 \otimes I_2 \otimes \ldots \otimes I_2 \otimes \tau^1 \otimes I_2 \otimes \ldots \otimes I_2$$

gewählt haben, und das direkte Produkt m Faktoren hat mit τ_k^1 an der k-ten Stelle.

Man überprüft nun, dass sich unsere Matrix L in der Form

$$L = V_1' V_2$$

schreiben lässt. Hierbei sind

$$V_2(S, S') = \delta_{\sigma_1,\sigma_1'} \delta_{\sigma_2,\sigma_2'} \ldots \delta_{\sigma_m,\sigma_m'} \prod_{k=1}^{m} e^{\nu \sigma_k \sigma_{k+1}}.$$

Wählt man also (8.2) als darstellende Gleichung für V_1', so ist V_2 diagonal und bekommt die Gestalt

$$V_2 = \exp(\nu \sum_{k=1}^{m} \tau_k^3 \tau_{k+1}^3).$$

Hier ist

$$\tau_k^3 = I_2 \otimes I_2 \otimes \ldots \otimes I_2 \otimes \tau^3 \otimes I_2 \ldots \otimes I_2,$$

wobei das Produkt wieder m Faktoren hat, τ^3 an der k'-ten Stelle steht und

$$\tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

ist. Setzen wir dies in gie Gleichung für $Z_{\Lambda,\beta}$ ein, so erhalten wir

$$Z_{\Lambda,\beta} = (2\sinh 2\nu)^{m^2/2} Tr((V_1V_2)^m) = (2\sinh 2\nu)^{m1/2} \sum_{j=1}^{2^m} \Lambda_j^m,$$

wobei die Λ_j die Eigenwerte der Matrix V_1V_2 sind und wir den Faktor $(2\sinh 2\nu)^{m/2}$ aus V_1' ausgeklammert haben, also

$$V_1' = (2\sinh 2\nu)^{m/2} V_1 \quad \text{bzw.} \quad V_1 = \exp(\nu^* \sum_{k=1}^{m} \tau_k^1).$$

Die Matrizen τ_k^1 und τ_k^3 sind aus der Matrizenmechanik (also der Quantenmechanik in Matrixschreibweise) wohlbekannt: Es sind zwei der drei Pauli-Matrizen. Die dritte ist

$$\tau_k^2 = -i\tau_k^3\tau_k^1 = I_2 \otimes \ldots \otimes I_2 \otimes \tau^2 \otimes I_2 \ldots \otimes I_2,$$

wo $i = \sqrt{-1}$ und

$$\tau^2 = -i\tau^3\tau^1 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

ist. Auch außerhalb der Quantenmechanik haben die Pauli-Matrizen ein paar nützliche Eigenschaften, die man schnell verifiziert (Übung):

a) $\tau_k^\alpha \tau_k^\beta = i\tau_k^\gamma$ für jede zyklische Permutation $(\alpha\beta\gamma)$ von (1,2,3).

b) $(\tau_k^\alpha)^2 = I_m$, die 2^m-dimensionale Einheitsmatrix, für $\alpha = 1, 2, 3$ und für $k = 1, \ldots, m$.

c) $\tau_k^\alpha \tau_k^\beta + \tau_k^\beta \tau_k^\alpha = 0$ (genauer $0_{2^m}^n$, die 2^m-dimensionale Einheitsmatrix), für $\alpha = 1, 2, 3,\ \beta = 1, 2, 3,\ \alpha \neq \beta$ und $k = 1, \ldots, m$.

d) Und schließlich $\tau_k^\alpha \tau_\ell^\beta = \tau_\ell^\beta \tau_k^\alpha$ für $k \neq \ell = 1, \ldots, m$ und $\alpha, \beta \in \{1, 2, 3\}$.

Der nächste Schritt bei der Berechnung von $Z_{\Lambda,\beta}$ besteht in der Definition von Matrizen P_k und Q_k, deren Zweck bald deutlich werden wird. Hierbei ist

$$P_k = \tau_1^1\tau_2^1 \ldots \tau_{k-1}^1\tau_k^3 \quad \text{und} \quad Q_1 = \tau_1^1\tau_2^1 \ldots \tau_{k-1}^1\tau_k^2.$$

Aus den obigen Punkten folgt dann

$$P_kQ_k = \tau_k^3\tau_k^2 = -i\tau_k^1.$$

Somit erhalten wir für

$$V_1 = \exp(i\nu^* \sum_{k=1}^{m} P_k Q_\ell).$$

Ähnlich berechnet man

$$P_{k+1}Q_2 = \tau_{k+1}^3 \tau_k^1 \tau_k^2 = i\tau_{k+1}^3 \tau_k^3 \quad \text{für} \quad k = 1, 2, \ldots, m-1.$$

Damit lassen sich die Exponenten in V_2 darstellen, bis auf den Randkern $\tau_1^3 \tau_m^3$, der sich als

$$\tau_1^3 \tau_m^3 = iP_1 Q_m U$$

mit

$$U = \tau_1^1 \tau_2^1 \ldots \tau_m^1$$

schreiben lässt. Somit kann man V_2 in der folgenden Form darstellen:

$$V_2 = \exp(i\nu P_1 Q_m U) \exp(-i\nu \sum_{k=1}^{m-1} P_{k+1} Q_\ell).$$

Obschon der Randterm störend aussieht, hilft er uns in der Folge $Z_{\Lambda,\beta}$ zu berechnen. Aus der Definition von U und den Rechenregeln für die τ_k^α erhält man

$$U^2 = I_{2^m} \quad \text{und} \quad (iP_1 Q_m U)^2 = I^{2^m}.$$

Damit erhalten wir

$$\begin{aligned}
\exp(i\nu P_1 Q_m U) &= I_{2^m} \cosh(\nu) + iP_1 Q_m U \sinh(\nu) \\
&= [\frac{1}{2}(I_{2^m} + U) + \frac{1}{2}(I_{2^m} - U)](\cosh(\nu) + iP_1 Q_m U \sinh(\nu)) \\
&= \frac{1}{2}(I_{2^m} + U)(\cosh(\nu) + iP_1 Q_m \sinh(\nu)) \\
&+ \frac{1}{2}(I_{2^m} - U)(\cosh(\nu) - iP_1 Q_m \sinh(\nu)) \\
&= \frac{1}{2}(I_{2^m} + U) \exp(i\nu P_1 Q_m) + \frac{1}{2}(I_{2^m} - U) \exp(-i\nu P_1 Q_m).
\end{aligned}$$

Somit können wir

$$V := V_2 V_1$$

schreiben als

$$V = \frac{1}{2}(I_{2^m} + U)V_+ + \frac{1}{2}(I_{2^m} - U)V_-,$$

wobei

$$V_\pm = \exp(-i\nu \sum_{k=1}^{m} P_{k+1} Q_k) \exp(i\nu^* \sum_{k=1}^{m} P_k Q_k)$$

mit der Übereinkunft, dass

$$P_{m+1} = \mp P_1 \quad \text{für} \quad V_\pm$$

gilt. Man mag sich bis hierher fragen, war wir gewonnen haben und wieso wir einen recht komplizierten Ausdruck für V durch einen anderen, anscheinend nicht weniger komplizierten ersetzen wollen. Dies hat teilweise historische Gründe. Wir folgen von nun an einer

Lösung von Thompson. Der Trick dabei ist, die Operatoren P_k und Q_k durch Operatoren a_k und deren Hermitesch adjugierte a_k^* zu ersetzen.

Wir setzen

$$a_k + a_k^* = \tau_1^1 \tau_2^1 \ldots + \tau_{k-1}^1 \tau_k^2 = Q_k$$

und

$$a_k - a_k^* = i\tau_1^1 \tau_2^1 \ldots \tau_{k-1}^1 \tau_k^3 = iP_k.$$

Aus den Eigenschaften der τ's erhalten wir

$$a_k a_{k'} + a_{k'}^* a_k = I_{2^m} \delta_{k,k'} \quad \text{und} \quad a_k a_{k'} + a_{k'} a_k = 0_{2^m}.$$

(Dies bedeutet in der Sprache der Physik, dass a_k^* als Erzeugungs- und a_k als Vernichtungsoperator aufgefasst werden können.) Wir können nun $V_\pm$ und U als Funktion von a_k und $a_{k'}$ ausdrücken:

$$V_\pm = \exp[\nu \sum_{k=1}^{m} (a_{k+1}^* - a_{k+1})(a_k^* + a_k)] \tag{8.3}$$

und

$$\begin{aligned} U &= \prod_{k=1}^{m} \tau_k^1 = (-i)^m \prod_{k=1}^{m} \exp\left(\frac{\pi P_k Q_k}{2}\right) \\ &= (-i)^m \exp(i\pi \sum_{k=1}^{m} (a_k^* a_k - \frac{1}{2})), \end{aligned}$$

wobei wir sowohl die Darstellung der τ's in Termen von P_k und Q_k benutzt haben, als auch die vorher vorgestellte Rotation für die a_k. Wegen $P_{m+1} = \mp P_1$ erhalten wir für die Darstellung von V_+

$$a_{m+1} = -a_1$$

und für V_-

$$a_{m+1} = a_1.$$

Wir werden nun einen anderen Weg als in einer Dimension 1 beschreiten und die Eigenwerte nicht zur Berechnung der Spur heranziehen. Hierzu bemerken wir, dass

$$(\frac{1}{2} I_{2^m} \pm U)$$

wechselseitig orthogonale Projektionen sind, die mit V_+ bzw. V_- vertauschen. Daher gilt

$$V^m = \frac{1}{2}(I_{2^m} + U)V_+^m + \frac{1}{2}(I_{2^m} - U)(V_-^m.$$

Die Zustandssumme ist dann durch

$$Z_m = (2\sinh 2\nu)^{m^2/2} Tr(V^m)$$

gegeben und zerfällt in 4 Teile. Betrachten wir zunächst $Tr(U_+^m)$. Mittels der diskreten Fourier-Transformation erhalten wir mit

$$\eta_q := \frac{1}{\sqrt{m}} e^{i\pi/4} \sum_k e^{-iqk} a_k,$$

dass

$$a_k = \frac{1}{\sqrt{m}} e^{i\pi/4} \sum_q e^{iq^k} \eta_q \tag{8.4}$$

folgt. Da $a_{m+1} = -a_1$ gelten soll, folgt, dass

$$q = \pm(2j-1)\pi/m$$

gilt (und wir haben aus Einfachhheitsgründen qm als gerade angenommen). Den Faktor $e^{i\pi/4}$ haven wir dabei allein aufgenommen, um schließlich reelle Koeffizienten zu erhalten. Man überprüft nun, dass auch die η_q die Term-Operator-Beziehungen erfüllen, d. h. es gilt

$$\eta_q \eta_q^* + \eta_{q'}^*, \eta_q = Id\, \delta_{q,q'}, \quad \eta_q \eta_{q'} + \eta_{q'} \eta_q = 0.$$

Setzt man (8.4) in (8.3) ein, so erhält man

$$V_\tau = \prod_{0<q<\pi} V_q.$$

Hierbei ist

$$V_q = \exp[2\nu(\cos \Sigma_q^3 - \sin q \Sigma_q^1] \exp(-2\nu^* \Sigma_q^3)$$

und

$$\begin{aligned} \Sigma_q^1 &:= \eta_{-q}\eta_q + \eta_q^* \eta_{-q}^* \\ \Sigma_q^2 &:= i(\eta_{-q}\eta_q - \eta_q^* \eta_{-q}^*) \\ \Sigma_q^3 &:= \eta_q^* \eta_q + \eta_{-q}^* \eta_{-q} - I_{2^m}. \end{aligned}$$

Schließlich läuft das Produkt in der Gleichung für V_+ über

$$q = \frac{\pi}{m}, \frac{3\pi}{m}, \frac{5\pi}{m}, \ldots, \frac{(m-1)\pi}{m}.$$

Aus den Fermischen Antikommutator-Beziehungen für η_q ergibt sich für die Σ_q^α:

$$\Sigma_q^\alpha \Sigma_{q'}^\beta - \Sigma_{q'}^\beta \Sigma_q^\alpha = 2i\delta_{q,q'} \Sigma_q^\gamma, \tag{8.5}$$

wobei $(\alpha\beta\gamma)$ eine zyklische Permutation der Indizes (123) ist. Weiter gilt (aus den gleichen Gründen)

$$\Sigma_q^\alpha \Sigma_q^\beta + \Sigma_q^\beta \Sigma_q^\alpha = 0_{2^m}$$

für $\alpha \neq \beta = 1, 2, 3$.

Aus den Antikommutator-Beziehungen lässt sich ableiten, dass $\eta_q^* \eta_q$ und $\eta_{-q}^* \eta_{-q}$ die Eigenwerte 0 und 1 besitzen und beide mit derselben Vielfachheit (Übung).

Aus der definierenden Gleichung für Σ_q^3 liest man daher ab, dass Σ_q^3 die Eigenwerte 1, 0 und -1 besitzt, alle mit derselben Vielfachheit. Aus Gleichung (8.5) erhalten wir daher, dass Σ_q^3 in der folgenden Darstellung diagonal ist

$$\sum_q^\alpha = I_4 \otimes I_4 \otimes \ldots \otimes I_4 \otimes \sigma^\alpha r^+ \otimes I_4 \otimes \ldots \otimes I_4.$$

Hierbei erstreckt sich das direkte Produkt über $m/2$ Terme und der Term $\sigma^\alpha r^+$ ist an der j-ten Stelle (wobei $q = (2j-1)\pi/m$ ist). Weiter ist I_4 die 4×4 Einheitsmatrix und σ^α und r^+ sind definiert durch

$$\sigma^\alpha = \begin{pmatrix} \tau^\alpha & 0_2 \\ 0_2 & \tau^\alpha \end{pmatrix}, \qquad r^+ = \begin{pmatrix} I_2 & 0_2 \\ 0_2 & 0_2 \end{pmatrix}.$$

I_2 und 0_2 sind die 2×2 Einheits- bzw. Nullmatrix und schließlich sind die $\tau^\alpha,\ \alpha = 1,2,3$ die Pauli-Matrizen

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Nutzt man die Multiplikationseigenschaften des direkten Produkts, erhält man aus der Gleichung für V_+ und den Gleichungen für V_q und Σ_q^α

$$V_+^m = \bigotimes_{j=1}^{m/2} U_{2j-1}^m.$$

Hierbei ist

$$V_\ell = \exp(2\nu[\cos(\frac{\ell\pi}{m})\sigma_+^3 \sin(\frac{\ell\pi}{m})\sigma_+^1]) \times \exp(-2\nu^*\sigma_+^3) \tag{8.6}$$

und

$$\sigma_+^\alpha = \sigma^\alpha r^+.$$

Verwendet man nun, dass sich die Spur eines direkten Produkts von Operatoren als Produkt der Spuren berechnen lässt, erhält man

$$Tr(V_+^m) = \prod_{j=1}^{m/2} Tr(V_{2j-1}^m).$$

Somit haben wir unser Problem heruntergekocht auf das Auffinden der Spur von 4×4 Matrizen. Um diese zu betrachten, stellen wir fest, dass die σ_+^α den Vertauschungsbeziehungen

$$\sigma_+^\alpha\sigma_+^\beta - \sigma_+^\beta\sigma_+^\alpha = 2i\sigma_+^\gamma$$

für $(\alpha\beta\gamma)$ eine zyklische Permutation von (123) genügen. Nun lässt sich gemäß der Baker-Hausdorff-Entwicklung das Produkt zweier Operatoren $\exp(A)\exp(B)$ in der Form $\exp(C)$ schreiben, wobei sich C als Summe von Kommutatoren von A und B schreiben lässt. Wendet man dies auf (8.6) an und zieht die vorhergehenden Vertauschungsrelationen in Betracht, so erhält man

$$V_\ell = \exp(\sum_{\alpha=1}^{3} C_\ell^\alpha \sigma_+^\alpha),$$

wobei die C_ℓ^α zu bestimmen sind. Bedenkt man ferner, dass

$$\sigma_+^\alpha\sigma_+^\beta + \sigma_+^\beta\sigma_+^\alpha = 2\delta_{\alpha,\beta}r^+ \ \alpha,\beta = 1,2,3,$$

gilt (Übung), so kann man durch die übliche Entwicklung der exp-Funktion die Beziehung

$$V_\ell = r^- + r^+\cosh(\gamma_\ell) + (\sum_{\alpha=1}^{3} C_\ell^\alpha \sigma_+^\alpha)\frac{\sinh\gamma_\ell}{\gamma_\ell} \tag{8.7}$$

mit

$$r^- = I_4 - r^+ \quad \text{und} \quad \gamma_c^2 = \sum_{\alpha=1}^{3} (C_\ell^\alpha)^2$$

gewinnen. Umgekehrt mann man auch in Gleichung (8.6) die exp-Funktionen entwickeln und

$$\begin{aligned} V_\ell &= r^- + r^+[\cosh 2\nu \cosh 2\nu^* - \sinh 2\nu \sinh 2\nu^* \cos(\frac{\ell\pi}{m})] - \sigma_1^+[\sinh 2\nu \cosh 2\nu^* \sin(\frac{\ell\pi}{m})] \\ &+\sigma_+^2[\sinh 2\nu^* \sinh 2\nu \sin(\frac{\ell\pi}{m})] + \sigma_+^3[\sinh 2\nu \cosh 2\nu^* \cos(\frac{\ell\pi}{m}) - \sinh 2\nu^* \cosh 2\nu] \end{aligned}$$

erhalten. Dieser Ausdruck muss natürlich derselbe sein wie (8.7), von wo man für die Koeffizienten von r^+ vermöge der Definition von ν^* bekommt:

$$\cosh \gamma_\ell = \cosh 2\nu \cosh 2\nu^* - \sinh 2\nu \sinh 2\nu^* \cos(\frac{\ell\pi}{m}) = \cosh 2\nu \cosh 2\nu^* - \cos(\frac{\ell\pi}{m}).$$

Lustigerweise genügt dies, um die Spur von V_+^m zu berechnen. Bedenkt man, dass aus den obigen Überlegungen

$$V_\ell^m = \exp(\sum_{\alpha=1}^{3} (nc_\ell^\alpha)\sigma_+^\alpha) = r^- + r^+ \cosh(n\gamma_\ell) + (\sum_{\alpha=1}^{3} C_\ell^* \sigma_+^\alpha) \frac{\sinh n\gamma_\ell}{\gamma_\ell}$$

folgt, und erinnert man sich der Definition für r^+ und r^- (und bedenkt, dass $Tr(\sigma_+^\alpha) = 0$, was auch eine kleine Übung ist), so erhält man

$$Tr(V_\ell^m) = 2[1 + \cosh(n\gamma_\ell)] = 4\cosh^2(\frac{n\gamma_\ell}{2}).$$

Hieraus lässt sich der erste Term für die Spur von V_+ berechnen. Es folgt nämlich

$$Tr(V_+^m) = \prod_{j=1}^{m/2} 4\cosh^2(\frac{n\gamma_{2j-1}}{2}).$$

Um den Spurterm für UV_+^n zu berechnen, verwendet man wieder die Diagonaldarstellung der Σ_q^α und schreibt U in der Form

$$U = \bigotimes_{j=1}^{m/2} (-e^{i\pi\sigma_+^3}) = \bigotimes_{j=1}^{m/2} (r^+ - r^-).$$

Hieraus erhält man

$$\begin{aligned} Tr(UV_+^m) &= \prod_{j=1}^{m/2} Tr[(r^+ - r^-)V_{2j-1}^m] \\ &= \prod_{j=1}^{m/2} 2(-1 + \cosh(n\gamma_{2j-1})) \\ &= \prod_{j=1}^{m/2} 4\sinh^2(\frac{n\gamma_{2j-1}}{2}). \end{aligned}$$

Um die Terme $Tr(V_-^m)$ und $Tr(UV_-^n)$ zu berechnen, ersetzt man $\sigma^\alpha r^+$ durch $\sigma^\alpha r^-$ und die Berechnungen gehen im wesentlichen mit einem Minuszeichen anstelle des Pluszeichens durch.

Wir setzen nun γ_k als die positive Lösung von

$$\cosh\gamma_k = \cosh 2\nu \cosh 2\nu^* - \cos(\frac{k\pi}{m})$$

fest und bemerken, dass

$$\gamma_{2m-k} = \gamma_k$$

für $k = 0, 1, \ldots, m$ gilt. Dann erhalten wir als Kombination der vorherigen Resultate

$$\begin{aligned} Z_m &= (2\sinh 2\nu)^{m^2/2} Tr(V^m) \\ &= \frac{1}{2}(2\sinh 2\nu)^{m^2/2} [\prod_{j=1}^m 2\cosh\left(\frac{m\gamma_{2j-1}}{2}\right) \\ &\quad + \prod_{j=1}^m 2\sinh\left(\frac{m\gamma_{2j-1}}{2}\right) + \prod_{j=1}^m 2\cosh\left(\frac{m\gamma_{2j}}{2}\right) \\ &\quad + \prod_{j=1}^m 2\sinh\left(\frac{m\gamma_{2j}}{2}\right)]. \end{aligned}$$

Dies ist der exakte Wert der Zustandssumme für das 2-dimensionale Ising-Modell. Man kann nun noch mit

$$\gamma_0 = 2(\nu - \nu^*), \qquad \gamma_{2r} \sim \gamma_{2r-1}\gamma_0$$

berechnen und kommt asymptotisch auf

$$Z_m \sim (2\sinh 2\nu)^{m^2/2} \exp(\frac{m}{2}\sum_{j=1}^m \gamma_{2j-1}) \times [1 + \prod_{j=1}^m \tanh\left(\frac{m\gamma_{2j-1}}{2}\right)]$$

(wobei man $2\cosh(x) \sim e^x$ für $x \to \infty$ verwendet). Dies ergibt

$$\lim_{m\to\infty} \frac{1}{m^2} \log Z_m = \frac{1}{2}\log(2\sinh 2\nu) + \lim_{m\to\infty} \frac{1}{2^m} \sum_{j=1}^m \gamma_{2j-1}.$$

Der erwähnte Phasenübergang im $2d$-Ising-Modell ergibt sich nun aus der Tatsache, dass, während alle γ_r für $r \geq 2$ mehr oder weniger symmetrisch in $\nu = \nu_c$ sind (wobei wir ν_c via $\sinh 2\nu_c = 1$ definieren),

$$\gamma_1 \sim \gamma_0 \quad \text{für} \quad \nu > \nu_c$$

und

$$\gamma_1 \sim -\gamma_0 \quad \text{für} \quad \gamma < \nu_c$$

gilt.

9 Vom Mittelwertmodell zum Ising-Modell: Das Kac-Ising-Modell

Wir haben in den vorhergehenden Kapiteln Phasenübergänge im Curie-Weiss-Modell und dem Ising-Modell untersucht. Hierbei ist das Curie-Weiss-Modell ein sogenanntes Mittelwertmodell, weil die Interaktion von σ_i mit den anderen Spins σ_j durch die durchschnittliche Interaktion mit einem "großen" Spin ersetzt wurde. Es stellt sich die Frage, was die Phasenübergänge in den beiden genannten Modellen miteinander zu tun haben. Hierzu betrachten wir ein Modell, das zwischen beiden interpoliert, das sogenannte Kac-Ising-Modell. Hierfür sei $\gamma > 0$ ein Parameter, den wir schließlich gegen 0 schicken wollen. Für $\Lambda \subseteq \mathbb{Z}^d$ endlich definieren wir die Hamiltonfunktion:

$$H_{\Lambda,\gamma}(\sigma) = -\gamma_2^d c_d \sum_{\substack{i,j \\ |i-j| \leq \gamma^{-d}}} \sigma_i \sigma_j - \sum_{\substack{i \in \Lambda \\ j \notin \Lambda \\ |i-j| \leq \gamma^{-d}}} \sigma_i \sigma_j \gamma^d c_d.$$

Hierbei ist $|\cdot|$ die sup-norm und c_d derart, dass

$$S \mathbb{1}_{|X| \leq 1} dx|(x) = \frac{1}{c_d}.$$

Dieser Hamiltonian stellt eine Art Übergang zwischen dem Ising-Hamiltonian und dem Curie-Weiss-Hamiltonian dar, da er bei festgehaltenem Λ für $\gamma = 1$ gleich der Energiefunktion im Ising-Modell ist, während er für $\gamma \to 0$ schließlich dem Curie-Weiss-Hamiltonian gleicht. Die wesentliche Schwierigkeit besteht nun darin, dass wir die Limiten in einer anderen Reihenfolge ausführen wollen, d. h. wir betrachten zuerst den thermodynamischen Limes $\Lambda \to \infty$ des Modells und nehmen dann den "meanfield"-Limes $\gamma \to 0$.

Wir definieren nun die lokalen Spezifikationen, die zu der Hamiltonfunktion $H_{\Lambda,\gamma}$ gehören:

$$\mu^\eta_{\Lambda,\beta,\gamma}(\sigma_\Lambda) = \frac{1}{Z^\eta_{\Lambda,\beta,\gamma}} e^{-\beta H_{\Lambda,\gamma}(\sigma_\lambda, \eta_{\Lambda^c})},$$

wobei

$$Z^\eta_{\Lambda,\beta,\gamma} = \sum_{\sigma'_\Lambda} e^{-\beta H_{\Lambda,\gamma}(\sigma_\Lambda, \eta_{\Lambda^c})}$$

die zugehörige Zustandssumme ist.

Wie üblich wird ein Gibbs-Maß im unendlichen Volumen ein Maß $\mu_{\beta,\gamma}$ sein, dass den DLR-Gleichungen

$$\mu_{\beta,\gamma} \mu^\bullet_{\Lambda,\beta,\gamma} = \mu_{\beta,\gamma}$$

genügt. Das zentrale Resultat dieses Abschnitts entstammt zwei Arbeiten von Cassandro und Presutti und Bovier und Zahradnik aus dem Jahr 1997 und besagt:

Satz 9.1 *Sei $d \geq 2$. Dann gibt es eine Funktion $f(\gamma)$ mit $\lim_{\gamma \downarrow 0} f(\gamma) = 0$, so dass für alle*

$$\beta > 1 + f(\gamma)$$

mindestens zwei unterschiedliche Gibbs-Maße im unendlichen Volumen existieren.

Bemerkung: Cassandro, Morra und Presutti zeigten schon 1995, dass es für $\beta < 1$ im Kac-Ising-Modell nur ein Gibbs-Maß im unendlichen Volumen gibt. Zusammen mit obigem Satz sehen wir also, dass

$$\lim_{\gamma \downarrow 0} \beta_{c,\gamma} = 1$$

gilt. Die kritische Temperatur im Kac-Ising-Modell konvergiert also gegen die kritische Temperatur im Curie-Weiss-Modell, obschon wir in beiden Modellen unterschiedliche Begriffen für "Phasenübergang" verwendet haben – ein sehr befriedigendes Resultat. Der Beweis beruht auf Konturargumenten, die wir ähnlich (aber weniger kompliziert) schon in den Kapiteln 5 und 7 kennengelernt haben. Dort haben wir schon gesehen, dass die Magnetisierung ein Ordnungsparameter des Systems war. Wir wollen zunächst eine lokalere Version dieser Größe einführen.

Sei ℓ eine Längenskala mit $1 \ll \ell \ll \frac{1}{\gamma}$. Wir unterteilen das Gitter $\mathbb{Z}^d$ in Blöcke der Kantenlänge ℓ. Wenn wir die Marke $X \in \mathbb{Z}^d$ eines Blocks mit dem Block identifizieren, so können wir schreiben

$$x := \{i \in \mathbb{Z}^d : |i - \ell x| \leq \ell/2\}.$$

Wir setzen dann als Magnetisierung auf x:

$$m_x(\sigma) := \frac{1}{\ell^d} \sum_{i \in x} \sigma_i.$$

Wir werden in der Folge annehmen, dass alle endlichen Volumina mit diesen Blöcken kompatibel sind, in dem Sinne, dass sie in diese Blöcke unterteilbar sind. Ebenso nehmen wir an, dass $\gamma \cdot \ell \in \mathbb{N}$ ist. Ferner sei $\mathcal{M}_\Lambda \subseteq \mathcal{F}_\Lambda$ die σ-Algebra, die durch $\{m_x(\sigma)\}_{x \in \Lambda}$ erzeugt wird.

Die Grundidee wird es nun sein, eine Art "Vergröberung" ("coarse graining") des Modells zu betrachten, indem wir versuchen, die Hamiltonfunktion durch eine Hamiltonfunktion in den "Blockspins" $m_x(\sigma)$ zu approximieren. Für diese neue Hamiltonfunktion führen wir dann eine Version des Peierlschen Arguments durch.

Für den ersten Schritt wird es sich als nützlich herausstellen, die interessanten Wahrscheinlichkeiten als Wahrscheinlichkeiten von Ereignissen aus $\mathcal{M}_V$, $V \subseteq \mathbb{Z}^d$ endlich, zu beschreiben. Für ein solches Ereignis $A \in \mathcal{M}_V$ und $\Lambda \supseteq V$ gilt

$$\mu^\eta_{\Lambda,\beta,\gamma}(A) = \sum_{\sigma_{\Lambda\setminus V}} \mu^\eta_{\Lambda,\beta,\gamma}(\sigma_{\Lambda\setminus V}) \mu^{\sigma_{\Lambda\setminus V},\eta_{V^c}}_{V,\beta,\gamma}(A) = \sum_{\sigma_{\Lambda\setminus V}} \mu^\eta_{\Lambda,\beta,\gamma}(\sigma_{\Lambda\setminus V}) \sum_{\substack{m_x, x\in V\\ \{m_x\}\subseteq A}} \mu^{\sigma_{\Lambda\setminus V},\eta_{\Lambda^c}}_{V,\beta,\gamma}(\{m_x\}_{x\in V}),$$

wobei wir im ersten Schritt die Verträglichkeit der Gibbs-Maße genutzt haben, im zweiten Schritt, dass $A \in \mathcal{M}_V$ vorausgesetzt war. Die Summe über die m_x läuft dabei über die Werte aus der Menge $\{-1, -1 + 2\ell^{-d}, \ldots, 1 - 2\ell^{-d}, 1\}$. In unserer Situation können wir dadurch, dass wir Λ hinreichend groß wählen, annehmen, dass $\mu^{\eta_{\Lambda\setminus V},\eta_{\Lambda^c}}_{V,\beta,\gamma}$ nicht von Λ^c abhängt. Wir werden daher den oberen Index η_{Λ^c} auch weglassen.

Wie angekündigt, können wir in unseren Kac-Modellen die Hamiltonfunktion approximativ durch eine Funktion ersetzen, die nur von den Block-spins m_x abhängt:

$$H_{V,\gamma}(\sigma_v, \sigma_{V^c}) = -\frac{1}{2} \sum_{x,y \in V} \sum_{\substack{i \in x\\ j \in y}} J_\gamma(i,j)\sigma_i\sigma_j - \sum_{x \in V, y \in V^c} \sum_{i \in x, j \in y} J_\gamma(i,j)\sigma_i\sigma_j.$$

Hierbei ist

$$J_\gamma(i) := \gamma^d J(\gamma_i)$$

und

$$J(v) = c_d 1\!\!1_{|x|\le 1}$$

gewählt. Somit gilt weiter

$$\begin{aligned}
H_{V,\gamma}(\sigma_v, \sigma_{V^c}) &= -\frac{1}{2}\sum_{\substack{x\in V\\ y\in V}} J_\gamma(\ell(x-y))\sum_{\substack{i\in x\\ j\in y}}\sigma_i\sigma_j - \sum_{\substack{x\in V\\ y\in V^c}} J_\gamma(\ell(x-y))\sum_{\substack{i\in x\\ j\in y}}\sigma_i\sigma_j \\
&\quad -\frac{1}{2}\sum_{x,y\in V}\sum_{\substack{i\in x\\ j\in y}}[J_\gamma(i-j) - J_\gamma(\ell(x-y))]\sigma_i\sigma_j \\
&\quad -\sum_{\substack{x\in V\\ y\in V^c}}\sum_{\substack{i\in x\\ j\in y}}[J_\gamma(i-j) - J_\gamma(\ell(x-y))]\sigma_i\sigma_j \\
&=: H^{(0)}_{V,\gamma,\ell}(m_V(\sigma_V) m_{V^c}(\sigma_{V^c})) + \Delta H_{V,\gamma,\ell}(\sigma_V, \sigma_{V^c}).
\end{aligned}$$

Hierbei haben wir unter Ausnutzung von

$$J_\gamma(\ell x) = \ell^{-d} J_{\ell\gamma}(x)$$

die folgenden Setzungen vorgenommen:

$$H^{(0)}_{V,\gamma,\ell}(m_v, m_{v^c}) := -\ell^d\frac{1}{2}\sum_{x,y\in V} J_{\gamma\ell}(x-y) m_x m_y - \ell^d \sum_{\substack{x\in V\\ y\in V^c}} J_{\gamma\ell}(x-y) m_x m_y$$

und

$$\Delta H_{V,\gamma,\ell}(\sigma_v, \sigma_{V^c}) = -\frac{1}{2}\sum_{x,y\in V}\sum_{\substack{i\in x\\ j\in y}}[J_\gamma(i-j) - J_\gamma(\ell(x-y))]\sigma_i\sigma_j - \sum_{\substack{x\in V\\ y\notin V}}\sum_{\substack{i\in x\\ j\in y}}[J_\gamma(i-j) - J_\gamma(\ell(x-y))]\sigma_i\sigma_j.$$

Wesentlich ist nun, dass der Restterm $\Delta H_{V,\gamma,\ell}$ für $\gamma \to 0$ verschwindet.

Lemma 9.2 *Für jedes $V \subseteq \mathbb{Z}^d$ gilt*

$$\sup_\sigma |\Delta H_{V,\gamma,\ell}(\sigma_v, \sigma_{V^c})| \le \kappa_d \gamma \ell |V|$$

mit einer Konstante κ_d, die nur von der Dimension d abhängt.

Beweis: Die genaue Ausführung des Beweises ist eine Übung. Der Beweis folgt allerdings unmittelbar aus der Tatsache, dass

$$J_\gamma(i-j) - J_\gamma(\ell(x-y))$$

in den obigen Summen nur dann verschieden von Null sein kann, wenn

$$|x-y| \approx \frac{1}{\gamma\ell}$$

gilt (ansonsten sind entweder beide Terme positiv und identisch ober beide Terme gleich 0). $\square$

Eine unmittelbare Folge des vorhergehenden Lemmas ist das folgende

Lemma 9.3 *Sei*

$$E_{V,\beta,\gamma,\ell}(m_v, m_{v^c}) := -\frac{1}{2} \sum_{x,y \in V} J_{\gamma\ell}(x-y) m_x m_y - \sum_{\substack{x \in V \\ y \in V^c}} J_{\gamma\ell}(x-y) m_x m_y + \frac{1}{\beta} \sum_{x \in V} I(m_x),$$

wobei $I(m)$ für $m \in [-1,1]$ die Entropiefunktion

$$I(m) = \frac{m+1}{2} \log(m+1) + \frac{1-m}{2} \log(1-m)$$

bezeichnet. Dann gilt für alle endlichen Volumina V und alle Konfigurationen m_V

$$\mu^{\sigma_{\Lambda \setminus V}}_{V,\beta,\gamma}(m_V) \le \frac{e^{-\beta \ell^d E_{V,\beta,\gamma,\ell}(m_V, m_{V^c}(\sigma_{V^c}))}}{\sum_{m_V} e^{-\beta \ell^d E_{V,\beta,\gamma,\ell}(m_V, m_{V^c}(\sigma_{V^c}))}} \times e^{+\kappa_d \beta \gamma \ell |V|}$$

und

$$\mu^{\sigma_{\Lambda \setminus V}}_{V,\beta,\gamma}(m_V) \ge \frac{e^{-\beta \ell^d E_{V,\beta,\gamma,\ell}(m_V, m_{V^c}(\sigma_{V^c}))}}{\sum_{m_V} e^{-\beta \ell^d E_{V,\beta,\gamma,\ell}(m_V, m_{V^c}(\sigma_{V^c}))}} \times e^{-\kappa_d \beta \gamma \ell |V|}$$

Beweis: Lemma 9.3 impliziert

$$\mu^{\sigma_{\Lambda \setminus V}}_{V,\beta,\gamma}(m_V) \le \frac{e^{-\beta \ell^d H^{(0)}_{V,\gamma,\ell}(m_V, m_{V^c})} \prod_{x \in V} \mathbb{E}_\sigma / \mathbb{1}_{m_x(\sigma) = m_x}}{\sum_{\tilde{m}_V} e^{-\beta \ell^d H^{(0)}_{V,\gamma,\ell}(\tilde{m}_V, \tilde{m}_{V^c}} \prod_{x \in V} \mathbb{E}_\sigma \mathbb{1}_{m_x(\sigma) = m_x}} e^{+\kappa_d \beta \gamma \ell |V|}$$

und

$$\mu^{\sigma_{\Lambda \setminus V}}_{V,\beta,\gamma}(m_V) \ge \frac{e^{-\beta \ell^d H^{(0)}_{V,\gamma,\ell}(m_V, m_{V^c})} \prod_{x \in V} \mathbb{E}_\sigma / \mathbb{1}_{m_x(\sigma) = m_x}}{\sum_{\tilde{m}_V} e^{-\beta \ell^d H^{(0)}_{V,\gamma,\ell}(\tilde{m}_V, \tilde{m}_{V^c}} \prod_{x \in V} \mathbb{E}_\sigma \mathbb{1}_{m_x(\sigma) = m_x}} e^{-\kappa_d \beta \gamma \ell |V|}$$

Nun ist

$$\mathbb{E}_\sigma \mathbb{1}_{m_x(\sigma) = m_x} = \begin{cases} 2^{-\ell^d} \binom{\ell^d}{\frac{m_x+1}{2} \ell^d}, & \text{falls } \frac{m_x+1}{2} \ell \in \mathbb{N} \\ 0 & \text{sonst} \end{cases}.$$

Wendet man die Stirlingsche Formel

$$2^{-\ell^d} \binom{\ell^d}{\frac{m_x+1}{2} \ell^d} \approx e^{-\ell^d I(m_x) + 0(\log \ell)}$$

für $\ell \to \infty$. Dies beweist die Behauptung. □

Bemerkung: Wir werden $\ell \to \infty$ gehen lassen, wenn $\gamma \to 0$ geht. Die Idee dabei ist, dass $E_{V,\beta,\gamma,\ell}$ als eine Art Ratenfunktion fungiert. Dies ergibt noch einige technische Probleme, die wir in der Folge lösen werden(im wesentlichen müssen wir sicherstellen, dass für unsere Ereignisse $E_{V,\beta,\gamma,\ell}$ genügend groß ist, so dass die Fehlerterme nicht dominierend werden).

Bevor wir fortfahren, wollen wir $E_{V,\beta,\gamma,\ell}$ mit Hilfe von

$$-m_x m_y = \frac{1}{2}(m_x - m_y)^2 - \frac{1}{2}(m_x^2 + m_y^2)$$

ein wenig umschreiben. Wir setzen

$$\tilde{E}_V(m_V, m_{V^c}) := \frac{1}{4}\sum_{x,y\in V} J_{\gamma\ell}(x-y)(m_x-m_y)^2 + \frac{1}{2}\sum_{\substack{x\in V\\ y\in V^c}} J_{\gamma\ell}(m_x-m_y)^2 + \sum_{x\in V} f_\beta(m_x),$$

wobei

$$f_\beta(x) := \frac{1}{\beta}I(m_x) - \frac{1}{2}m_x^2$$

die freie Energie des Curie-Weiss-Modells ist. Dann gilt (unter Weglassung der überflüssigen Indizes)

$$E_V(m_V, m_{V^c}) = \tilde{E}_V(m_V, m_{V^c}) - C_V(m_{V^c}).$$

Hierbei hängt

$$C_V(m_{V^c}) := \frac{1}{2}\sum_{\substack{x\in V\\ y\notin V}} J_{\gamma\ell}(x-y)m_y^2$$

nur von den Variablen in V^c ab. Der Vorteil von $\tilde{E}$ ist, dass man die bevorzugten Konfigurationen schon ablesen kann: Es sind solche, die annähernd konstant sind und zwar konstant gleich dem Minimum der "Curie-Weiss-Funktion" $f_\beta(m)$. Wir werden nun versuchen, eine Version des Peierlschen Arguments in diesem Modell anzuwenden. Die Hauptschwierigkeit liegt dabei in zwei Punkten begründet: Zum einen nehmen die Blockspins mehr als nur zwei Werte an, zum anderen wechselwirken die Spins m_x immer noch über lange Distanzen (das ist auch notwendig, weil anderenfalls der Approximationsfehler $|\Delta H(\sigma)|$ nicht gegen0 konvergierte). Beides führt dazu, dass man Konturen nicht einfach naiv wie im Peierlschen Argument definieren kann. Stattdessen werden wir Konturen mit Hilfe "lokaler MIttelwerte", $\Phi_x(m)$ und "lokaler Varianzen", $\psi_x(m)$, definieren. Diese sind so definiert:

$$\phi_x(m) := \sum_J J_{\gamma\ell}(x-y)m_y \quad \text{und} \quad \psi_x(m) := \sum_J J_{\gamma\ell}(x-y)(m_y-\phi_y(m))^2.$$

Mit diesen Größen sei

$$\tilde{\Gamma} := \{x : |\phi_x(m) - m^*(\beta)| > \xi m^*(\beta) \quad \text{oder} \quad \psi_x(m) > (\xi m^*(\beta))^2\}.$$

Hierbei haben wir die Größe $m^*(\beta)$ schon im Curie-Weiss-Modell kennengelernt. Es ist die größte Lösung der Gleichung

$$x = \tanh(\beta x)$$

und somit die Stelle des nicht-negativen Minimums der freien Energie f_β. Wir erinnern

$$m^*(\beta)\begin{cases} = 0 & \text{für } \beta \le 1 \\ > 0 & \text{für } \beta > 1 \end{cases}.$$

Außerdem lässt sich zeiten

Übung: $\lim_{\beta\uparrow\infty} m^*(\beta) = 1$ und $\lim_{\beta\downarrow 1} \frac{(m^*(\beta))^2}{3(\beta-1)} = 1$.

Wir werden zunächst m^* für $m^*(\beta)$ schreiben. Die Konstante $\xi < 1$ werden wir geeignet wählen. Die Grundidee bei der Konstruktion von $\tilde{\Gamma}$ wird etwas klarer, wenn man sich – ähnlich zum Ising-Modell – vorstellt, man habe Randbedingungen mit $\phi_x(m(\eta)) \approx m^*$ und Spins nahe dem Ursprung mit $\phi_0(m(\sigma)) < 0$ gegeben. Dann muss es ein Gebiet geben, für das ϕ einen Wert nahe 0 annimmt. Dieses Gebiet gehörte dann zu $\tilde{\Gamma}$. Für spätere Zwecke werden wir diese Menge zunächst regularisieren. Hierzu teilen wir das Gitter $\mathbb{Z}^d$ erneut in Blöcke, diesmal in solche, die mit der Reichweite der Wechselwirkung vergleichbar sind. Wieder identifizieren wir die Punkte dieses gröberen Gitters mit den Boxen

$$U := \{x \in \mathbb{Z}^d,\, |x - \frac{u}{\gamma\ell}| \leq \frac{1}{2\gamma\ell}\}.$$

Wir schreiben $u(x)$ für die Marke des eindeutigen Blocks, der x enthält. Wir nennen Mengen, die Vereinigungen solcher Blöcke n sind, reguläre Mengen. Wir setzen

$$\underline{\Gamma}_0 := \{x : u(x) \cap \tilde{\underline{\Gamma}} \neq \emptyset\}.$$

Für ein $n \geq 1$ ($n \in \mathbb{N}$) (das wir später wählen werden), definieren wir

$$\underline{\Gamma} := \{x : \mathrm{dist}(x, \underline{\Gamma}_0) \leq \frac{n}{\gamma\ell}\},$$

wobei $\mathrm{dist}(\cdot,\cdot)$ die Distanz ist, die zur sup-Norm gehört. Wir werden n so wählen, dass n von β abhängt und

$$n \to \infty \qquad \text{wenn} \qquad \beta \to 1$$

gilt. $\underline{\Gamma}$ ist gerade so konstruiert, dass es eine reguläre Menge ist. Zusammenhangskomponenten von $\underline{\Gamma}$ zusammen mit der Angabe der Werte $m_x,\ x \in \underline{\Gamma}$ heißen Konturen. Sie werden mit Γ bezeichnet. Den Rand einer solchen Kontur definieren wir als

$$\partial\Gamma := \{x \in \underline{\Gamma} : \mathrm{dist}(x, \underline{\Gamma}^c) \leq \frac{n}{\gamma\ell}\}.$$

Definitionsgemäß gilt

$$\partial\Gamma \cap \underline{\Gamma}_0 = \emptyset.$$

Wir bezeichnen mit $D^\pm$

$$D^\pm := \{x : |\phi_x \mp m^*| \leq \xi m^*\} \cap \underline{\Gamma}^c.$$

Diese nennen wir $\pm$-korrekt. Jede Zusammenhangskomponente von $\partial\Gamma$ hat entweder eine Verbindung zu D^+ oder zu D^-. Wir nennen diese Zusammenhangskomponenten $\partial\Gamma^+$ bzw. $\partial\Gamma^-$.

Für eine zusammenhängende Menge $\underline{\Gamma}$ bezeichnen wir mit int $\underline{\Gamma}$ die einfache zusammenhängende Menge, die man erhält, wenn man die "Lücke" in $\underline{\Gamma}$ füllt. int $\underline{\Gamma}$ hießt das Innere einer Kontur. Den Rand von int $\underline{\Gamma}$ nennen wir auch den *äußeren Rand von* $\underline{\Gamma}$. Die Zusammenhangskomponente von $\partial\Gamma$, die auch Rand von int $\underline{\Gamma}$ ist, heißt auch *äußerer Rand von* Γ. Wir schreiben hierfür auch $\partial\Gamma^{\mathrm{ext}}$. Die Gesamtidee des Beweises verläuft mit diesen Definitionen ähnlich zum Beweis des Peierlschen Arguments im Ising-Modell: Mit Randbedingungen, die $\phi_x(\eta) \approx m^*$ genügen, benötigt das Ereignis

$$\{|\phi_0(x) - m^*| > \xi m^*\}$$

die Existenz einer Kontur $\underline{\Gamma}$ mit $0 \in \operatorname{int} \underline{\Gamma}$. Also müssen wir "nur" zeigen, dass solche Konturen geringe Wahrscheinlichkeit besitzen. Da diese Wahrscheinlichkeit im wesentlichen durch die Energie der zugrundeliegenden Konturen bestimmt ist, benötigen wir eine untere Schranke für die Energie der Konfigurationen, die mit $\underline{\Gamma}$ kompatibel sind und eine obere Schranke für die Energie einer geeigneten "Vergleichskonfiguration", die die Kontur $\underline{\Gamma}$ nicht besitzt. Die Idee bei der Einführung der Größen ϕ und ψ war es, dass die untere Schranke sich in diesen Größen ausdrücken lässt. Nun liegt die Schwierigkeit in der Tatsache, dass die Variablen m_x (wie schon erwähnt) über lange Reichweiten wechselwirken und "beide" stetig sind. Wir müssen daher "Sicherheitszonen" konstruieren, die sicherstellen, dass das Innere einer Kontur und ihr Äußeres genügend unkorreliert sind.

Wir beginnen nun damit, das "genügend unkorreliert" zu beweisen. Hierzu benötigen wir ein paar Eigenschaften der Konfiguration m auf $\partial\Gamma$, die $\mathbb{E}_{\partial\Gamma}$ für gegebene Randbedingungen minimiert.

Definition 9.4 *Eine Konfiguration m_V^{opt} heißt optimal, falls sie*

$$m_V \mapsto E(m_V, m_{V^c})$$

für gegebenes m_{V^c} minimiert.

Ein wesentlicher Punkt ist nun, dass Konfigurationen weit weg von $\underline{\tilde{\Gamma}}$ "beinahe konstant" sein müssen.

Lemma 9.5 *Sei*

$$dist(x, \underline{\tilde{\Gamma}}) > \frac{1}{\gamma\ell}.$$

Dann gilt

a) $\sum_J J_{\gamma\ell}(x-y)(m_y \pm m^*)^2 < 4\xi(m^*)^2$

und

b) für jedes $V \subseteq \underline{\tilde{\Gamma}}$ gilt

$$\sum_{J\in V} J_{\gamma\ell}(x-y)|m_y \pm m^*| \leq 2\xi m^* \sqrt{\sum_{J\in V} J_{\gamma\ell}(x-y)}.$$

Hierbei hängt das Vorzeichen von m^* davon ab, ob $\phi_x(m)$ in dem Gebiet positiv oder negativ ist.

Beweis: a) folgt, da wir für x mit $\operatorname{dist}(x, \underline{\tilde{\Gamma}}) > \frac{1}{\gamma\ell}$ wissen, dass

$$|\sum_J J_{\gamma\ell}(x-y)m_y - m^*| \leq 2\xi m^*$$

gilt.

b) folgt aus a) mit Hilfe von Cauchy-Schwartz. $\square$

Lemma 9.6 *Sei V eine reguläre Menge. Dann gibt es ein $\xi_d > 0$, das nur von der Dimension d abhängt, mit der folgenden Eigenschaft: Ist m_{V^c} eine Randbedingung mit überwiegend $\oplus$-Spins, für die Aussagen des vorhergehenden Lemmas gelten (mit $\xi \leq \xi_d$), dann ist für alle $x \in V$*

$$|m_x^{opt} - m^*| \leq \frac{m^*}{2}.$$

Analoges gilt für überwiegend $\ominus$-Randbedingungen.

Beweis: Wir behandeln die Variablen m_y als kontinuierliche Variablen. Dies ist einfacher und haben wir eine Lösung in $\mathbb{R}$ gefunden, so gibt es eine Approximation in der Menge der zulässigen Werte für m_y, so dass sich die betreffenden Energien um höchstens $\frac{|\Gamma|}{\ell^d} \to 0$ unterscheiden. Betrachtet man die Definition E, so ergibt sich aus der Minimalität von m_V

$$\frac{d}{dm_y}\mathbb{E}(m_V, m_{V^c}) = 0$$

für alle $y \in V$. Das bedeutet

$$\phi_y(m) = \frac{1}{\beta} I'(m_y)$$

(dabei verschwindet der Faktor $\frac{1}{2}$ in der Definition von E, da alle Konfigurationen $x \in V$, $y \in V$ doppelt in der Summe auftauchen). Nach Definition von I ist dies gleichbedeutend mit

$$m_y = \tanh(\beta\phi_y(m)).$$

Wir nehmen nun stillschweigend an, dass $\phi_y(m) > 0$ gilt. m^* ein ein (attraktiver) Fixpunkt von

$$m \mapsto \tanh \beta_m,$$

der die Punkte mit $m > 0$ unter Iteration anzieht. Somit gilt

$$|\tanh(\beta\phi_y(m)) - m^*| \leq |\phi_y(m) - m^*|$$

und insbesondere

$$\begin{aligned} \tanh(\beta\phi_y(m)) &> \phi_y(m), \text{ falls } \phi_y(m) < m^* \quad \text{und} \\ \tanh(\beta\phi_y(m)) &< \phi_y(m), \text{ falls } \phi_y(m) > m^*. \end{aligned}$$

Zunächst zeigen wir

$$m_x^{\text{opt}} \geq \frac{m^*}{2}.$$

Sei $x \in V$ ein Punkt mit

$$m_x = \inf_{J \in V}\{m_y : m_Y \leq m^*\}.$$

Ist $m_x = m^*$, ist nichts zu zeigen. Umgekehrt kann aber $m_x < m^*$ nur für x nahe am Rand, also mit $\text{dist}(x, \partial V) < \frac{1}{\gamma\ell}$ gelten (für x, die von den Randspins nicht beeinflusst wrden, kann man die Energie nur kleiner machen, indem man alle Spins auf m^* setzt). Für x mit $\operatorname{dist}(x, \partial V) < \frac{1}{\gamma\ell}$ kann man schreiben:

$$\begin{aligned} m_x - m^* &\geq \sum_{J \in V} J_{\gamma\ell}(x-y)(m_y - m^*) + \sum_{J \in V^c} J_{\gamma\ell}(x-y)(m_y - m^*) \qquad (9.1) \\ &\geq (m_x - m^*) \sum_{J \in V} J_{\gamma\ell}(x-y) - 2\xi m^* \sqrt{\sum_{J \in V^c} J_{\gamma\ell}(x-y)}, \end{aligned}$$

wobei wir in der zweiten Ungleichung die Minimalität von m_x und Lemma 9.6 verwendet haben. Also

$$m_x - m^* \geq \frac{-2\xi m^*}{\sqrt{\sum_{J \in V^c} J_{\gamma\ell}(x-y)}},$$

wobei wir benutzen, dass

$$\sum_{J \in V^c} J_{\gamma\ell}(x-y) \leq \sum_{J \in V} J_{\gamma\ell}(x-y)$$

gilt. Auf der anderen Seite gilt (9.1) für alle $y \in V$, so dass man, setzt man diese Abschätzung in die erste Zeile von (9.1) ein, erhält:

$$m_x - m^* \geq (m_x - m^*) \sum_{y \in V} \sum_{t \in V} J_{\gamma\ell}(x-y) J_{\gamma\ell}(y-z) - 4\xi m^*.$$

Aus diesen beiden Abschätzungen sieht man, dass wir fertig sind, wenn wir

$$\sqrt{\sum_{y \in V^c} J_{\gamma\ell}(x-y)} \geq 4\xi$$

oder

$$1 - \sum_{y \in V} \sum_{z \in V} J_{\gamma\ell}(x-y) J_{\gamma\ell}(y-z) \geq 8\xi$$

zeigen können. Dies folgt aus der Tatsache, dass V aus Würfeln aufgebaut ist, deren Kantenlänge gerade so groß ist wie die Reichweite der Wechselwirkung. In der Tat gilt

$$1 - \sum_{y \in V} \sum_{z \in V} J_{\gamma\ell}(x-y) J_{\gamma\ell}(y-z) = \sum_{y \in V^c} J_{\gamma\ell}(x-y) + \sum_{y \in V} \sum_{z \in V^c} J_{\gamma\ell}(x-y) J_{\gamma\ell}(y-z).$$

Der zweite Term dieser Summe kann also für reguläre V nicht zu klein sein, so lange $\operatorname{dist}(x, V^c) \leq \frac{1}{\gamma\ell}$ gilt. Man überzeugt sich, dass im schlimmsten Falle x von einer "Ecke" von V^c Abstand $\frac{r}{\gamma\ell}$ hat. Aber selbst erhält man

$$\begin{aligned} &\sum_{y \in V} \sum_{z \in V^c} J_{\gamma\ell}(x-y) J_{\gamma\ell}(y-z) \geq 2^{-(d-1)} \int_0^1 (r+s)^{d-1} (1-s)^d ds \\ \geq\ & 2^{-(d+1)} \int_0^1 s^{d-1} (1-s)^d ds = 2^{-(d+2)} \frac{((d-1)!)^2}{(2d-1)!}, \end{aligned}$$

so dass die Behauptung stimmt, wenn man 4ξ kleiner wählt als diese rechte Seite.

Somit folgt $m_x \geq m^*/2$ in V, was natürlich auch unsere Eingangsannahme $\phi_x(m) > 0$ rechtfertigt. Also ist

$$m_x^{\text{opt}} \geq m^*/2.$$

Gleichermaßen zeigt man $m_x^{\text{opt}} \leq \frac{3}{2}m^*$. Dies beweist das Lemma. □

In der Folge werden wir die nachfolgende Definition benötigen:

Definition 9.7 *Eine reguläre Menge V heißt n-Schicht-Oval, wenn V von der Form*

$$V = \{x \in \tilde{V}^c : dist(x, \tilde{V}) \leq \frac{k}{\gamma\ell}\}$$

für eine zusammenhängende, aus Blöcken n bestehende Menge $\tilde{V}$. Die Mengen

$$V_k \equiv \{x \in \tilde{V}^c : \frac{k-1}{\gamma\ell} < dist(x, \tilde{V}) < \frac{k}{\gamma\ell}\}$$

heißen die k-te Schicht von V.

Beispiel 9.8 *Definitionsgemäß sind die Mengen $\partial\Gamma$ n-Schicht-Ovale.*

Der Beweis des folgendes Lemmas ist eine kleine Übung.

Lemma 9.9 *Sei*

$$f_\beta(m) = \frac{I(m)}{\beta} - \frac{m^2}{2}.$$

Dann gilt für alle $m \in [-1, 1]$

$$f_\beta(m) - f_\beta(m^*) \geq C(\beta)(|m| - m^*)^2,$$

wobei

$$C(\beta) := \frac{\log\cosh(\beta m^*)}{\beta(m^*)^2} - \frac{1}{2} > 0$$

die Eigenschaft hat, dass

$$\lim_{\beta \downarrow 1} \frac{C(\beta)}{(m^*)^2} = \frac{1}{12}$$

gilt.

Beweis: Übung. □

Lemma 9.10 *Sei V ein n-Schicht-Oval mit $n \geq \frac{r}{c(\beta)}$. Dann gibt es eine Schicht V_k in V, so dass gilt:*

$$\sum_{x \in V_k} (m_x^{opt})^2 \leq 2^{-r-3}(m^*)^2(|V_1| + |V_n|).$$

Beweis: Setze

$$u_x := |m_x| - m^* \quad \text{und} \quad \|u_{V_k}\|_2^2 := \sum_{x \in V_k} (u_x)^2$$

(und analog für andere Funktionen). Dann folgt aus der Darstellung von $\tilde{E}$, dass für jede Konfiguration gilt:

$$\tilde{E}_{V|V_1|V_2}(m_{V|V_1|V_n}, m_{V_1 \cup V_n}) \geq \sum_{k=2}^{n-1} c(\beta) \|u_{V_k}\|_2^2 + \sum_{x \in V|V_1|V_n} f_\beta(m^*).$$

Andererseits können wir auch eine Konfiguration betrachten, die auf V_1 und V_n gleich m^{opt} ist und gleich m^* auf $V|V_1|V_n$. Hierfür gilt

$$\tilde{E}_{V|V_1|V_2}(m_{V|V_1|V_n} = m^*, m^{\mathrm{opt}}_{V_1 \cup V_n}) = \frac{1}{2} \sum_{\substack{x \in V|V_1|V_1 \\ y \in V_1 \cup V_n}} J_{\gamma\ell}(x-y)(m^{\mathrm{opt}}_y - m^*)^2 + \sum_{x \in V|V_1|V_n} f_\beta(m^*).$$

Nach Definition von m^{opt} gilt

$$\begin{aligned}
0 &\geq \tilde{E}_{V|V_1|V_2}(m^{\mathrm{opt}}_{V|V_1|V_n}, m^{\mathrm{opt}}_{V_1 \cup V_n}) - \tilde{E}_{V|V_1|V_2}(m^{\mathrm{opt}}_{V|V_1|V_n} = m^*, m^{\mathrm{opt}}_{V_1 \cup V_n}) \\
&\geq \sum_{k=2}^{n-1} c(\beta)\|u_{V_k}\|_2^2 - \frac{1}{2} \sum_{\substack{x \in V|V_1|V_2 \\ y \in V_1 \cup V_2}} J_{\gamma\ell}(x-y)(m^{\mathrm{opt}}_y - m^*)^2 \\
&\geq \sum_{k=2}^{n-1} c(\beta)\|u_{V_k}\|_2^2 - \frac{1}{2}(\|u^{\mathrm{opt}}_{V_1}\|_2^2 + \|u_{V_n}\|_2^2).
\end{aligned}$$

Also gilt für alle $q < \frac{n}{2}$:

$$\begin{aligned}
&qc(\beta) \inf_{k=2,\dots,q+1} [\|u^{\mathrm{opt}}_{V_k}\|_2^2 + \|u^{\mathrm{opt}}_{V_{n=1+k}}\|_2^2] \\
\leq &\sum_{k=2}^{n-1} c(\beta)[\|u^{\mathrm{opt}}_{V_k}\|_2^2 + \|u^{\mathrm{opt}}_{V_{n+1-k}}|_2^2] \leq \frac{1}{2}\|u^{\mathrm{opt}}_{V_1}\|_2^2 + \frac{1}{2}\|u^{\mathrm{opt}}_{V_n}\|_2^2.
\end{aligned}$$

Von hier aus erhalten wir

$$\inf_{k=2,\dots,q+1} [\|u_{V_k}\|_2^2 + \|u_{V_{n+1-k}}\|_2^2] \leq \frac{1}{2qc(\beta)} [\|u^{\mathrm{opt}}_{V_1}\|_2^2 + \|u^{\mathrm{opt}}_{V_n}\|_2^2].$$

Wählt man $q := [\frac{1}{c(\beta)}$, so erhält man die Existenz eines $2 \leq k \leq q+1$, so dass

$$[\|u_{V_k}\|_2^2 + \|u_{V_{n+1-k}}\|_2^2] \leq \frac{1}{2}[\|u^{\mathrm{opt}}_{V_1}\|_2^2 + \|u^{\mathrm{opt}}_{V_n}\|_2^2].$$

Wiederholt man diese Prozedur und bedenkt zudem, dass

$$\frac{1}{2}[\|u^{\mathrm{opt}}_{V_1}\|_2^2 + \|U^{\mathrm{opt}}_{V_n}\|_2^2] \leq \frac{1}{8}(m^*)^2(|V_1| + |V_n|),$$

erhalten wir die Behauptung des Lemmas. □

Wir sind nun in der Lage, unsere Referenz-Konfiguration zu konstruieren und eine obere Schranke für ihre Energie anzugeben: Gegeben sei eine Kontur Γ und eine kompatible Konfiguration m auf $\underline{\Gamma}$ und $\tilde{\underline{\Gamma}}$, nennen wir m^{opt} die Konfiguration auf $\underline{\Gamma}$, die die Energie unter diesen Randbedingungen optimiert, d. h. minimiert. Diese Konfiguration ist auch im Sinne der vorhergehenden Definition optimal. Alwo wissen wir aus dem letzten Lemma, dass es in jeder Zusammenhangskomponente $\partial\underline{\Gamma}_i^+$ des Random von Γ eine Schicht $\mathcal{L}_i^\pm$ der Dirchte $\frac{1}{\gamma\ell}$ in $\partial\Gamma_i^\pm$ gibt mit

$$\|m^{\text{opt}}_{\mathcal{L}_i^\pm} \mp m^*\|_m^2 \leq 2^{-r-3}(m^*)^2[|V_1(\partial\underline{\Gamma}_i^\pm)| + V_n(\partial\underline{\Gamma}_i^\pm)|].$$

Für ein gegebenes $\mathcal{L}_i^\pm$ zerlegen wir $\partial\underline{\Gamma}_i^\pm$ in 2 Mengen

$$\partial\Gamma_{i,in}^\pm := \{x \in \partial\underline{\Gamma}_i^\pm \backslash \mathcal{L}_i^\pm : \text{dist}(x, D^\pm) > \text{dist}(\mathcal{L}_i^\pm, D^\pm)\}$$

und

$$\partial\Gamma_{i,\text{out}}^\pm := \partial\underline{\Gamma}^\pm \backslash \partial\Gamma_{i,\text{in}}^\pm.$$

OBdA sei nun der äußere Rand unserer Kontur mit einer positiven Randbedingung versehen. Wir definieren unsere Referenzkonfiguration m^{ref} als

$$m_x^{\text{ref}} = \begin{cases} m_x^{\text{opt}}, & x \in \partial\underline{\Gamma}_{i,\text{out}}^+ \\ -m_x^{\text{opt}}, & x \in \partial\Gamma_{i,\text{out}}^- \\ m^* & \text{für alle anderen } x \in \underline{\Gamma} \\ m_x, & x \in D^+ \\ -m_x, & x \in D^- \end{cases}.$$

Lemma 9.11 *Für jede äußere Konfiguration, die mit m^{ref} verträglich ist, gilt*

$$\begin{aligned} \tilde{E}_{\underline{\Gamma}}(m_\Gamma^{ref}, m_{\Gamma^c}^{ref}) \quad &\geq \quad \sum_{i,\pm} \tilde{E}^\pm_{\partial\underline{\Gamma}_{i,out}}(m^{opt}_{\partial\underline{\Gamma}_{i,out}}, m^{ref}_{\underline{\Gamma}^c}) \\ &\quad + \sum_{i,\pm} 2^{-r-3}(m^*)^2[|V_1(\partial\underline{\Gamma}_i^\pm)| + |V_n(\partial\underline{\Gamma}_i^\pm)|] + \sum_{x \in \underline{\Gamma}|\partial\underline{\Gamma}_{out}} f_\beta(m^*). \end{aligned}$$

Beweis: Das ist nur ein Zusammenfügen der vorhergehenden Abschätzungen. □

Wir wenden uns nun dem zweiten Schritt zu: Einer Energieschranke für alle Konfigurationen, die mit die Kontur Γ enthalten. Dazu ein Hilfsresultat:

Lemma 9.12 *Seien $U, V, W \subseteq \mathbb{Z}^d$ disjunkt, dergestalt, dass für alle $y \in U \cup W$*

$$\sum_{X \in U \cup V \cup W} J_{\gamma\ell}(x-y) = 1$$

und für jedes $y \in U$

$$\sum_{X \in V \cup W} J_{\gamma\ell}(x-y)$$

gilt. Dann folgt

$$
\begin{aligned}
& \tilde{E}_{V \cup U \cup W}(m_{U \cup V \cup W}, m_{(U \cup V \cup W)^c}) \\
\geq \; & \frac{1}{4} \sum_{X \in U} \psi_x(m) + \frac{1}{2} \sum_{X \in U \cup W} [f_\beta(m_x) + f_\beta(\phi_x(m))] + \sum_{x \in V} f_\beta(m^*).
\end{aligned}
$$

Beweis: Das ist eine einfache, aber längliche Rechnung, die hier nicht vorgeführt werden soll. □

Dieses Lemma erlaubt es uns, die Energie einer Konfiguration nur in Ausdrücken von $\phi_x(m)$ und $\psi_x(m)$ zu beschreiben. Nimmt man nämlich für V die Schichten $\mathcal{L}_i^\pm$ und die Gebiete "zwischen" $\mathcal{L}_i^\pm$, erhalten wir für jede Konfiguration

$$
\begin{aligned}
\tilde{E}_{\underline{\Gamma}}(m_{\underline{\Gamma}}, m_{\underline{\Gamma}^c}) \;\; \geq \;\; & \sum_{i,\pm} \tilde{E}^+_{\partial \underline{\Gamma}_{i,\mathrm{out}}}(m^+_{\partial \Gamma_{i,\mathrm{out}}}, m_{\underline{\Gamma}^c}) + \frac{1}{2} \sum_{x \in \underline{\Gamma} \setminus \partial \underline{\Gamma}_{\mathrm{out}}} [f_\beta(\phi_x(m)) - f_\beta(m^*)] \\
& + \sum_{\substack{x \in \underline{\Gamma} \setminus \partial \underline{\Gamma}_{\mathrm{out}} \\ \mathrm{dist}(x, \partial, \Gamma^-_{\mathrm{out}}) > \frac{1}{\gamma \ell}}} \quad .
\end{aligned}
$$

Nun haben $\phi(\cdot)$ und $\psi(\cdot)$ angenehme Stetigkeitseigenschaften.

Lemma 9.13 *Es gibt eine Konstante $c_d < +\infty$, so dass für jede Kontur Γ und jedes*

$$y \in \underline{\hat{\Gamma}} : \{y : dist(y, \underline{\tilde{\Gamma}}) \leq \frac{(\xi -^*)^2}{\gamma_{c_d} \gamma \ell}\}$$

gilt

$$\| \phi_y(m)| - m^*| \geq \frac{\xi m^*}{2} \quad oder \quad \psi_x(m) \geq \frac{(\xi m^*)^2}{2}.$$

Beweis: Das überlegt man sich selbst. □

Lemma 9.14 *Es gibt $c'_d < +\infty$, so dass für jede Kontur Γ gilt*

$$|\underline{\Gamma}| \leq c'_d \frac{(n+1)^d}{(\xi m^*)^{2d}} |\underline{\hat{\Gamma}}|.$$

Beweis: $\frac{|\underline{\Gamma}|}{|\underline{\hat{\Gamma}}|}$ ist maximal, falls $|\underline{\tilde{\Gamma}}| = 1$. Dann ist aber die Behauptung wahr. □

Kombiniert man nun die Abschätzungen für die Energie der Referenzkonfiguration m^{ref} mit den Abschätzungen für die Konfiguration, die mit der Anwesenheit einer Kontur verträglich sind, ergibt sich

Proposition 9.15 *Sei* $\Gamma = (\underline{\Gamma}, m)$ *eine Kontur mit festem* $\underline{\Gamma}$. *Dann gibt es eine Referenzkonfiguration* m^{ref}, *die* Γ *nicht enthält, so dass gilt*

$$E_{\underline{\Gamma}}(m_{\underline{\Gamma}}, m_{\underline{\Gamma}^c}) - E_{\underline{\Gamma}}(m^{ref}_{\underline{\Gamma}}, m^{ref}_{\underline{\Gamma}^c}) \geq \frac{1}{8}\frac{c(\beta)}{c_d}\frac{(\xi m^*)^{2d+2}}{(n+1)^d}|\Gamma| - \frac{1}{8}(m^*)^2 2^{-nc(\beta)}|\underline{\Gamma}|.$$

Wählt nun s, dass

$$2^{-nc(\beta)} = \frac{1}{2}\frac{c(\beta)(\xi m^*)^{2d}}{c_d(n+1)^d}$$

(diese Gleichung hat natürlich nur eine Lösung $\xi \in \mathbb{R}$, wir berechnen mit n^*, $\lceil \xi \rceil := n^*$), so erhalten wir

$$E_{\underline{\Gamma}}(m_{\underline{\Gamma}}, m_{\underline{\Gamma}^c}) - E_{\underline{\Gamma}}(m^{\text{ref}}_{\underline{\Gamma}}, m^{\text{ref}}_{\underline{\Gamma}^c}) \geq \frac{1}{16}\frac{c(\beta)(\xi m^*)^{2d+2}}{c_d(n+1)^d}|\Gamma|.$$

Man überlegt sich, dass

$$n^* \leq c\frac{1}{c(\beta)}\log[\frac{c(\beta)(\xi m^*)^{2d}}{2c_d}]$$

für eine Konstante C, falls $c(\beta)$ klein genug ist. Die Abschätzungen für die Energiedifferenz oben sind natürlich nur hilfreich, wenn die Schranken wesentlich größer sind als die Fehlerterme, die man aus der Blockspinapproximation erhält. Dafür muss

$$\frac{1}{16}\frac{c(\beta)}{c_d}\frac{(\xi m^*)^{2d+2}}{(n+1)^d} > \tilde{c}_d\,\gamma\ell$$

gelten. Hierbei wird die rechte Seite klein, da $\gamma \to 0$ konvergent und $2\gamma \to 0$ geht. ℓ schließlich bestimmt sich dadurch, dass der Logarithmus der Anzahl von möglichen $\underline{\Gamma}$ mit festem Volumen $|\underline{\Gamma}|$ und die Anzahl der m, die mit einem solchen $|\underline{\Gamma}|$ kompatibel sind, klein ist gegen die Energiedifferenz. Da sich besagte Anzahl durch

$$\ell^{d|\Gamma|}C^{|\Gamma|}$$

abschätzen lässt, führt diese Überlegung zu der Bedingung

$$\beta\ell^d[\frac{1}{16}\frac{c(\beta)}{c_d}\frac{(\xi m^*)^{2d+2}}{c_d(n+1)^d} - c_d\gamma\ell] > d\log\ell + \log C.$$

Hierzu muss ℓ hinreichend klein sein, z. B.

$$\ell = \frac{1}{\gamma}\frac{1}{c_d}\frac{1}{32}\frac{(c(\beta)(\xi m^*)^{2d+2}}{c_d(n+1)^d},$$

wodurch man, setzt man es in die vorhergehende Gleichung ein, ein Verhältnis von β und γ ergibt. Es ist klar, dass dies erfüllt werden kann, so lange $\beta > 1$ ist und γ hinreichend klein. Somit haben wir insgesamt gezeigt

$$\mathbb{P}(m : m \text{ ist verträglich mit festem } \Gamma) \leq e^{-c|\Gamma|\log\ell|}.$$

Und insgesamt

$$\mathbb{P}(\exists\,\Gamma : 0 \in \operatorname{int}\Gamma) \leq Ce^{-c\beta n^d|\log C|}.$$

Dies beweist den Satz. $\square$

10 Ungeordnete Systeme und Spingläser – einleitende Bemerkungen

Der Erfolg des Ising-Modells und verwandter Modelle bei der Modellierung des Ferromagnetismus animierte Physiker zu Beginn der 70er Jahre, ähnliche Modelle zur Modellierung amorpher Substanzen auszuprobieren. Die entstehenden Probleme sind heutzutage mathematisch noch weitestgehend ungelöst. Der Ansatz ist es hierbei, statt Interaktion, die eine global einheitliche Orientierung der Spins bevorzugen, solche zu wählen, die teils eine gleichsinnige Orientierung der Spins favorisieren (ferromagnetische Interaktion), teils eine ungleichgerichtete (antiferromagnetische Interaktion). Das intuitive Bild hierbei ist in etwa das folgende: Für kleine β sollten die Spins unter dem Gibbs-Maß alle in etwa gleich wahrscheinlich sein – man spricht von der PARAMAGNETISCHEN PHASE. Bei tiefen Temperaturen, d. h. großen β, halten sich die Spinkonfigurationen dagegen in der Nähe der Minima der Energiefunktion auf. Diese haben aber – anders als im Ising-Modell – in der Regel die Magnetisierung null (oder nahezu null). Das historisch erste Modell eines sogenannten Spinglases ist auch zugleich eines der sinnigsten. Es ist die ungeordnete Version des Ising-Modells, das sogenannte Edwards-Anderson-Modell. Es wurde von Edwards und dem Nobelpreisträger Anderson 1975 vorgeschlagen. Die lokale Hamiltonfunktion ist hierbei

$$H_\Lambda(\sigma) = -\sum_{\substack{<i,j> \\ i,j\in\Lambda}} \sigma_i\sigma_j J_{ij}.$$

Hierbei ist $\Lambda \subseteq \mathbb{Z}^d$ endlich und $\sigma \in \{-1,+1\}^\Lambda$. Die J_{ij} sind i.i.d. Zufallsvariablen, die unabhängig von σ gewählt werden und vor diesen. J_{ij} hat dabei Erwartungswert 0, also etwa

$$J_{ij} \sim \mathcal{N}(0,\sigma^2) \quad \text{oder} \quad J_{ij} \sim \mathrm{Ber}(\frac{1}{2}) \quad \text{auf} \quad \{-1,+1\}.$$

Die Hauptschwierigkeit liegt dabei darin, dass es sehr viele σ_i gibt, die aufgrund der J_{ij} und der anderen σ_j sowohl Nachbarn haben, die $\sigma_i = +1$ favorisieren als auch solche, die $\sigma_i = -1$ bevorzugen. Die Grundzustände des Systems, d. h. die Zustände niedrigster Energie, sind nicht mehr direkt ablesbar. Dieses Phänomen heißt Frustration (ebenso wie der Gemütszustand vieler Mathematiker, die sich damit beschäftigen).

Schon bald nachdem das Edwards-Anderson-Modell vorgestellt worden war, wurde klar, dass es einer mathematischen Lösung nicht zugänglich ist – ja, mehr als 30 Jahre nach seiner Entwicklung sind noch verschiedene Heuristiken über ganz zentrale Fragen seines Verhaltens im Umlauf, etwa:

- Ab welcher Dimension ist ein Phasenübergang in dem Modell vorhanden?
- Wie verhält sich die Anzahl der Grundzustände typischerweise (exponentiell in $|\Lambda|$, polynomiell in $|\Lambda|$, endlich, ...)?

Daher entwickelten Sherrington und Kirkpatrick kurz nach Edwards und Anderson eine mean-field-Variante des Edwards-Anderson-Modells.

10.1 Das Sherrington-Kirkpatrick-Modell

Wie oben beschrieben stellt dieses Modell eine mean-field-Version des Edward-Anderson-Modells dar. Seine Hamiltonfunktion ist gegeben durch

$$H_{N,h}(\sigma) = -\frac{1}{\sqrt{N}} \sum_{i<j} J_{ij}\sigma_i\sigma_j - h\Sigma\sigma_i,$$

für $\sigma \in \{-1,+1\}^N$ und $h \in \mathbb{R}$. Die J_{ij} sind hierbei in der Regel i.i.d. und man nimmt an, dass

$$J_{ij} \sim \mathcal{N}(0,1)$$

gilt. Allerdings sollte dies keine Rolle spielen, solange sie nur symmetrisch und zentriert sind und genügend viele Momente von J_{ij} existieren. Das zugehörige Gibbs-Maß ist gegeben durch

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{-\beta H_{N,h}(\sigma)}}{Z_{N,\beta,h}}$$

mit

$$Z_{N,\beta,h} = \sum_{\sigma'} e^{-\beta H_{N,h}(\sigma')}.$$

Wie immer ist die freie Energie, also eine Version von

$$\frac{1}{N}\log Z_{N,\beta,h}$$

von Interesse. Man beachte, dass dies nun eine Zufallsgröße (in den (J_{ij})) ist. Will man wie üblich eine Zahl erhalten, so muss man über die J_{ij} mitteln. Dies kann auf zwei Arten geschehen:

1. $f^{an}_{\beta,h} := \lim_{N\to\infty} \frac{1}{N}\log \mathbb{E}_J Z_{N,\beta,h}$;
2. $f^{q}_{\beta,h} := \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_J \log Z_{N,\beta,h}$.

Hierbei bezeichnet $\mathbb{E}_J$ den Erwartungswert bezüglich der J_{ij}; die erste Größe nennt man die "annealte" freie Energie, die zweite Größe die "gequenchte" freie Energie. Es ist unmittelbar deutlich, dass man an der gegenechten freien Energie ein viel größeres Interesse haben sollte. Außerdem tendieren die beiden Größen aufgrund der Jensenschen Ungleichung dazu, verschieden zu sein. Interesanterweise gibt es im SK-Modell einen Temperaturbereich (den sogenannten Hochtemperaturbereich), in dem

$$f^{an}_{\beta} = f^{q}_{\beta}$$

gilt (dies ist für $\beta < \beta_c = 1$ der Fall). Dies hat den Vorteil, dass sich f^{an}_{β} leichter berechnen lässt:

Satz 10.1 *Es gilt im SK-Modell für alle* $\beta \in \mathbb{R}^+$

$$f^{an}_{\beta,0} = \beta^2/4 + \log 2.$$

Beweis: Der ist nicht tiefsinnig. Die "wichtigste" Eigenschaft eines Gaußschen $\mathcal{N}(0, \sigma^2)$-verteilten Zufallsvariablen Y ist

$$\mathbb{E}e^{tY} = e^{t^2\sigma^2/2}.$$

In der Tat gilt ja

$$\begin{aligned}
\mathbb{E}e^{tY} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{ty} e^{-y^2/2\sigma^2} dy \\
&= e^{t^2\sigma^2/2} \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{1}{2\sigma^2}(t\sigma^2 - y)^2} dy \\
&= e^{t^2\sigma^2/2} \frac{1}{\sqrt{2\pi}} \int e^{-z^2/2} dz \\
&= e^{t^2\sigma^2/2}
\end{aligned}$$

mit $\dot{z} = \frac{t\sigma^2 - y}{\sigma^2}$. Damit folgt

$$\begin{aligned}
\mathbb{E}_J Z_{N,\beta,0} &= 2^N \mathbb{E}_\sigma \mathbb{E}_J e^{\frac{\beta}{\sqrt{N}} \sum_{iyj} J_{ij}\sigma_i\sigma_j} \\
&= 2^N \mathbb{E}_\sigma \prod_{i<j} \mathbb{E}_J e^{\frac{\beta}{\sqrt{N}}\sigma_i\sigma_j J_{ij}} \\
&= 2^N \mathbb{E}_\sigma \prod_{i<j} e^{\frac{\beta^2}{2N}} \\
&= 2^N \mathbb{E}_\sigma e^{\frac{\beta^2\binom{N}{2}}{2N}}.
\end{aligned}$$

Also ist

$$\begin{aligned}
f_{\beta,0}^{an} &= \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_J Z_{N,\beta,0} + \log 2 \\
&= \lim_{N\to\infty} \frac{1}{N} \frac{\beta^2 N(N-1)}{4N} + \log 2 = \beta^2/4 + \log 2.
\end{aligned}$$

□

Wir wollen nun eine Methode kennenlernen, auch f_β^q zu berechnen. Allerdings ist diese nicht heuristisch.

10.2 Die Replica-Methode

Die Grundidee der Replica-Methode ist relativ simplel: $\mathbb{E}_J \log Z_{N,\beta,h}$ ist nicht einfach zu berechnen, also ersetzt man $\log Z_{N,\beta,h}$ durch eine andere Funktion von $Z_{N,\beta,h}$, die einfacher zu berechnen ist. Dazu erinnert man sich, dass gemäß der l'Hospitalschen Regel

$$\log x = \frac{x^n - 1}{n}|_{n=0}$$

gilt. Setzt man stillschweigend voraus, dass man Erwartungswerte mit Limiten vertauschen darf, dann sollte

$$\mathbb{E}_J \log Z_{N,\beta,h} = \lim_{n\to 0} \frac{\mathbb{E}_J Z_{N,\beta,h}^n - 1}{n}$$

gelten. Für ganzzahliges n lässt sich $\mathbb{E}_J Z^n_{N,\beta,h}$ aber annäherungsweise berechnen. Dazu stellt man sich Z^n einfach als Zustandssumme von n unabhängigen Systemen der Größe N vor und erhält

$$2^{-N}\mathbb{E}Z^n = \int \mathbb{E}_\sigma e^{\beta/\sqrt{N}\sum_{i<j} J_{ij}\sum_{\alpha=1}^n \sigma_i^\alpha\sigma_j^\alpha + \beta h \sum_{i=1}^N\sum_{\alpha=1}^n \sigma_i^\alpha} d\mathbb{P}_J.$$

Benutzt man erneut, dass für eine Gaußsche Variable Y gilt

$$\mathbb{E}e^{tY} = e^{t\mathbb{E}Y^2/2}, \tag{10.1}$$

so erhält man, dass der obige Integral gleich

$$\begin{aligned} & C\cdot\mathbb{E}_\sigma \exp\left(\frac{1}{N}\sum_{i<j}(\frac{1}{2}\beta^2\sum_{\alpha_1,\alpha_2}\sigma_i^{\alpha_1}\sigma_j^{\alpha_1}\sigma_i^{\alpha_2}\sigma_j^{\alpha_2}) + \beta h\sum_\alpha\sum_i\sigma_i^\alpha\right) \\ = \ & \text{Const. } e^{\frac{N\beta_2 n}{4}}\mathbb{E}_\sigma\exp\{\frac{\beta^2}{2N}\sum_{\alpha_1<\alpha_2}(\sum_i \sigma_i^{\alpha_1}\sigma_i^{\alpha_2})^2 + \beta h\sum_i\sum_\alpha\sigma_i^\alpha\}. \end{aligned}$$

Definiere q_{α_1,α_2} als Gaußsche Zufallsvariablen, über deren Kovarianz noch zu reden sein wird. Durch ein wenig Herumrechnen, wobei man (10.1) benötigt, erhält man

$$2^{-N}\mathbb{E}Z^n = \int C\cdot\exp\{\frac{N}{4}\beta^2 n - \frac{N\beta^2}{2}\sum_{\alpha_1<\alpha_2} q^2_{\alpha_1\alpha_2} + N\log\mathbb{E}_\sigma e^L\}dq_{\alpha_1\alpha_2},$$

wobei

$$L = \beta^2\cdot\sum_{\alpha_1<\alpha_2} q_{\alpha_1,\alpha_2}\sigma_1^{\alpha_1}\sigma_1^{\alpha_2} + \beta\sum_\alpha h\sigma_1^\alpha$$

ist. Die Sattelpunktmethode besagt nun grob gesprochen, dass dieses Integral ungefähr denselben Wert liefert, wenn man das Maximum der Exponenten anschaut. Bezeichnen wir diesen wie den Exponenten selbst, vertauschen die Limiten $N\to\infty$ und $n\to\infty$ und entwickeln die Exponentialfunktion bis zur ersten Ordnung, so ergibt sich

$$\begin{aligned} 2^{-N}\mathbb{E}Z^n &\approx \exp\{\frac{N}{4}\beta^2 n - \frac{N}{2}\beta^2\sum_{\alpha_1<\alpha_2} q^2_{\alpha_1,\alpha_2} + N\log\mathbb{E}_\sigma e^L\} \\ &\approx 1 + Nn[\frac{\beta^2}{4} - \frac{\beta^2}{4n}\sum_{\alpha_1\neq\alpha_2} q^2_{\alpha_1,\alpha_2} + \frac{1}{n}\log\mathbb{E}_\sigma e^L]. \end{aligned}$$

Und somit

$$f^{an}_{\beta,h} = \lim_{n\to 0}\frac{\mathbb{E}_j Z^n - 1}{nN} = \lim_{n\to\infty}\{\frac{-\beta^2}{4n}\sum_{\alpha_1\neq\alpha_2} q^2_{\alpha_1,\alpha_2} + \frac{\beta^2}{4} + \frac{1}{n}\log\mathbb{E}_\sigma e^L\}.$$

Hierbei müssen die q_{α_1,α_2} Maximalwerte liefern, was sich in den Gleichungen

$$q_{\alpha_1,\alpha_2} = \frac{1}{\beta^2}\frac{\partial}{\partial q_{\alpha_1,\alpha_2}}\log\mathbb{E}_\sigma e^L = \frac{\mathbb{E}_\sigma\sigma_1^{\alpha_1}\sigma_1^{\alpha_2}e^L}{\mathbb{E}_\sigma e^L} \tag{10.2}$$

ausdrücken lässt.

Bis hierher ist "physikalisch" alles in Ordnung, wenn auch nicht mathematisch. Auf der mathematischen Seite haben wir zum einen den Limes $n \to 0$ durch das Integral gezogen (was nicht einsichtig ist) und zum anderen die Limiten $N \to \infty$ und $n \to 0$ vertauscht. Vor allem hierfür fehlt jede mathematische Berechtigung. Am schlimmsten ist allerdings, dass die obigen Berechnungen nur für ganzzahlige $n \in \mathbb{N}$ ihren Wert besitzen. Nun den Limes $n \to 0$ zu nehmen, entspricht der Vermutung, dass nur eine Funktion, deren Werte für $n \in \mathbb{N}$ man kennt, nur auf EINE Art in 0 fortsetzen kann. Diese Annahme ist so nicht berechtigt. Die heuristischen Berechnungen werden folgendermaßen fortgesetzt: Um (10.2) zu lösen, muss die Abhängigkeit der q_{α_1,α_2} von α_1 und α_2 bekannt sein. Naiverweise könnte man vermuten, dass die Indizes α_1, α_2 die Größe von q_{α_1,α_2} gar nicht beeinflussen, so dass

$$q_{\alpha_1,\alpha_2} \equiv q$$

gilt. In der Tat haben die Replicas ja keine physikalische Bedeutung, sondern wir haben sie aus technischen Gründen eingeführt.

Tatsächlich aber führt diese Annahme nach einigen länglichen Rechnungen, die hier nicht vorgeführt werden sollen, dazu, dass man die Entropie bei $T = 0$, d. h. $\beta = \infty$ als negativ bestimmen kann. Dies ist inkonsistent mit der üblichen Definition der Entropie und wurde zunächst auf das Vertauschen der Limiten ($n \to 0$ und $N \to \infty$) zurückgeführt. Dieses Resultat führte dazu, dass der Ansatz

$$q_{\alpha_1,\alpha_2} \equiv q$$

(der sogenannte Replica-symmetrische Ansatz) jenseits einer Kurve im $h-\beta$-Phasendiagramm, die insbesondere den Fall $h = 0$, β groß ausgrenzt, verworfen wurde. Diese Kurve heißt die Almeida-Thonless-Linie. Jenseits dieser Linie wurde postuliert, dass die q_{α_1,α_2} sehr wohl von α_1 und α_2 abhängen. Dieses Phänomen heißt die Brechung der Replica-Symmetrie.

1980 beschrieb Parisi in einem genialen Ansatz, wie die q_{α_1,α_2} zu wählen sind. Von einer Overlap-Struktur für die q_{α_1,α_2} der Form

$$\begin{pmatrix} 0 & q \\ q & 0 \end{pmatrix}$$

gelangt er über

$$\left(\begin{array}{cc|cc} 0 & q_1 & & q_0 \\ q_1 & 0 & & \\ \hline q_0 & & & q_1 \\ & & q_1 & \end{array}\right)$$

und

$$\left(\begin{array}{cc|ccc} 0 & q_1 & & q_0 & \\ q_2 & 0 & & & \\ \hline & & 0 & q_2 & \\ & q_0 & q_2 & 0 & q_1 \\ & & & q_1 & q_2 \end{array}\right)$$

zu immer feineren Strukturen.

Dabei wird q_i als eine Treppenfunktion gewählt, die die Identität approximiert. Dieser Ansatz mag nicht nur Mathematikern als absolut willkürlich erscheinen. Er führt jedoch zu einer Differentialgleichung für die freie Energie, deren Richtigkeit 2004 von M. Talagrand mathematisch bewiesen wurde. Diese ist nicht sonderlich einleuchtend und soll hier daher weder dargestellt noch hergeleitet werden.

Wir wenden uns nun ungeordneten Modellen zu, die leichter analysierbar sind.

11 Das Random Energy Modell (REM)

11.1 Das Gesetz der großen Zahlen

Wir haben im vorhergehenden Kapitel schon zwei Modelle mit zufälligen Energiefunktionen kennengelernt und auch gesehen, dass beide relativ schwierig zu lösen sind. Wir wollen daher in der Folge einfachere Modelle studieren.

Das allereinfachste, was einem dabei einfallen kann, ist, wieder

$$S_N = \{-1, +1\}^N$$

zu betrachten und darauf eine vollkommen zufällige Energiefunktion, in dem Sinne, als dass man jedem $\sigma \in S_N$ eine Zufallsvariable zuweist. Genauer ist die Energiefunktion des Random Energy Modells (REM) gegeben durch

$$H_N(\sigma) = -\sqrt{N} X_\sigma.$$

Dabei sind die X_σ i.i.d. $\mathcal{N}(0,1)$-verteilte Zufallsvariablen und der Faktor $\sqrt{N}$ wurde eingeführt, damit die freie Energie pro Punkt endlich ist und nicht identisch verschwindend. Ehrlicherweise sollte hier bemerkt werden, dass beim REM nur noch wenig von der Struktur wechselwirkender Spins übriggeblieben ist. Dennoch ist es ein nettes Spielzeugmodell, um sich mit einigen Eigenheiten ungeordneter Systeme vertraut zu machen.

Wir beginnen mit dem Studium der Grundzustandsenergie, also dem Maximum der X_σ. Da die X_σ i.i.d. sind, ist dies nicht besonders schwierig. Als Vorbereitung benötigen wir

Lemma 11.1 *Für alle $u > 0$ gilt*

$$\frac{1}{u} e^{-u^2/2}(1 - \frac{2}{u^2}) \leq \int_u^\infty e^{-x^2/2} dx \leq \frac{1}{u} e^{-u^2/2}.$$

Beweis: Übung. □

Naheliegenderweise werden wir dieses Lemma benutzen, um die Extrema von Gaußvariablen zu beschränken.

Proposition 11.2 *Für die Energie des Grundzustandes im REM gilt:*

$$\lim_{N\to\infty} \max_{\sigma \in S_N} -\frac{1}{N} H_N(\sigma) = \lim_{N\to\infty} \max_{\sigma \in S_N} \frac{1}{\sqrt{N}} X_\sigma = \sqrt{2 \log 2}.$$

Hierbei gilt die Konvergenz fast sicher und in L^1.

Beweis: Aufgrund der vorausgesetzten Unabhängigkeit gilt

$$\mathbb{P}[\max_{\sigma \in S_N} X_\sigma \leq u] = (1 - \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-x^2/2} dx)^{2^N}.$$

Offenbar muss das Integral rechts von der Ordnung 2^{-N} sein, damit man etwas bekommt, was weder trivial gegen 0 noch trivial gegen 1 konvergiert. Wir setzen $u_n(x)$ für $x > -\frac{\log N}{\log 2}$ fest durch

$$\frac{2^N}{\sqrt{2\pi}} \int_{u_n(x)}^{\infty} e^{-z^2/2} dz = e^{-x}.$$

Man kann das obige Lemma verwenden, um zu zeigen

$$u_n(x) = \sqrt{2N \log 2} + \frac{x}{\sqrt{2N \log 2}} - \frac{\log(N \log 2) + \log 4\pi}{2\sqrt{2N \log 2}}.$$

Dies ist eine Übung (man verwendet, dass das Lemma im wesentlichen besagt, dass

$$\frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-z^2/2} dz \equiv \frac{1}{u\sqrt{2\pi}} e^{-u^2/2}).$$

Somit erhalten wir für solche x

$$\mathbb{P}[\max_{\sigma \in S_N} X_\sigma \leq u_N(x)] = (1 - 2^{-N} e^{-x})^{2N}.$$

Somit konvergiert die Zufallsvariable

$$Y_N = u_N^{-1}(\max_{\sigma \in S_N} X_\sigma)$$

gegen eine doppelt exponentielle Verteilung. Somit konvergiert auch

$$\mathbb{P}[\max_{\sigma \in S_N} \frac{1}{\sqrt{N}} X_\sigma \leq \sqrt{2 \log 2} + \frac{x}{N\sqrt{2 \log 2}} - \frac{\log(N \log 2) + \log 4\pi}{2N\sqrt{2 \log 2}}] \to e^{-e^{-x}}.$$

Da aber alle Summanden auf der rechten Seite des Eriegnisses bis auf $\sqrt{2 \log 2}$ gegen 0 gehen, bedeutet dies die Konvergenz von $\max_\sigma \frac{1}{\sqrt{N}} X_\sigma$ gegen $\sqrt{2 \log 2}$ in Wahrscheinlichkeit. Aufgrund der exponentiellen Struktur der doppelten Exponentialverteilung folgt sogar die fast sichere Konvergenz gegen 0 mit Hilfe des Borel-Cantelli-Lemmas. Schließlich folgt aus derselben Struktur die L^1-Konvergenz, wie man leicht überprüft. □

Wir werden als nächstes die Partitions-Funktion bzw. die freie Energie des Modells untersuchen. Die Zustandssumme bzw. Partitionsfunktion ist wie immer gegeben durch

$$Z_{N,\beta} = \sum_{\sigma \in S_N} e^{-\beta H_N(\sigma)} = \sum_{\sigma \in S_N} e^{\beta \sqrt{N} X_\sigma}.$$

Wir führen die Größe

$$\Phi_{N,\beta} := \frac{1}{N} \log 2^{-N} Z_{N,\beta}$$

ein, die bis auf triviale Faktoren (die nichts an der Differenzierbarkeit der Funktion ändern) mit der freien Energie übereinstimmt. Wir hoffen, $\Phi_{\beta,N}$ im Limes $N \to \infty$ berechnen zu können. Ähnlich wie in der vorangegangenen Diskussion über das SK-Modell kann man hoffen, dass

$$Z_{N,\beta} \sim \mathbb{E}_J Z_{N,\beta}$$

gilt (wobei wir nicht weiter spezifizieren wollen, was "∼" bedeutet). In diesem Falle ließen sich $\Phi_{\beta,N}$ leicht berechnen und wäre gleich der "annealten" freien Energie.

Lemma 11.3 *Für die annealte freie Energie im REM gilt*

$$f_\beta^{an} := \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_J 2^{-N} Z_{N,\beta} = \beta^2/2.$$

Beweis: Der Beweis verläuft völlig analog zum Fall des SK-Modells und ist dem Leser als Übung überlassen. □

Natürlich kann man auch wieder vermuten, dass

$$\lim_{N\to\infty} \Phi_{N,\beta} \neq \beta^2/2,$$

da sich Logarithmen bekannterweise nicht schmerzlos durch Erwartungswerte ziehen lassen. Die Wahrheit liegt, wie der folgende Satz besagt, irgendwo in der Mitte:

$$\lim_{N\to\infty} \Phi_{N,\beta} = \beta^2/2$$

ist ein echtes Hochtemperaturphänomen, das bei genügend tiefen Temperaturen zusammenbricht.

Satz 11.4 *Im REM gilt für die freie Energie das folgende*

$$\lim_{N\to\infty} \mathbb{E}\Phi_{N,\beta} = \begin{cases} \beta^2/2 & \textit{für } \beta \le \beta_c \\ \frac{\beta_c^2}{2} + (\beta-\beta_c)\beta_c & \textit{für } \beta > \beta_c \end{cases}.$$

Hierbei ist $\beta_c = \sqrt{2\log 2}$.

Beweis: Von mehreren alternativen Beweisen geben wir hier denjenigen, der Ideen verwendet, die auch in der neueren Spinglasliteratur wieder auftauchen. Wir beginnen mit der schon erwähnten Jensenschen Ungleichung, derzufolge

$$\mathbb{E}_J \log Z_{N,\beta} \le \log \mathbb{E}_J Z_{N,\beta},$$

womit wir

$$\mathbb{E}_J \Phi_{N,\beta} \le \beta^2/2$$

für alle N erhalten (hierzu bemerke man, dass der Wert von f_β^q im vorhergehenden Lemma schon für jedes N angenommen wird). Für tiefe Temperaturen können wir aber eine noch bessere Schranke finden, wenn wir bemerken, dass $\mathbb{E}\Phi_{\beta,N}$ nicht quadratisch wächst. Dazu berechnen wir nach dem Satz von der majorisierten Konvergenz

$$\frac{d}{d\beta}\mathbb{E}_J \Phi_{N,\beta} = \mathbb{E}_J \frac{d}{d\beta}\Phi_{N,\beta}.$$

$$
\begin{aligned}
\mathbb{E}_J \frac{d}{d\beta}\Phi_{\beta,N} &= \mathbb{E}_J \frac{d}{d\beta}\frac{1}{N}\log\frac{1}{2^N}\sum_{\sigma\in S_N} e^{\beta\sqrt{N}X_\sigma} \\
&= \frac{\sqrt{N}}{N}\mathbb{E}_J \frac{\frac{1}{2^N}\sum_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma}}{\frac{1}{2^N}\sum_\sigma e^{\beta\sqrt{N}X_\sigma}} \\
&= \frac{1}{\sqrt{N}}\mathbb{E}_J \frac{\mathbb{E}_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma}}{2^{-N}Z_{N,\beta}} \\
&\leq \frac{1}{\sqrt{N}}\mathbb{E}_J \frac{\max_\sigma X_\sigma \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}}{2^{-N}Z_{N,\beta}} \\
&= \frac{1}{\sqrt{N}}\mathbb{E}_J \max_{\sigma\in S_N} X_\sigma \leq \sqrt{2\log 2}(1+\frac{c}{N}),
\end{aligned}
$$

wobei man für die letzte Abschätzung Proposition 11.2 benutzt und c eine Konstante ist. Damit wird klar, dass $\beta^2/2$ für große β eine zu große Schranke für

$$
\lim_{N\to\infty} \mathbb{E}_J \Phi_{N,\beta}
$$

ist, denn die Funktion hat eine maximale Ableitung von $\sqrt{2\log 2}$ (wir ignorieren von nun an alle Korrekturen, die im Limes sowieso verschwinden). Es ist auch klar, dass diese Information erst ab

$$
\beta \geq \beta_c := \sqrt{2\log 2}
$$

effektiv wird, denn für kleinere Werte von β ist die Ableitung von $\beta^2/2$ sowieso kleiner als $\sqrt{2\log 2}$. Somit erhalten wir die obere Schranke

$$
\lim_{N\to\infty} \mathbb{E}_J \Phi_{\beta,N} \leq \begin{cases} \beta^2/2 & \beta \leq \beta_c \\ \beta_c^2/2 + (\beta-\beta_c)\sqrt{2\log 2} & \beta > \beta_c \end{cases}.
$$

Es bleibt die untere Schranke zu beweisen. Dazu stellen wir fest:

$$
\begin{aligned}
\frac{d^2}{d\beta^2}\Phi_{\beta,N} &= \frac{1}{\sqrt{N}}\frac{d}{d\beta}\frac{\mathbb{E}_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma}}{2^{-N}Z_{N,\beta}} \\
&= \frac{\mathbb{E}_\sigma X_\sigma^2 e^{\beta\sqrt{N}X_\sigma} 2^{-N}Z_{N,\beta} - (\mathbb{E}_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma})^2}{2^{-2N}Z_{N,\beta}^2}
\end{aligned}
$$

Betrachtet man das zum REM gehörige Gibbs-Maß

$$
\mu_{N,\beta}(\sigma) = \frac{e^{\beta\sqrt{N}X_\sigma}}{Z_{N,\beta}},
$$

so ergibt sich

$$
\frac{d^2}{d\beta^2}\Phi_{N,\beta} = \mathbb{V}_{\mu_{N,\beta}}(X_\sigma) \geq 0.
$$

Das Wichtige an dieser Feststellung ist, dass dies bedeutet, dass $\frac{d}{d\beta}\Phi_{N,\beta}$ und somit auch $\frac{d}{d\beta}\mathbb{E}_J\Phi_{N,\beta}$ nicht fällt. Dies wiederum hat zur Konsequenz, dass wir die Konvergenz

$$
\mathbb{E}\Phi_{N,\beta} \to \beta^2/2
$$

nur im Hochtemperaturbereich $\beta < \beta_c$ zeigen müssen. Für $\beta \geq \beta_c$ gilt die Behauptung dann automatisch, da die obere Schranke schon die kleinstmögliche Art ist, wie $\mathbb{E}\Phi_{N,\beta}$ wachsen kann. Darüber hinaus haben wir festgestellt, dass $\beta^2/2$ gerade gleich $\frac{1}{N}\log 2^{-N}\mathbb{E}Z_{N,\beta}$ ist. Da $Z_{N,\beta}$ eine Summe von i.i.d. Zufallsvariablen ist (aber abhängig von N), sollte man das mit der üblichen Technik, z. B. einer Chebyshev-Ungleichung, bewerkstelligen können. Hierzu benötigen wir die Varianz von $Z_{N,\beta}$ und dafür das zweite Moment. Wenn wir dies ohne weitere Bedenken berechnen, erhalten wir

$$\begin{aligned}
2^{-2N}\mathbb{E}_J Z_{N,\beta}^2 &= \mathbb{E}_J(\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma})^2 \\
&= \mathbb{E}_\sigma\mathbb{E}_{\sigma'}\mathbb{E}_J e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})} \\
&= 2^{-2N}(\sum_{\sigma\neq\sigma'}(\mathbb{E}_J(e^{\beta\sqrt{N}X_\sigma}))^2 + \sum_\sigma \mathbb{E}_J e^{2\beta\sqrt{N}X_\sigma}),
\end{aligned}$$

wobei wir die Unabhängigkeit der X_σ verwendet haben. Nach den Rechenregeln für Gaußsche Integrale erhalten wir

$$\begin{aligned}
2^{-2N}\mathbb{E}_J Z_{N,\beta}^2 &= 2^{-2N}[\sum_{\sigma\neq\sigma'} e^{N\beta^2} + \sum_\sigma e^{2N\beta^2}] \\
&= e^{N\beta^2}[(1-2^{-2N}) + 2^{-N}e^{N\beta^2}],
\end{aligned}$$

da es gerade 2^N Indizes σ gibt und $2^{2N}-2^N$ Paare $\sigma\neq\sigma'$. Ist $\beta < \sqrt{\log 2}$, so ist der zweite Summand in der Klammer exponentiell klein und wir können folgendermaßen abschätzen:

$$\begin{aligned}
&\mathbb{P}[|\log\frac{Z_{N,\beta}}{\mathbb{E}_j Z_{N,\beta}}| > \varepsilon N] \\
=\ &\mathbb{P}[\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} < e^{-\varepsilon N} \quad \text{oder} \quad \frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} > e^{\varepsilon N}] \\
\leq\ &\mathbb{P}[(\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} - 1)^2 > (1-e^{-\varepsilon N})^2] \\
\leq\ &\mathbb{E}_J(\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} - 1)^2/(1-e^{-\varepsilon N})^2
\end{aligned}$$

nach der Markov-Ungleichung. Die letzte Zeile ist nun gleich

$$\frac{\mathbb{E}_J Z_{N,\beta}^2}{(\mathbb{E}_J Z_{N,\beta})^2} - 1/(1-e^{-\varepsilon N})^2 \leq \frac{2^{-N} + 2^{-N}e^{N\beta^2}}{(1-e^{-\varepsilon N})^2}$$

nach der vorhergehenden Rechnung, da $(\mathbb{E}Z_{N,\beta})^2 = e^{N\beta^2}$. Für $\beta < \sqrt{\log 2}$ konvergiert dies sogar exponentiell schnell gegen 0. Der einzige Nachteil dieses Verfahrens besteht darin, dass wir die behauptete Konvergenz nicht nur für $\beta^2 < \log 2$ sondern für $\beta < 2\log 2$ beweisen müssen. Für $\beta^2 \geq \log 2$ explodiert aber der Varianzton. Der Ursprung dieses Problems ist nicht ganz leicht zu erkennen. Er liegt in der Tatsache verborgen, dass wir für die Berechnung von $2^{-2N}\mathbb{E}_J Z_{N,\beta}^2$ insbesondere

$$\mathbb{E}_J e^{2\sqrt{N}\beta X_\sigma}$$

berechnen müssen (und dies mit dem Resultat $e^{2N\beta^2}$ auch getan haben). Dies aber bedeutet – wenn man das Maximum des Integrals betrachtet – dass die Hauptmasse des Integrals von dem Bereich

$$X_\sigma \sim 2\beta\sqrt{N}$$

stammt. Umgekehrt stammt der größte Anteil von $\mathbb{E}Z_{N,\beta}$ (mit dessen Quadrat ja verglichen wird) von Termen

$$X_\sigma \sim \beta\sqrt{N}.$$

Man bekommt den Verdacht, dass man die Explosion der Varianz verhindert kann, ohne etwas an durchschnittlichen Verhalten zu ändern, wenn man die X_σ passend abschneidet. Sei $c \geq 0$ und wir betrachten

$$\tilde{Z}_{N,\beta}(c) := \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma < c\sqrt{N}}.$$

Wir wollen sehen, wieviel Masse wir verlieren: Es ist klar, dass sich $\mathbb{E}_J\tilde{Z}_{N,\beta}(c)$ von $2^{-N}\mathbb{E}_J Z_{N,\beta}$ nur um

$$\int_{c\sqrt{N}}^\infty \frac{1}{\sqrt{2\pi}}e^{-y^2/2}e^{\beta\sqrt{N}y}dy$$

unterscheidet. Verwendet man die Abschätzungen aus dem vorhergehenden Lemma für die Normalverteiung, sieht man, dass für $\beta < c$ gilt

$$\begin{aligned}
\int_{c\sqrt{N}}^\infty \frac{1}{\sqrt{2\pi}}e^{-y^2/2}e^{\beta\sqrt{N}y}dy &= e^{\beta^2N/2}\int_{c\sqrt{N}}^{\int} \frac{1}{\sqrt{2\pi}}e^{-1/2(y-\beta\sqrt{N})^2}dy \\
&= e^{\beta^2N/2}\int_{(c-\beta)\sqrt{N}}^\infty \frac{1}{\sqrt{2\pi}}e^{-1/2z^2}dz \\
&= e^{\beta^2N/2}e^{-(c-\beta)^2N/2}\frac{1}{\sqrt{2\pi N}(c-\beta)}(1+\mathcal{O}(\frac{1}{N})).
\end{aligned}$$

Ergo

$$\mathbb{E}_J\tilde{Z}_{N,\beta}(c) = e^{N\beta^2/2}(1 - \frac{e^{-N(c-\beta)^2/2}}{\sqrt{2\pi N}(c-\beta)}(1+\mathcal{O}(\frac{1}{N}))).$$

Somit ändert das Abschneiden im wesentlichen nichts am Erwartungswert. Berechnen wir aber $\mathbb{E}_J\tilde{Z}^2_{N,\beta}$, so ergibt sich:

$$\mathbb{E}_J\tilde{Z}^2_{N,\beta}(c) = (1-2^{-N})(\mathbb{E}_J\tilde{Z}_{N,\beta}(c))^2 + 2^{-N}\mathbb{E}e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma < c\sqrt{N}}.$$

Führt man die obige Rechnung erneut durch, so erhält man für den zweiten Summanden

$$\mathbb{E}_J e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma<c\sqrt{N}} = \begin{cases} e^{2\beta^2N}, & \text{falls } 2\beta < c \\ \frac{1}{2^N}\frac{e^{2c\beta N - \frac{1}{2}c^2N}}{(2\beta-c)\sqrt{2\pi N}}, & \text{sonst} \end{cases}$$

(bis auf Terme vernachlässigbarer Ordnung). Somit erhalten wir (erneut bis auf vernachlässigbare Terme) für

$$c/2 < \beta < c$$

die Gleichung

$$\frac{2^{-N}\mathbb{E}_j e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma<c\sqrt{N}}}{(\mathbb{E}_J\tilde{Z}_{N,\beta})^2} = \frac{e^{-\beta^2N+2cN-c^2/2N}2^{-2N}}{(2\beta-c)\sqrt{2\pi N}}$$

(letzteres wieder bis auf vernachlässigbare Faktoren). Die rechte Seite ist nun gleich

$$e^{-N(c-\beta)^2-N(2\log 2-c^2)/2}.$$

Wählt man daher

$$\beta < c < \sqrt{2\log 2},$$

bekommt man

$$\mathbb{E}_J \left[\frac{\tilde{Z}_\beta - \mathbb{E}_J \tilde{Z}_{N,\beta}}{\mathbb{E}_J \tilde{Z}_{N,\beta}} \right]^2 \leq e^{-N\kappa(c,\beta)}$$

für ein

$$\kappa(c,b) > 0.$$

Also erhält man mit Chebyshev

$$\mathbb{P}[|\tilde{Z}_{N,\beta}(c) - \mathbb{E}_J \tilde{Z}_{N,\beta}(c)| > \delta \mathbb{E} \tilde{Z}_{N,\beta}(c)] \leq \frac{1}{\delta^2} e^{-N\kappa(c,\beta)}.$$

Insbesondere sieht man also, dass

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}_J \log \tilde{Z}_{N,\beta}(c) = \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_J Z_{N,\beta}(c)$$

für jede Wahl von c mit

$$\beta < c < \sqrt{2\log 2} = \beta^c.$$

Da aber

$$Z_{N,\beta} \geq Z_{N,\beta}(c)$$

für alle c gilt, folgt daraus bei geeigneter Wahl von c die Ungleichung

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_J Z_{N,\beta} \geq \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}_J \tilde{Z}_{N,\beta}(c) = \beta^2/2.$$

Dies ergibt die fehlende Ungleichung (die andere Richtung hatten wir mit Hilfe der Jensenschen Ungleichung hergeleitet) und beweist den Satz. $\square$

11.2 Fluktuationen

Auch hier kann man als Wahrscheinlichkeitstheoretiker quasi reflexsartig nach dem Fluktuationsverhalten fragen: Schließlich wissen wir ja schon aus anderen Modellen wie dem Curie-Weiss-Modell, dass Phasenübergänge oft auch in dem Verhalten der zentralen Grenzwertsätze widergespiegelt werden. Das Ergebnis ist einigermaßen überraschend. Als Vorbereitung benötigen wir aber noch ein wenig Kenntnis über Punktprozesse.

Grob gesprochen ist ein Punktprozess so etwas wie eine zufällige Verteilung von Punkten im Raum. Nennen wir diesen für den Augenblick E und nehmen an, dass er lokal kompakt Hausdorffsch mit abzählbarer Basis ist. (Typisch wird $E = \mathbb{R}^d$ sein.) Sei $\mathcal{E}$ die Borel σ-Algebra auf E. Für $x \in E$ sei ε_x das Dirac-Maß auf E mit Masse in x:

$$\varepsilon_x = \begin{cases} 1 & x \in A \\ & & A \in \mathcal{E} \quad . \\ 0 & x \in A \end{cases}$$

Ein Punkt-Maß ist dann eine Konfiguration von Punkten (die wir mit ihrem Dirac-Maßen identifizieren), die lokal "nicht zu sehr klumpt". Formaler: Seien $\{x_i,\, i \geq 1\}$ Punkte in E (nicht notwendig verschieden). Dann ist

$$m = \sum_{i=1}^{n} \varepsilon_{x_i}$$

ein Punkt-Maß, falls

$$m(K) < +\infty \quad \text{für alle} \quad K \in \mathcal{E} \quad \text{kompakt}$$

(d. h. m ist ein Radon-Mass). Die Menge

$$S_m = \{x \in E : m(\{x\}) \neq 0\}$$

ist der Träger von m. Ein Punkt-Prozess ist nun ein zufälliges Punkt-Maß. Genauer: Sei

$$M_p = \{m \in \mathcal{M}(E) : m \text{ ist ein Punktmaß}\}$$

und sei $\mathcal{M}_p(E)$ über $M_p(E)$ die kleinste σ-Algebra, so dass die Auswertungen $F \mapsto m(F)$ allesamt messbar sind für $F \in \mathcal{E}$. Genauer sei $\mathcal{M}_p(E)$ die kleinste σ-Algebra, die alle Mengen der Form

$$\{m \in M_p(E) : m(F) \in B\}$$

umfasst. Hierbei seien $f \in \mathcal{E}$ und $B \in \mathcal{B}^1$ beliebig.

Definition 11.5 *Ein* Punkt-Prozess *ist eine messbare Abbildung*

$$N : (\Omega, \mathcal{A}, \mathbb{P}) \to (M_p(E), \mathcal{M}_p(E))$$

(also ein $M_p(E)$-wertige Zufallsvariable) auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

$$\mathbb{P}_N = \mathbb{P} \circ N^{-1}$$

ist die Verteilung der Punktprozesse.

Eine wichtige Klasse von Punkt-Prozessen sind die Poisson-Punkt-Prozesse.

Definition 11.6 *Gegeben sei ein Radon-Maß μ auf $(E, \mathcal{E})$. Ein Punkt-Prozess N heißt* Poissonscher Punktprozess *mit Integrationsmaß μ, falls für alle $k \in \mathbb{N}_0$ gilt*

$$\mathbb{P}[N(F) = k] = \begin{cases} \frac{e^{-\mu(F)}(\mu(F))^k}{k}, & \textit{falls } \mu(F) < \infty \\ 0, & \textit{falls } \mu(F) = 0 \end{cases}$$

und für alle $L \in \mathbb{N}$ und alle paarweise disjunkten $F_1, \ldots, F_L \in \mathcal{E}$ sind

$$N(F_i), \quad i \leq L,$$

unabhängige Zufallsvariablen.

Der Poissonsche Punktprozess auf $\mathbb{R}$ mit Intensitätsmaß $e^{-x}dx$ $(= \mu(dx))$ spielt im folgenden Satz eine zentrale Rolle:

Satz 11.7 *(Bovier, Kourkova, Löwe) (2002)): Die Zustandssumme $Z_{N,\beta}$ im REM hat die folgenden Fluktuationen:*

1. *Für $\beta < \sqrt{\log 2/2}$ gilt*
$$e^{N/2(\log 2-\beta^2)} \log \frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

2. *Für $\beta = \sqrt{\log 2/2}$ gilt*
$$e^{N/2(\log 2-\beta^2)} \log \frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\frac{1}{2}).$$

3. *Sei $\alpha := \frac{\beta}{\sqrt{2\log 2}}$. Für $\sqrt{\frac{\log 2}{2}} < \beta < \sqrt{2\log 2}$ gilt dann*
$$e^{N/2(\sqrt{2\log 2}-\beta)^2+\alpha/2[\log(N\log 2)+\log 4\pi]} \log \frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} \to \int_{-\infty}^{\infty} e^{\alpha z}(\mathcal{P}(dz) - e^{-z}dz).$$
Dabei ist $\mathcal{P}$ der Poissonsche Punkt-Prozess auf $\mathbb{R}$ mit Intensitätsmaß $\mu(dx) = e^{-x}dx$.

4. *Falls $\beta = \sqrt{2\log 2}$, so gilt*
$$e^{\frac{1}{2}[\log(N\log 2)+\log 4\pi]}\left(\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} - \frac{1}{2} + \frac{\log(n\log 2)+\log 4\pi}{4\sqrt{\pi N\log 2}}\right)$$
$$\xrightarrow{\mathcal{D}} \int_{-\infty}^{0} e^{z}(\mathcal{P}(dz) - e^{-z}dz) + \int_{0}^{\infty} e^{z}\mathcal{P}(dz).$$

5. *Falls $\beta > \sqrt{2\log 2}$, so gilt*
$$e^{-N[\beta\sqrt{2\log 2}-\log 2]+\frac{\alpha}{2}[\log(N\log 2)+\log 4\pi]} Z_{N,\beta} \underset{\mathcal{D}}{\longrightarrow} \int_{-\infty}^{\infty} e^{dz}\mathcal{P}(dz),$$
und
$$N(\Phi_{N,\beta} - \mathbb{E}\Phi_{N,\beta}) \xrightarrow{\mathcal{D}} \log \int_{-\infty}^{+\infty} e^{\alpha z}\mathcal{P}(dz) - \mathbb{E}\log \int_{-\infty}^{\infty} e^{\alpha z}\mathcal{P}(dz).$$

Bemerkung: Satz 11.7 zeigt einen zweiten Phasenübergang in REM, der nur auf der Ebene der CLTs sichtbar ist: Bei $\beta = \sqrt{\frac{\log 2}{2}}$ "springt" die Grenzverteilung gleich zweimal. Der Phasenübergang bei $\beta = \sqrt{2\log 2}$ bleibt auf dieser Ebene weiterhin sichtbar.

Dieses Resultat lässt sich auf andere Klassen von Zufallsvariablen X_σ ausdehnen (siehe die Arbeit von Ben Arous, Bogachev und Molchanov [1] aus dem Jahr 2004).

Beweis: Wir gehen schrittweise vor:
Zu 1: Man beachte, dass $Z_{N,\beta}$ eine Summe von i.i.d. Zufallsvariablen ist, so dass man leicht den Verdacht bekommt, dass die Behauptung aus einem der gängigen Zentralen Grenzwertsätze folgt. Dies ist in der Tat der Fall. Wir schreiben

$$\log \frac{Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}} = \log(1 + \frac{Z_{N,\beta} - \mathbb{E}Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}}).$$

Wir entwickeln den Logarithmus

$$\log(1+x) = x + \mathcal{O}(x^2)$$

und stellen fest, dass der 2. Ordnungsterm auf unserer Skala in Verteilung gegen 0 konvergiert (Übung). Somit bleibt, einen CLT für eine richtig skalierte Version von

$$\frac{Z_{N,\beta} - \mathbb{E}Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}}$$

zu beweisen. Da wir $\mathbb{E}Z_{N,\beta}$ berechnen können, folgt

$$\frac{Z_{N,\beta} - \mathbb{E}Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}} = 2^{-N} \sum_{\sigma \in \{-1,+1\}^N} (e^{\beta\sqrt{N}X_\sigma - N\beta^2/2} - 1).$$

Um wieder auf Summen (zentrierter) Zufallsvariablen zu kommen, setzen wir

$$Y_N(\sigma) = \frac{e^{\beta\sqrt{N}X_\sigma - N\beta^2/2} - 1}{e^{N\beta^2/2}\sqrt{1 - e^{-N\beta^2}}}.$$

Dann gilt

$$\mathbb{E}Y_N = 0 \quad \text{und} \quad \mathbb{V}Y_N = 1$$

und

$$\frac{Z_{N,\beta} - \mathbb{E}Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}} = e^{-\frac{N}{2}(\log 2 - \beta^2)}\sqrt{1 - e^{-N\beta^2}} \frac{1}{2^{N/2}} \sum_{\sigma \in \{\pm 1\}^N} Y_N(\sigma).$$

Da der Wurzelterm gegen 1 konvergiert, bekommen wir unsere Behauptung, wenn wir

$$\frac{1}{2^{N/2}} \sum_{\sigma \in \{\pm 1\}^N} Y_N(\sigma) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

zeigen können. Dies folgt sofort aus den CLT für Dreiecksschemata, falls die Lindeberg-Bedingung gilt, d. h. falls für alle $\varepsilon > 0$

$$\lim_{N \to \infty} \mathbb{E}Y_N^2(\sigma) \mathbb{1}_{\{|Y_N(\sigma)| \geq \varepsilon 2^{N/2}\}} = 0$$

gilt. Dies rechnen wir nach:

$$\mathbb{E}Y_N^2 \mathbb{1}_{\{|_N(\sigma)| \geq \varepsilon 2^{N/2}\}} = \frac{e^{-2N\beta^2}}{\sqrt{2\pi}(1 - e^{-N\beta^2})} \int_{\sqrt{N}(\frac{\log 2}{2\beta} + \beta) + \frac{\log \varepsilon}{\sqrt{N}\beta} + 0(\frac{1}{\sqrt{N}})} e^{2\sqrt{N}\beta z - z^2/2} dz + o(1)$$

$$= \frac{1}{\sqrt{2\pi}(1 - e^{-N\beta^2})} \int_{\sqrt{N}(\frac{\log 2}{2\beta} - \beta) + \frac{\log \varepsilon}{\sqrt{N}\beta} + 0(\frac{1}{\sqrt{N}})} e^{-z^2/2} dz$$

nach der üblichen Substitution (man beachte die Veränderung der Integrationsgrenze).

Nach den (inzwischen bekannten) Abschätzungen für Gaußsche Integrale ist dies von der Ordnung

$$\text{Const.}\ \frac{1}{\sqrt{N}(\frac{\log 2}{2\beta}-\beta)+\frac{\log\varepsilon}{\sqrt{N\beta}}}e^{-\frac{N}{2}(\frac{\log 2}{2\beta}-\beta)}.$$

Es ist unmittelbar klar, dass dies dann und nur dann verschwindet, wenn

$$\frac{\log 2}{2\beta}>\beta, \quad \text{d. h.} \quad \beta^2<\frac{\log 2}{2}$$

gilt. Für diesen Bereich haben wir damit den gewünschten CLT bewiesen.

2. und 3: Bedenkt man, dass die Feller-Bedingung zusammen mit dem CLT äquivalent ist zur Lindeberg-Bedingung, und dass überdies die Feller-Bedingung trivialerweise für jede Folge von i.i.d. Zufallsvariablen gilt, so sieht man, dass die vorhergehende Berechnung nicht nur einen CLT für $\beta^2<\frac{\log 2}{2}$ zeigt, sondern auch, dass selbiger für $\beta^2\geq\frac{\log 2}{2}$ nicht gelten kann. Da wir mit $Z_{N,\beta}$ nach wie vor eine Summe von i.i.d. Zufallsvariablen vor uns haben, kann dieses Zusammenbrechen des CLT nur auf eine zu große Wahrscheinlichkeit für Ausreißer zurückzuführen sein (gerade diese versucht man ja mit der Lindeberg-Bedingung zu kontrollieren).

In diesem Sinne ist das folgende zentrale Resultat der Extremwerttheorie sehr interessant:

Satz 11.8 *Sei $\mathcal{P}_N$ der folgende Punktprozess auf $\mathbb{R}$:*

$$\mathcal{P}_N:=\sum_{\sigma\in\{-1,+1\}^N}\varepsilon_{u_N^{-1}}(X_\sigma).$$

Hierbei ist u_N (wie schon vorher) definiert als

$$u_N(x)=\sqrt{2N\log 2}+\frac{x}{\sqrt{2N\log 2}}-\frac{\log(N\log 2)+\log 4\pi}{2\sqrt{2N\log 2}}$$

und $u_N^{-1}(\cdot)$ ist die Inverse von u_N (die natürlich existiert). Dann konvergiert $\mathcal{P}_N$ schwach gegen einen Poissonschen Punkt-Prozess auf $\mathbb{R}$ mit Intensitätsmaß e^{-x}.

Ein Beweis findet sich in dem Buch von Resnick [4].

Bemerkung: Wenn man die Funktionsweise der Funktion u_n^{-1} genauer betrachtet, so wird klar, dass $\mathcal{P}_N$ um das fast sichere Maximum $\sqrt{N2\log 2}$ der X_σ in den Prozess "hineinzoomt" und sich dort die Verteilung der Punkte anschaut. Dabei werden die meisten Punkte gegen $-\infty$ gedrängt (sie sind eben nicht annähernd Maxima), der Rest konvergiert gegen den beschriebenen Poissonprozess. Entsprechend definieren wir

$$Z_{N,\beta}^x:=\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}\mathbb{1}_{\{X_\sigma\leq u_N(x)\}}$$

und schreiben

$$\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}}=1+\frac{Z_{N,\beta}^x-\mathbb{E}_J Z_{N,\beta}^x}{\mathbb{E}_J Z_{N,\beta}}+\frac{Z_{N,\beta}-Z_{N,\beta}^x-\mathbb{E}_J[Z_{N,\beta}-Z_{N,\beta}^x]}{\mathbb{E}_J Z_{N,\beta}}.$$

Wir betrachten den letzten Summanden zuerst. Sei

$$\begin{aligned} W_N(x) := \frac{Z_{N,\beta} - Z^x_{N,\beta}}{\mathbb{E}Z_{N,\beta}} &= e^{-N(\log 2+\beta^2/2)} \sum_{\sigma\in\{\pm1\}^N} e^{\beta\sqrt{N}X_\sigma} 1\!\!1_{\{X_\sigma > u_n(x)\}} \\ &= e^{-N(\log 2+\beta^2/2)} \sum_{\sigma\in\{\pm1\}^N} e^{\beta\sqrt{N}u_n(u_N^{-1}(X_\sigma))} 1\!\!1_{\{X_\sigma > u_n(x)\}} \\ &= \frac{1}{C(N,\beta)} \sum_{\sigma\in\{\pm1\}^N} e^{\alpha u_N^{-1}(X_\sigma)} 1\!\!1_{\{u_n^{-1}(X_\sigma) > x\}}. \end{aligned}$$

Hierbei ist

$$C(N,\beta) := e^{\frac{N}{2}(\sqrt{2\log 2}-\beta)^2 + \frac{\alpha}{2}[\log(N\log 2)+\log 4\pi]}$$

und

$$\alpha = \frac{\beta}{\sqrt{2\log 2}},$$

und wir haben die explizite Struktur von $u_n(\cdot)$ verwendet. Verwenden wir nun Satz 11.8, d. h.

$$\mathcal{P}_N \Rightarrow \mathcal{P},$$

so folgt die Konvergenz in Verteilung von

$$\frac{1}{C(N,\beta)} \sum_\sigma e^{\alpha u_N^{-1}(X_\sigma)} 1\!\!1_{\{u_n^{-1}(X_\sigma) > x\}},$$

falls man zeigen kann, dass $e^{\alpha x}$ integrierbar ist (auf $[x,\infty)$ gegen den Poissonschen Punkt-Prozess mit Intensität e^{-x}. Das ist aber problemlos fast sicher der Fall, denn fast sicher hat eine Realisierung eines solchen Prozesses einen maximalen Punkt (mit Hilfe von Borel-Cantelli) und somit ist $e^{\alpha x}$ fast sicher immer integrierbar. Man kann jedoch berechnen, dass der Erwartungswert des Integrals für $\beta \geq \sqrt{2\log 2}$ unendlich wird – man sieht also den Phasenübergang in den Tieftemperaturbereich. Wir haben also gesehen, dass das folgende Lemma gilt:

Lemma 11.9 *Für $W_N(x)$ definiert wie oben gilt*

$$C(\beta,N)W_N(x) \xrightarrow{\mathcal{D}} \int_x^\infty e^{\alpha z}\mathcal{P}(dz),$$

wobei $\mathcal{P}(\cdot)$ den Poissonschen Punktprozess mit Intensitätsmaß $e^{-z}dz$ beschreibt.

Das Folgende lässt sich über das Integral auf der rechten Seite aussagen:

Übung: Man zeige, dass
(i) $\mathbb{V}(\int_x^\infty e^{\alpha z}\mathcal{P}(dz)) < +\infty$, falls $\beta^2 < \frac{\log 2}{2}$, d. h. falls der CLT gilt.
(ii) $\lim_{x\downarrow -\infty} \mathbb{E}\int_x^\infty e^{\alpha z}\mathcal{P}(dz) = \infty$.

Wir betrachten nun den anderen Summanden:

$$\frac{Z^x_{N,\beta} - \mathbb{E}Z^x_{N,\beta}}{\mathbb{E}Z_{N,\beta}} =: V_N(x).$$

Im allgemeinen ist es gar nicht einfach, die Grenzverteilung der $V_N(x)$ zu bestimmen.

Lemma 11.10 *Für $\alpha > 1/2$ und alle $k \geq 2$ gilt*

$$\lim_{N\to\infty} C(N,\beta)^k \mathbb{E}[V_N(x)]^k = \sum_{i=1}^{k} \frac{1}{i!} \sum_{e_1\geq 2,\dots,e_i\geq 2} \frac{k!}{\ell_1!\dots\ell_i!} \frac{e^{(k_\alpha - i)x}}{(\ell_1\alpha-1)\dots(\ell_i\alpha-1)}.$$

Für $\alpha = \frac{1}{2}$ erhalten wir für alle ungeraden k

$$\lim_{N\to\infty} \mathbb{E}[V_N(x)]^k \, e^{kN(\sqrt{2\log 2}-\beta)^2/2} = 0$$

und für alle geraden k

$$\lim_{N\to\infty} \mathbb{E}[V_N(x)]^k \, e^{kN(\sqrt{2\log 2}-\beta)^2/2} = \frac{k!}{(k/2)!2^k} = \frac{(k-1)!!}{2^k/2}$$

(dabei ist $k!! = k(k-2)\dots 1$).

Bemerkung: Im Falle $\alpha = \frac{1}{2}$ erhält man für $V_N(x)$ offenbar die Momente der $\mathcal{N}(0,\frac{1}{2})$-Verteilung. Tatsächlich ist dies genau der Grund für das Auftreten dieser Verteilung im zu beweisenden Satz.

Beweis: Da sich $V_N(x)$ aus $Z^x_{N,\beta}$ zusammensetzt, müssen wir im Beweis des Lemmas die Momente von

$$T_N(\sigma) := 2^{-N} e^{\beta\sqrt{N}X_\sigma} \mathbb{1}_{\{X_\sigma \leq u_n^{-1}(x)\}}$$

berechnen. Dies geht wieder durch quadratische Ergänzung im Exponenten – diesmal in den Teilabschätzungen der Gauß-Variablen. Man erhält die folgende Asymptotik (wobei $a_n \sim b_n$ bedeutet $\frac{a_n}{b_n} \to 1$):

1. Falls $k\beta < \sqrt{2\log 2}$

$$\mathbb{E}T_N^k(\sigma) \sim 2^{-kN} e^{k^2\beta^2 N_k}.$$

2. Falls $k\beta = \sqrt{2\log 2}$

$$\mathbb{E}T_N^k(\sigma) \sim \frac{2^{-kN} e^{k^2\beta^2 N/2}}{2} = \frac{2^{-N} e^{(k_\alpha-1)}}{2} e^{k[(\beta\sqrt{2\log 2}-\log 2)N]}.$$

3. Falls $k\beta > \sqrt{2\log 2}$

$$\mathbb{E}T_N^k(\sigma) \sim \frac{2^{-N} e^{-(k_\alpha-1)}}{(k_\alpha-1)} \cdot e^{k[(N\beta\sqrt{2\log 2}-\log 2)-\frac{\alpha}{2}[\log(N\log 2)]+\log 4\alpha]}.$$

Nun gilt

$$\begin{aligned}
(\mathbb{E}_J Z_{N,\beta})^k \mathbb{E}[V_N(x)]^k &= \mathbb{E}(\sum_\sigma [T_N - \mathbb{E}_J TN(\sigma)])^k \\
&= \sum_{\sigma_1,\dots,\sigma_k} \mathbb{E} \prod_{i=1}^{k} [T_N(\sigma_i) - \mathbb{E}_J TN(\sigma_i)] \\
&= \sum_{i=1}^{k} \sum_{\substack{\ell_1\geq 2,\dots,\ell_i\geq 2 \\ \sum \ell_j = k}} \frac{k!}{\ell_1!\dots\ell_i!} \binom{2^N}{i} \\
&\quad \mathbb{E}_J[T_N(\sigma_1) - \mathbb{E}_J(T_N(\sigma_1))]^{\ell_1} \dots \mathbb{E}[T_N(\sigma_i) - \mathbb{E}(T_N(\sigma_i)]^{\ell_i},
\end{aligned}$$

wobei wir rein kombinatorische Überlegungen verwendet haben und dass

$$T_N(\sigma) - \mathbb{E}T_N(\sigma)$$

zentral ist. Das Wesentliche ist nun, dass für $k\beta \geq \sqrt{2\log 2}$ die zentrierten Momente von T_N im wesentlichen wie die gewöhnlichen Momente von T_N wachsen, d. h. für $\ell \geq 2$ gilt

$$\mathbb{E}_J[T_N - \mathbb{E}_J T_N]^\ell = \sum_{j=0}^{\ell} (-1)^j \binom{\ell}{j} \mathbb{E}_J(T_N(\sigma)^{\ell-j})(\mathbb{E}(T_n(\sigma)))^j \sim \mathbb{E}_J T_n^\ell.$$

In der Tat haben wir für $\sqrt{\frac{\log 2}{2}} \leq \beta < \sqrt{2\log 2}$ $j \neq \ell - 1, \ell$

$$\frac{\mathbb{E}_J T_N^{\ell-j}(\sigma)[\mathbb{E}_J T_N(\sigma)]^j}{\mathbb{E}_J T_N^\ell(\sigma)} = e^{N_j(\beta^2/2 - \beta\sqrt{2\log 2})} 0(N^{\alpha j/2}$$

und für $\ell \geq 2$ $j = \ell - 1,\ \ell$

$$\mathbb{E}_J T_N^{\ell-j}(\sigma)[\mathbb{E}_J T_N(\sigma)]^j = e^{N\ell(\beta^2/2 - \beta\sqrt{2\log 2}) + N\log 2} 0(N^{\alpha\ell/2}) \leq e^{-N\log^2/2} 0(N^{\alpha\ell/2}).$$

Schließlich gilt für $\beta \geq \sqrt{2\log 2},\ \ \ell \geq 2,\ \ j \geq 1$

$$\mathbb{E}T_N^{\ell-j}(\sigma)[\mathbb{E}T_N(\sigma)]^j = 0(2^{-N}).$$

Somit setzen wir für $\ell \geq 2$ und $\beta > \sqrt{\frac{\log 2}{2}}$ bzw. für $\ell \geq 3$ und $\beta = \sqrt{\frac{\log 2}{2}}$

$$\mathbb{E}_J[T_N(\sigma) - \mathbb{E}_J T_N(\sigma)]^\ell \sim \mathbb{E}T_N^\ell(\sigma) \sim \frac{2^{-N} e^{x(l\alpha - 1)}}{k\alpha - 1}[2^{-N} e^{N\beta\sqrt{2\log 2} - \frac{\alpha}{2}[\log(N\log 2) + \log 4\pi]}].$$

Setzt man dies in die Formel für $(\mathbb{E}_J Z_{N,\beta})^k \mathbb{E}V_N^k(x)$ ein, so ergibt sich die Behauptung des Lemmas für $\alpha > \frac{1}{2}$.

Für $\beta = \sqrt{\frac{\log 2}{2}}$ verläuft die Argumentation ein wenig anders. Hier berechnet man für $\ell = 2$

$$\mathbb{E}_J[T_N(\sigma) - \mathbb{E}T_N(\sigma)]^2 \sim \frac{2^{-N} e^{-x}}{2}[2^{-N} e^{N\beta\sqrt{2\log 2}} e^{\alpha x}]^2.$$

Und man überprüft, dass der Term

$$j = k/2\ \ell_1 = \ell_2 = \ldots = \ell_i = 2$$

für den höchsten Beitrag für das Moment

$$(\mathbb{E}_J Z_{N,\beta})^k \mathbb{E}_J V_N^k(x)$$

liefert. Setzt man dies ein, so bekommt man auch die zweite Behauptung des Lemmas für $\alpha = 1/2$. □

Schreiben wir nun $V(x, \alpha)$ für die Zufallsvariable, deren Momente (im Falle $\alpha > 1/2$) durch das vorhergehende Lemma beschrieben sind, so erhalten wir

Lemma 11.11 • *Für* $\sqrt{\frac{\log 2}{2}} < \beta$ *gilt*

$$C(N,\beta)V_N(x) \xrightarrow{\mathcal{D}} V(x,\alpha)$$

(hierbei hat $V(x,\alpha)$ *den Erwartungswert 0).*

- *Für* $\beta = \sqrt{\frac{\log 2}{2}}$ *gilt*

$$e^{\frac{N}{2}(\sqrt{2\log 2}-\beta)^2} V_N(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{2}).$$

Beweis: Dies folgt standardmäßig aus dem vorhergehenden Lemma, wenn die Verteilung rechts eindeutig durch ihre Momente festgelegt ist. Das ist für die Normalverteilung bekannt und kann für die Verteilung von V überprüft werden. □

Schickt man nun im Falle $\beta = \sqrt{\frac{\log 2}{2}}$ x gegen ∞, so verschwindet der Integralterm und man erhält die Behauptung des Satzes für $\beta = \sqrt{\frac{\log 2}{2}}$.

Den dritten Teil bekommt man aus dem folgenden Lemma:

Lemma 11.12 *Für* $\sqrt{\frac{\log 2}{2}} < \beta < \sqrt{2\log 2}$ *und beliebiges* x *gilt*

$$\begin{aligned} &e^{\frac{N}{2}[\sqrt{2\log 2}-\beta]^2+\frac{\alpha}{2}[\log(N\log 2)+\log 4\pi]} \log \frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} \\ &\xrightarrow{\mathcal{D}} V(x,\alpha) + \int_x^\infty e^{\alpha z}\mathcal{P}(dz) - \int_x^\infty e^{\alpha z} e^{-z} dz, \end{aligned}$$

wobei $V(x,\alpha)$ *und* $\mathcal{P}(\cdot)$ *unabhängig sind.*

Beweis: Das wäre unmittelbar aus dem Vorhergehenden klar, wären $V_N(x)$ und $W_N(x)$ unabhängig. Obwohl sie dies nicht sind, sind sie nicht sehr weit davon entfernt. Um dies einzusehen, beobachtet man, dass die Zerlegung

$$\frac{Z_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} = 1 + \frac{Z^x_{N,\beta} - \mathbb{E}_J Z^x_{N,\beta}}{\mathbb{E}_J Z_{N,\beta}} + \frac{Z_{N,\beta} - Z^x_{N,\beta} - \mathbb{E}(Z_{N,\beta} - Z^x_{N,\beta})}{\mathbb{E}_J Z_{N,\beta}}$$

in unabhängige Summanden zerfällt, falls man die Anzahl n_x der X_σ, die größer sind als $u_N(x)$ fixiert. Allerdings kann man das vorhergehende Korollar auch bedingt beweisen (Übung). Somit kann man den Limes auch wie gewünscht bilden. □

Ist nun $\alpha > \frac{1}{2}$, so ist

$$\mathbb{E}V(x,\alpha)^2 = e^{x(2\alpha-1)}/(2\alpha-1);$$

somit konvergiert

$$V(x,\alpha) \xrightarrow{\mathcal{D}} 0,$$

wenn $x \to -\infty$ gilt. Also

$$V(x,\alpha) \stackrel{\mathcal{D}}{=} \lim_{y\to\infty} \int_{-y}^{x} e^{\alpha x} \mathcal{P}(dz) - \int_{-y}^{x} e^{\alpha z} e^{-z} dz,$$

d. h. das Poisson-Integral

$$\int_{-\infty}^{\infty} e^{\alpha z} (\mathcal{P}(dz) - e^{-z} dz)$$

bekommt einen Sinn. Somit ist 3. bewiesen.

4. und 5: Das Gesehene genügt beinahe, um auch 4. und 5. zu beweisen. Mit den obigen Notationen schreiben wir

$$Z_{N,\beta} = Z^x_{N,\beta} + (Z_{N,\beta} - Z^x_{N,\beta}).$$

Rechnen wir den Unterschied zwischen $Z_{N,\beta}$ und $Z^x_{N,\beta}$ aus, so ergibt sich (Übung):

$$Z_{N,\beta} - Z^x_{N,\beta} = e^{N[\beta\sqrt{2\log 2} - \log 2] - \frac{\alpha}{2}[\log(N\log 2) + \log 4\pi]} e^{\alpha u_N^{-1}(X_\sigma)} \times \sum_{\sigma} \mathbb{1}_{\{u_N^{-1}(X_\sigma) > x\}}.$$

Somit konvergiert für jedes $x \in \mathbb{R}$

$$(Z_{N,\beta} - Z^x_{N,\beta}) e^{-N[\beta\sqrt{2\log 2} - \log 2] + \frac{\alpha}{2}[\log(N\log 2) + \log 4\pi]} \xrightarrow{\mathcal{D}} \int_x^{\infty} e^{\alpha z} \mathcal{P}(dz). \tag{11.1}$$

Schauen wir zuerst den Fall $\beta > \sqrt{2\log 2}$ an. Aus den vorhergehenden Berechnungen der Momente der $T_N(\sigma)$ ergibt sich

$$e^{-N[\beta\sqrt{2\log 2} - \log 2] + \frac{\alpha}{2}[\log(N\log 2) + \log 4\pi]} \mathbb{E}_J Z^x_{N,\beta} \sim \frac{2^{-N} e^{x(\alpha-1)}}{\alpha - 1}.$$

Schiebt man x gegen $-\infty$, so konvergiert dies gegen 0, somit konvergiert auf der Skala von (11.1) der Beitrag von $Z^x_{N,\beta}$ gegen 0. Daher folgt 5.

Schließlich betrachten wir den Fall $\beta = \sqrt{2\log 2}$. Geht man vor wie vorher, so erhält man

$$\mathbb{E}_J Z^0_{N,\beta} = \frac{2^N}{\sqrt{2\pi}} \int_{-\infty}^{u_N(0) - \sqrt{N 2\log 2}} e^{-z^2/2} dz = 2^N \left(\frac{1}{2} - \frac{\log(N\log 2) + \log 4\pi}{4\sqrt{N\pi\log 2}} + 0\left(\frac{(\log N)^2}{N}\right)\right).$$

Wir zerlegen diesmal

$$Z_{N,\beta} = Z_{N,\beta} - Z^0_{N,\beta} + \mathbb{E} Z^0_{n,\beta} + (Z^0_{N,\beta} - \mathbb{E} Z^0_{N,\beta}).$$

Mit den Abschätzungen für $\mathbb{E} Z^0_{N,\beta}$ sehen wir, dass

$$\frac{\mathbb{E} Z^0_{N,\beta}}{\mathbb{E} Z_{N,\beta}} \sim \frac{1}{2}.$$

Entlang der Überlegungen im Beweis zu 3. zeigt man zudem

$$\frac{Z_{N,\beta} - Z^0_{N,\beta}}{\mathbb{E} Z_{N,\beta}} = W_N(0) \to 0 \text{ f.s.}$$

(man beachte, dass trotzdem $\mathbb{E}_J W_N(0) = \frac{1}{2}$). Somit sehen wir genau, dass

$$e^{1/2[\log(N\log 2)+\log 4\pi]} W_N(0) \to \int_0^\infty e^z \mathcal{P}(dz)$$

(man beachte, dass die Limesvariable unendlichen Erwartungswert hat, aber fast sicher endlich ist). Schließlich folgt mit obigem Korollar

$$e^{1/2[\log(N\log 2)+\log 4\pi]} \frac{Z^0_{N,\beta} - \mathbb{E}Z^0_{N,\beta}}{\mathbb{E}Z_{N,\beta}} \to V(0,1).$$

Das gleiche Argument führt dazu, dass nun

$$V(0,1) \overset{\mathcal{D}}{=} \int_{-\infty} e^z(\mathcal{P}(dz) - e^{-z}dz).$$

Somit folgen auch die Aussagen 4. und somit der Satz. □

11.3 Das REM-Gibbs-Maß

Wir wollen nun die zum REM gehörigen Gibbs-Maße näher untersuchen. Diese sind gegeben durch

$$\mu_{N,\beta}(\sigma) = \frac{e^{-\beta\sqrt{N}X_\sigma}}{\sum_{\sigma'} e^{-\beta\sqrt{N}X_\sigma}}.$$

Um zu verstehen, was wir überhaupt untersuchen wollen, sollte man sich vor Augen führen, dass es sich bei den $\mu_{N,\beta}$ um ZUFÄLLIGE Maße handelt. Deren schwache Konvergenz wollen wir untersuchen. Hierzu bilden wir den Spinraum $\{\pm 1\}^N$ vermöge

$$r_N(\sigma) := 1 - \sum_{i=1}^N (1-\sigma_i) 2^{-i-1}$$

in dem Intervall $(0,1]$ ab und definieren das Punktmaß $\tilde{\mu}_{N,\beta}$ auf $(0,1]$ durch

$$\tilde{\mu}_{N,\beta} := \sum_{\sigma\in\{\pm 1\}^N} \delta_{r_N(\sigma)} \mu_{N,\beta}(\sigma).$$

Wir werden die Konvergenz der $\mu_{N,\beta}$ mit Hilfe der $\tilde{\mu}_{N,\beta}$ untersuchen. Hierbei kann man ein Bild erwarten, dass im Hochtemperaturbereich dem Verhalten bei $\beta = 0$ entspricht, bei dem alle Spins die gleiche Wahrscheinlichkeit tragen, während im Tieftemperaturbereich das Hauptgewicht auf wenigen Spin-Konfigurationen σ liegen sollte (dieses Bild haben wir bei der Analyse der freien Energie gewonnen). Tatsächlich gilt:

Satz 11.13 *Für $\beta \le \sqrt{2\log 2}$ gilt*

$$\tilde{\mu}_{N,\beta} \to \lambda\!\!\lambda^1|_{[0,1]} \quad \mathbb{P}\text{-}f.s.$$

Bemerkung: Da die unter $\mu_{N,0}$ gleichverteilten σ natürlich auch gegen das Lebesgue-Maß konvergieren, entspricht das Hochtemperatur-Verhalten also unseren Vorstellungen.

Beweis: Es ist zu zeigen, dass für jede endliche Ansammlung von Intervallen $\Delta_1, \ldots, \Delta_k \subset [0,1]$ die Familie von

$$(\tilde{\mu}_{N,\beta}(\Delta_1), \ldots, \tilde{\mu}_{N,\beta}(\Delta_k))$$

gegen

$$(|\Delta_1|, \ldots, |\Delta_k|)$$

konvergiert. Es genügt, dies für $k = 1$ zu zeigen. Sei $n \in \mathbb{N}$ und $N \geq n$. Sei

$$\pi_n := \{+1, -1\}^N \to \{+1, -1\}^n$$

die kanonische Projektion. Wir werden oft

$$\sigma_n := \pi_n \sigma$$

schreiben. Für $\sigma \in \{\pm 1\}^N$ sei

$$a_N(\sigma) := r_n(\pi_n \sigma)$$

und

$$\Delta_n(\sigma) := (a_n(\sigma) - 2^{-n}, a_n(\sigma)].$$

Offenbar überdeckt die Vereinigung dieser Intervalle $(0, 1]$. Die Konstruktion dieser Intervalle erlaubt die folgende Rechnung

$$\tilde{\mu}_{N,\beta}(\Delta_n(\sigma)) = \mu_{N,\beta}(\{\sigma' \in \{\pm 1\}^N : \pi_n(\sigma') = \pi_n(\sigma)\}).$$

Zunächst zeigen wir, dass die Masse der $\Delta_n(\sigma)$ gut durch die Lebesgue-Verteilung approximierbar sind.

Lemma 11.14 *Sei $\beta' := \sqrt{\frac{N}{N-n}}\beta$. Dann gilt für alle $\sigma \in \{\pm 1\}^N$:*
(i) Falls $\beta' \leq \sqrt{\frac{\log 2}{2}}$, so gilt

$$|\tilde{\mu}_{N,\beta}(\Delta_n(\sigma)) - 2^{-n}| \leq \frac{1}{2^n} e^{-(N-n)(\log 2 - \beta'^2)} Y_{N-n},$$

wobei Y_N eine beschränkte Varianz hat ($N \in \mathbb{N}$). (ii) Ist $\sqrt{\frac{\log 2}{2}} < \beta' < \sqrt{2 \log 2}$, so gilt

$$|\tilde{\mu}_{N,\beta}(\Delta_n(\sigma)) - 2^{-n}| \leq \frac{1}{2^n} e^{-(N-n)\frac{(\sqrt{2\log 2} - beta)}{2} - \frac{\alpha(\log(N-n)}{2}} Y_{N-n},$$

wobei Y_N beschränkten Erwartungswert hat. (iii) Für $\beta = \sqrt{2 \log 2}$ gilt für jedes feste n

$$|\tilde{\mu}_{N,\beta}(\Delta_n(\sigma)) - 2^{-n}| \to 0$$

in Wahrscheinlichkeit.

Bemerkung: Man beachte die exponentielle Geschwindigkeit der Konvergenz in (i) und (ii). Da sich das System bei $\beta = \sqrt{2\log 2}$ im Phasenübergang befindet, ist so ein Verhalten unter (iii) nicht zu erwarten.

Beweis: (Skizze) Der Beweis dieses Lemmas folgt einfach aus dem Hochtemperatur-CLT für $Z_{N,\beta}$ wie wir es in Abschnitt 11.2 untersucht haben. Hierzu bemerken wir nur, dass die

$$Z_{N,\beta}(\sigma_n) := \mathbb{E}_{\sigma'} e^{\beta\sqrt{N}X_{\sigma'}} \mathbb{1}_{\pi_n(\sigma')=\sigma_n}$$

unabhängig sind und verteilt sind wie

$$2^{-n} Z_{N-n,\beta'}.$$

Aber

$$\tilde{\mu}_{N,\beta}(\Delta_n(\sigma_n)) = \frac{Z_{N,\beta}(\sigma_n)}{\sum_{\sigma'_n\in\{\pm1\}^n} Z_{N,\beta}(\sigma'_n)}.$$

Nun sind alle 2^n Zufallsvariablen $Z_{N,\beta}(\sigma_n)$ unabhängig und man hat daher hervorragende Kontrolle über deren Konvergenzgeschwindigkeit. □

Nun können wir auch den Beweis des Satzes fertigstellen: Wir können nun das Maß aller Intervalle $\Delta_n(\sigma)$ approximieren. Nun können wir auch allgemeinere Mengen durch Vereinigungen solcher Intervalle approximieren und dies beweist den Satz. □

Das Tieftemperaturverhalten dieser Maße ist noch ein wenig spannender. Es sei $\mathcal{R}$ der Poissonsche Punktprozess auf dem Streifen $(0,1] \times \mathcal{R}$ mit Intensitätsmaß

$$\frac{1}{2} dyxe^{-x}dy.$$

Falls $(X_k, Y_k),\ k = 1, 2, \ldots$ die Atome dieses Prozesses sind, sei $\mathcal{M}_\alpha$ ein neuer Punktprozess auf $(0,1] \times (0,1]$ mit Atomen (Y_k, w_k), wobei

$$w_k := \frac{e^{\alpha X_k}}{\int e^{\alpha x}\mathcal{R}(dy, dx)}$$

für $\alpha > 1$. Mit dieser Notation gilt

Satz 11.15 *Sei $\beta > \sqrt{2\log 2}$ und*

$$\alpha := \frac{\beta}{\sqrt{2\log 2}},$$

dann gilt

$$\tilde{\mu}_{N,\beta} \xrightarrow{\mathcal{D}} \tilde{\mu}_\beta := \int_{(0,1]\times(0,1]} \mathcal{M}_\alpha(dy, dw)\delta yw.$$

Beweis: Sei $u_N(x)$ wieder definiert als

$$u_N(x) := \sqrt{2N\log 2} + \frac{x}{\sqrt{2N\log 2}} - \frac{\log(N\log 2) + \log 4\pi}{2\sqrt{2N\log 2}}.$$

Definiere den Punktprozess $\mathcal{R}_N$ auf $(0,1] \times \mathcal{R}$ durch

$$\mathcal{R}_N := \sum_{\sigma\in\{\pm 1\}^N} \delta_{(r_N(\sigma), u_N^{-1}(X_\sigma))}.$$

Ein Standardresultat der Extremwerttheorie ergibt

$$\mathcal{R}_N \xrightarrow{\mathcal{D}} \mathcal{R}, \quad N \to \infty,$$

wobei die Konvergenz im Sinne der schwachen Konvergenz auf dem Raum der σ-endlichen Maße ausgestattet mit der Topologie der vagen Konvergenz gilt. Nun ist

$$\mu_{N,\beta}(\sigma) := \frac{e^{\alpha u_n^{-1}(X_\sigma)}}{\sum_{\sigma'} e^{\alpha U_N^{-1}(X_{\sigma'})}} = \frac{e^{\alpha U_N^{-1}(X_\sigma)}}{\int e^{\alpha x}\mathcal{R}_N(dy,dx)}.$$

Da

$$\int \mathcal{R}_N(dy,dx)e^{\alpha x} < +\infty \quad \text{f.s.},$$

können wir den Punktprozess $\mathcal{M}_{N,\alpha}$ via

$$\mathcal{M}_{N,\alpha} := \sum_{\sigma\in\{-1,+1\}^N} \delta_{\left(r_N(\sigma), \frac{e^{\alpha U_N^{-1}(X_\sigma)}}{\int e^{\alpha x}\mathcal{R}_N(dy,dx)}\right)}$$

definieren (dies ist ein Punktprozess auf $(0,1] \times (0,1]$. Dann ist

$$\tilde{\mu}_{N,\beta} = \int \delta_y w \mathcal{M}_{N,\alpha}(dy,dw).$$

Der einzige nicht-triviale Punkt im Beweis der Konvergenz ist es zu zeigen, dass der Betrag im Nenner, der von Atomen mit $u_N(X_\sigma) < x$ verschwindet, wenn $x \to -\infty$ konvergiert. Dies haben wir schon im vorhergehenden Abschnitt gezeigt. Somit folgt

$$\mathcal{M}_{N,\alpha} \xrightarrow{\mathcal{D}} \mathcal{M}_\alpha$$

und somit der Satz. $\square$

12 Das Hopfield-Modell

12.1 Die Zugänge zum Hopfield-Modell

a) Der Spinglaszugang
Fünf Jahre nach der Entwicklung des Sk-Modells durch Sherrington und Kirkpatrick stellten Postur und Fijohn in drei Arbeiten eine andere Version eines Spinglases vor. Während im Sk-Modell die Hamiltonfunktion die Form

$$H_N^{Sk}(\sigma) = -\frac{1}{\sqrt{N}} \sum_{i<j} \sigma_i \sigma_j J_{ij}$$

mit J_{ij} i.i.d. und standardnormalverteilt, schlagen Postur und Fijohn als Hamiltonfunktion

$$H_N(\sigma) = -\frac{1}{N} \sum_{i<j} \sigma_i \sigma_j \sum_{\mu=1}^{M} \xi_i^n \xi_j^n$$

vor. Hierbei sind die ξ_i^n i.i.d. Zufallsvariable mit

$$\mathbb{P}(\xi_i^n = +1) = \mathbb{P}(\xi_i^n = -1) = \frac{1}{2}.$$

Dass $\frac{1}{N}$ die "richtige" Skalierung ist, sieht man, wenn man $M = M(N) = N$ wählt. In diesem Fall wäre

$$H_N(\sigma) = -\frac{1}{N} \sum_{i<j} \sigma_i \sigma_j \sum_{\mu=1}^{N} \xi_i^n \xi_j^n = -\frac{1}{\sqrt{N}} \sum_{i<j} \sigma_i \sigma_j \frac{1}{\sqrt{N}} \sum_{\mu=1}^{N} \xi_i^n \xi_j^n.$$

Für festes (i,j) sind die (ξ_i^n, ξ_j^n), $\mu = 1, \ldots, N$ i.i.d. Zufallsvariablen mit Erwartungswert 0 und Varianz 1. Nach dem CLT konvergieren die $\frac{1}{\sqrt{N}} \sum_\mu \xi_i^n \xi_j^n$ also gegen (unkorrelierte) $\mathcal{N}(0,1)$-verteilte Zufallsvariablen J_{ij}. Zumindest in diesem Limes haben wir also in Verteilung ein SK-Modell. Der Vorteil des soeben skizzierten Modells liegt u. a. darin, dass sich mit dem Parameter M die Komplexität des Modells bestimmen lässt, je größer M ist, desto komplexer ist das Modell. Fijohn und Postur untersuchten das Modell für endliches M und fanden dort allerdings nicht die spinglasartigen Strukturen, sondern nur eine Art "Überlagerung" von M Curie-Weiß-Modellen.

b) Der Zugang über neuronale Netzwerke
1983 beschäftigte sich John Hopfield, ein Physiker vom Caltech, mit der Modellierung neuronaler Aktivitäten und entdeckte dabei das später nach ihm benannte Modell zum zweiten Mal.

Ein Gehirn ist an sich ein chaotisch wirkendes System, bei dem ca. 10^9 bis 10^{10} Neuronen miteinander verbunden sind und wechselwirken. Diese Wechselwirkungen werden über sogenannte Synopsen ausgetaucht. Über diese Synopsen tauschen die Neuronen elektrische Signale aus. Dabei hängen die von einem Neuron ausgesandten Signale (komplex) von den empfangenen Signalen ab. Da selbst ein einzelnes Neuron schon bio-chemisch sehr komplex ist, ist die Frage, wie sich ein System sehr vieler solcher Neuronen analysieren lässt. Der Schlüssel ist zunächst ein Modell, das verschiedene Vereinfachungsschritte vorsieht. Zunächst werden die Signale der Neuronen in zwei Klassen eingeteilt: “schnell” und “langsam”. Dies bedeutet, dass man einem Neuron einen Ising-artigen Zustandsraum $S = \{+1, -1\}$ zuweist. Darüber hinaus war schon zu Hopfields Zeiten bekannt, dass das Ausgangssignal von dem gemeinsamen, gewichteten Eingangssignal abhängt. Dies kann modelliert werden als

$$h_i = f_i(\{\sigma_j\}, j \in (N(i))).$$

Hierbei ist $(N(i))$ die Menge aller Neuronen, die auf Neuron i feuern und f_i ist eine (lokale) Funktion, die die Art der Abhängigkeit des Ausgangssignals vom Eingangssignal bestimmt. Die einfachste Art, die f_i zu wählen, ist sicherlich eine lineare Funktion:

$$f_i(\{\sigma_j\}, j \in N(i)) = \sum_{j \in N(i)} J_{ij}\sigma_j.$$

Da die J_{ij} Neuronen i sowohl dazu bewegen können sollen zu feuern als auch nicht zu feuern, sollten sie beide Vorzeichen annehmen können. Natürlich müssen die J_{ij} irgendwie mit der zu speichernden Information zusammenhängen. Wir verfolgen hierbei ein Konzept, das schon D. Hebb 1949 vorschlug. Die Idee hierbei ist, dass das Netzwerk “autoassoziativ” sein soll, d. h. falls das Netzwerk sich nahe einer gelernten Eingabe befindet, wird dieses gelernte Muster verstärkt. Hierzu stellen wir uns vor, dass das Netzwerk die Muster $\xi^1, \ldots, \xi^M$ lernen soll. Diese stellen wir uns schon in der Form vor, in der sie vom Netzwerk rezepiert werden, also

$$\xi^n = (\xi_i^n)_{i=1}^N \quad \text{mit} \quad \xi_i^n \in \{-1, +1\} \quad \text{für alle} \quad i = 1, \ldots, N,\ \mu = 1, \ldots, M.$$

Wir sehen, dass in diesem Fall

$$h_i = \sum_{j \in N(i)} \sum_{\mu=1}^{M} \xi_i^n \xi_j^n \sigma_j,$$

d. h. in dem Fall, dass wir uns IN einem gelernten Muster befinden, d. h. z. B.

$$\sigma = \xi^1, \quad \text{d. h.} \quad \sigma_i = \xi_i^1 \quad \forall\, i,$$

so ist

$$h_i = \sum_{j \in N(i)} \sum_{\mu=1}^{M} \xi_i^n \xi_j^n \xi_j^1 = |N(i)|\xi_i^1 + \sum_{j \in N(i)} \sum_{\mu=2}^{M} \xi_i^n \xi_j^n \xi_j^1,$$

das “lokale Feld” h_i, hat also eine das gelernte Muster ξ^1 verstärkende Wirkung. Schließlich nehmen wir aus Gründen der Einfachheit auch an, dass alle Neuronen untereinander kommunizieren, d. h., dass $N(i) = \{1, \ldots, N\}$ für alle i gilt. Dann ist

$$h_i = \sum_{j=1}^{N} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \sigma_j.$$

Wir folgen nun Hopfield und nehmen an, dass die Dynamik des Netzwerkes durch eine Markov-Kette in kontinuierlicher Zeit gegeben ist. Genauer statten wir alle Neuronen mit einer Uhr aus, deren Alarm nach einer exp(1)-verteilten Zeit klingelt (und alle diese Uhren seien unabhängig). Wenn der Alarm von Neuron i losgeht, so wechselt σ_i sein Vorzeichen und zwar mit Wahrscheinlichkeiten:

$$\mathbb{P}(\sigma_i(t)) = \pm 1) = \frac{e^{\pm\beta h_i/N}}{e^{\beta h_i/N} + e^{-\beta h_i/N}}.$$

Es ist leicht nachzurechnen, dass das stationäre Maß dieser Markov-Kette ein Gibbs-Maß

$$\mu_N(\sigma) = \frac{e^{-\beta H_N(\sigma)}}{\sum_{\sigma'} e^{-\beta H_N(\sigma')}}$$

mit

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j=1}^{N} \sigma_i \sigma_j \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu =: -\frac{1}{N} \sum_{i,j=1}^{N} \sigma_i \sigma_j J_{ij}$$

ist. Genau diese Hamiltonfunktion haben wir schon eben kennengelernt.

c) Ein Zugang über ein sozio-ökonomisches Modell

Eine Bevölkering bestehend aus N Individuen soll bezüglich ihrer Meinungsbildung modelliert werden. Hierbei gebe es a priori nur zwei Meinungen "dafür" $(= +1)$ oder "dagegen" $(= -1)$, wobei "für oder gegen was" keine Rolle spielt. Hierbei ist wieder jedes Individuum i mit einer exp(1)-Uhr ausgestattet und wenn deren Alarm klingelt, entscheidet sich i erneut. Hierbei ist seine Entscheidung, nennen wir sie σ_i, a priori unvoreingenommen, d. h. er oder sie entscheidet sich a priori mit gleichen Wahrscheinlichkeiten für $+1$ oder für -1. Allerdings ist jedes Individuum einem Gruppendruck ausgesetzt. Im einfachsten Fall ist i versucht, sich eher der Mehrheit als der Minderheit anzuschließen (wir nehmen an, dass diese Meinungen beispielsweise durch Massenmedien publik gemacht werden). Dies führt zu einem Curie-Weiß-artigen Model. In einer verfeinerten Version teilen wir die Bevölkering in soziale Klassen ein. Diese seien der Einfachheit halber binär und paritätisch verteilt gewählt (dies ist nicht zwingend notwendig, erleichtert aber die Arbeit und soll für die Zwecke der Vorlesung so angenommen werden). Man habe also M soziale Kategorien, etwa

$$\begin{aligned} \xi_i^1 &= +1 \quad \hat{=} \quad i \quad \text{ist männlich} \\ \xi_i^1 &= -1 \quad \hat{=} \quad i \quad \text{ist weiblich} \\ \xi_i^2 &= +1 \quad \hat{=} \quad i \quad \text{ist reich} \\ \xi_i^2 &= -1 \quad \hat{=} \quad i \quad \text{ist arm} \end{aligned}$$

etc. Wenn Individuum i sich entscheiden muss, wird er/sie nun versuchen sich ähnlich zu den Menschen zu entscheiden, mit denen er in vielen sozialen Kategorien übereinstimmt. D. h. er wichtet die Meinung von Individuum j bei seiner Meinungsbildung, je nachdem in wievielen Faktoren i und j übereinstimmen. Eine Art, den äußeren Einfluss auf i dann zu messen ist

$$\bar{h}_i = \frac{1}{N} \sum_{j=1}^{N} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \sigma_j.$$

Nimmt man wieder eine Markovsche Dynamik mit

$$\mathbb{P}(\sigma_i(+) = \pm 1) = \frac{e^{\pm\beta h_i}}{e^{\beta h_i} + e^{-\beta h_i}}$$

an, so ist das invariante Maß wiederum das Gibbs-Maß

$$\mu_{N,\beta}(\sigma) = \frac{e^{-\beta H_N(\sigma)}}{\sum_{\sigma'} e^{-\beta H_N(\sigma)}}$$

mit

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j=1}^{N} \sum_{\mu=1}^{M} \sigma_i \sigma_j \xi_i^\mu \xi_j^\mu,$$

und wir erhalten wieder dasselbe Modell.

12.2 Die Speicherkapazität des Hopfield-Modells

In diesem Abschnitt werden wir uns auf den Aspekt des Hopfield-Modells als neuronales Netz konzentrieren. Hierbei hatten wir gesehen, dass das Hopfield-Modell als AUTOASSOZIATIVER Speicher fungieren soll, d. h. wenn ihm ein "gelerntes" Muster ξ^μ vorgelegt wird, soll das Netzwerk dies wiedererkennen. Nehmen wir für den Beginn $M = 1$ an, so funktioniert diese Idee in der Tat: Dann ist nämlich in ξ^1 das lokale Feld

$$h_i(\xi^1) = \sum_{j=1}^{N} \xi_i^1 \xi_j^1 \xi_j^1 = N\xi_i^1.$$

Nehmen wir nun an, dass die Temperatur $T = 0$, d. h. $\beta = +\infty$ ist, so ist

$$\mathbb{P}(\sigma_i(+) = \pm 1) = \frac{e^{\pm\beta N \xi_i^1}}{e^{\beta N \xi_i^1} + e^{-\beta N \xi_i^1}}.$$

Dies ist 1 oder 0, je nachdem, ob $\xi_i^1 = +1$ oder $\xi_i^1 = -1$ ist. M.a.W.: Startet man in $\sigma = \xi^1$, so behält man exakt dieses Muster bei, das Netzwerk erkennt dieses Muster wieder. Startet man mit $M > 2$ Mustern (und sagen wir wieder in $\sigma = \xi^1$) und nimmt zusätzlich an, dass diese orthogonal sind, d. h., dass

$$< \xi^\mu, \xi^\nu >= \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu = N\delta_{\mu,\nu}$$

gilt, so kommen ähnliche Ideen zum Tragen. In diesem Fall ist

$$\begin{aligned} h_i(\xi^1) &= \sum_{j=1}^{N} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \xi_j^1 \\ &= N\xi_i^1 + \sum_{\mu=1}^{M} \xi_i^\mu \sum_{j=1}^{N} \xi_j^\mu \xi_j^1 \\ &= N\xi_i^1 \end{aligned}$$

und wieder werden die Muster wiedererkannt. In der Regel werden aber zu lernende Muster nicht orthogonal sein (und sie zu orthogonalisieren ist aufgrund der Grundannahmen für neuronale Netze verboten). In der Literatur der neuronalen Netze hat es sich bewährt anzunehmen, dass die ξ^μ unabhängig und identisch verteilt sind und aus i.i.d. Komponenten ξ_i^μ mit

$$\mathbb{P}(\xi_i^\mu = +1) = \mathbb{P}(\xi_i^\mu = -1) = \frac{1}{2}$$

bestehen. Eine Realisierung solcher $\xi^1, \dots, \xi^M$ erfüllt natürlich in der Regel nicht die Orthogonalitätskriterien, allerdings gilt

$$\frac{1}{N} < \xi^\mu, \xi^\nu >= \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \xi_i^\nu \to \delta_{\mu,\nu} \quad \mathbb{P}\text{-f.s.}$$

für alle μ, ν. Wir fragen uns, ob dies genügt, um das Netzwerk die (ξ^μ) speichern zu lassen und wie groß M in diesem Fall sein kann. Hierfür ist zum einen zu bemerken, dass die zu erwartenden Aussagen nur wahrscheinlichkeitstheoretischer Natur sein können, da die (ξ^μ) ja zufällig sind. Zum anderen wollen wir unter "wiedererkennen" das folgende verstehen: Wir wählen wieder $T = 0$, d. h. $\beta = +\infty$. Wir haben schon gesehen, dass die Dynamik in diesem Fall den Spin σ_i einfach in die Richtung des Vorzeichens des lokalen Feldes h_i klappt, d. h. die Dynamik ist gegeben durch $T = (T_i)_{i=1,\dots,N}$ mit

$$T_i(\sigma) = sgn(\sum_{j=1}^N \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \sigma_j).$$

Die Mindestanforderung, die wir nun an das Netzwerk für das "Wiedererkennen" stellen wollen ist, dass ein gespeichertes Muster ξ^μ stabil ist unter T. Die ersten diesbezüglichen Resultate sind

Satz 12.1 *Es sei* $M = c\frac{N}{\log N}$.

a) Ist $c < \frac{1}{2}$, *so gilt für jedes feste* μ

$$\mathbb{P}(T_i(\xi^\mu) = \xi_i^\mu \quad \forall\, i = 1, \dots, N) \to 1,$$

wenn $N \to \infty$.

b) Ist $c \le \frac{1}{4}$, *so gilt*

$$\mathbb{P}(T_i(\xi^\mu) = \xi_i^\mu \quad \forall\, i = 1, \dots, N,\ \forall\, \mu = 1, \dots, M) \to 1,$$

wenn $N \to \infty$.

c) Ist $c < \frac{1}{6}$, *so gilt*

$$\mathbb{P}(\liminf\{T_i(\xi^\mu) = \xi_i^\mu \quad \forall\, i = 1, \dots, N\}) \to 1.$$

Beweis: Der Satz folgt im wesentlichen aus der exponentiellen Chebyshev-Ungleichung. In der Tat genügt es ja zunächst, die Stabilität von ξ^1 zu betrachten. Hierzu kann wiederum angenommen werden, dass

$$\xi_i^1 = 1 \quad \forall\, i = 1, \ldots, N$$

gilt, da man sonst den i-ten Spin "dreht", d. h. mit ξ_i^1 multipliziert. Nun ist

$$\begin{aligned} & T_i(\xi^1) = \xi_i^1 \\ \Leftrightarrow \quad & sgn(\sum_{j=1}^N \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^1) = 1 \\ \Leftrightarrow \quad & \sum_{j=1}^N \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \geq 0 \end{aligned}$$

(der Fall "$= 0$" erfordert eigentlich noch eine gesonderte Betrachtung, kann hier aber fortgelassen werden, da er mit einer Wahrscheinlichkeit auftritt, die für $N \to \infty$ verschwindet). Nun hat $\sum_{j=1}^N \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu$ aber die Tendenz, positiv zu sein, da der Summand für $\mu = 1$ $\sum_{j=1}^N \xi_i^1 \xi_j^1 = N$ ist. Ergo

$$\begin{aligned} & \mathbb{P}(T_i(\xi^1) \neq \xi_i^1) \\ = \quad & \mathbb{P}(\sum_{j=1}^N \sum_{\mu=2}^M \xi_i^\mu \xi_j^\mu \leq -N) \\ \leq \quad & e^{-tN} \mathbb{E} e^{t \sum_{j=1}^N \sum_{\mu=2}^M \xi_i^\mu \xi_j^\mu} \\ = \quad & e^{-tN} (e^{t \xi_i^\mu \xi_j^\mu})^{NM}. \end{aligned}$$

Hierbei ist $t > 0$ und wir haben ausgenutzt, dass die $\xi_i^\mu \xi_j^\mu$ für $j = 1, \ldots, N$ und $\mu = 1, \ldots, M$ i.i.d. Bernoulli-Variablen sind (hierfür ist die Struktur der ξ_i^μ wichtig). Da

$$\cosh(x) \leq e^{x^2/2}$$

gilt, folgt

$$\mathbb{P}(T_i(\xi^1) \neq \xi_i^1) \leq e^{-tN} (\cosh(t))^{NM} \leq e^{-tN} e^{t^2 NM/2},$$

also

$$\mathbb{P}(\exists\, i = 1, \ldots, N : T_i(\xi^1) \neq \xi_i^1) \leq N e^{-tN + t^2/2NM}.$$

Das optimale t ist durch $t = \frac{1}{M}$ gegeben:

$$\mathbb{P}(\exists i : T_i(\xi^1) \neq \xi_i^1) \leq N e^{-\frac{1}{2} NM}.$$

Wählt man nun $M = c \frac{N}{\log N}$, so sehen wir

$$\mathbb{P}(\exists i : T_i(\xi^1) \neq \xi_i^1) \leq N^{1 - \frac{1}{2c}},$$

was für $c < \frac{1}{2}$ gegen 0 konvergiert. Ebenso sieht man

$$\mathbb{P}(\exists_\mu \exists i : T_i(\xi^\mu) \neq \xi_i^\mu) \leq NM e^{-\frac{1}{2} NM}.$$

Setzt man wieder $M = c\frac{N}{\log N}$, so konvergiert die rechte Seite für $c \leq \frac{1}{4}$ gegen 0, dies beweist Teil b).

Schließlich ist (wieder mit $M = c\frac{N}{\log N}$)

$$NMe^{-\frac{1}{2}NM} \leq cN^{2-\frac{1}{2c}}$$

für $c < \frac{1}{6}$ von der Form $cN^{-\kappa}$ mit $\kappa > 1$, somit summierbar und der letzte Teil des Satzes folgt aus dem Borel-Cantelli-Lemma. □

Nun ist natürlich das pure Wiedererkennen eines einmal gelernten Musters nicht unbedingt das, was man eine ASSOZIATIVE Fähigkeit nennen würde (vielmehr zeigt das Netzwerk nur, dass es kein Alzheimer hat). Wir wollen nun sehen, was passiert, wenn ein Teil des Eingabemusters gestört ist:

Satz 12.2 *Für jedes μ sei $\tilde{\xi}^\mu \in S_{rN}(\xi^\mu)$, wobei*

$$S_r(x) = \{y \in \{-1, +1\}^N : d(x, y) = r\}$$

$d(\cdot, \cdot)$ der Haming-Abstand ist (also die Anzahl der unterschiedlichen Koordinaten), $0 \leq r < \frac{1}{2}$ und wir der Einfachheit halber annehmen, dass rN eine ganze Zahl ist. Dann gilt: Ist

$$M = c \cdot (1 - 2r)^2 \frac{N}{\log N}$$

und

a) $c < \frac{1}{2}$, so gilt für jedes feste μ

$$\mathbb{P}(T_i(\tilde{\xi}^\mu) = \xi_i^\mu \quad \forall\, i = 1, \ldots, N) \to 1$$

mit $N \to \infty$.

b) $c \leq \frac{1}{4}$, so gilt

$$\mathbb{P}(T_i(\tilde{\xi}^\mu) = \xi_i^\mu \quad \forall\, i = 1, \ldots, N,\ \forall\, \mu = 1, \ldots, M) \to 1,$$

wenn $N \to \infty$.

c) $c < \frac{1}{6}$, so gilt

$$\mathbb{P}(\liminf\{T_i(\tilde{\xi}^\mu) = \xi_i^\mu\}) = 1.$$

Bemerkung: Da mit ξ^μ auch $-\xi^\mu$ gelernt wird, ist eine "Zerstörungsrate" von rN mit $r < \frac{1}{2}$ das beste, was wir für eine erfolgreiche Rekonstruktion noch erhoffen können.

Beweis: Nun ist für festes ξ^μ mit $\xi_i^\mu = +1 \quad \forall\, i$ (o.B.d.A.)

$$T_i(\tilde{\xi}^\mu) = sgn(\sum_{j=1}^{N} \sum_{\nu=1}^{M} \xi_i^\nu \xi_j^\nu \tilde{\xi}_j^\mu) = N(1 - 2r) + \sum_{j=1}^{N} \sum_{\nu \neq \mu} \xi_i^\nu \xi_j^\nu \tilde{\xi}_j^\mu.$$

Also für alle $t > 0$

$$\mathbb{P}(T_i(\tilde{\xi}^\mu) \neq \xi_i^\mu) \leq e^{-tN(1-2r)} \times \mathbb{E}e^{t\sum_j \sum_{\nu\neq\mu} \xi_i^\nu \xi_j^\nu \tilde{\xi}_j^\mu}.$$

Der Rest des Beweises folgt nun dem vorhergehenden Beweis. $\square$

Bemerkung: Man kann zeigen, dass man die Speicherkapazität des Hopfield-Netzwerkes wieder auf die bekannten Werte von $\frac{N}{2\log N}$, $\frac{N}{4\log N}$ bzw. $\frac{N}{6\log N}$ anheben kann, wenn mehrere Schritte der Dynamik T zulässt. Dies soll aber hier nicht weiter verfolgt werden.

Wir wollen nun sehen, dass die jeweils unter a) gefundene Schranke in den letzten beiden Sätzen schon optimal ist. Erhöht man M über $\frac{N}{2\log N}$, so vergisst das Netzwerk alles. Dieser Abschnitt ist auch methodisch interessant, weil wir das Konzept der negativen Assoziertheit kennenlernen. Zu diesem Zweck nennen wir

$$f : \mathbb{R}^n \to \mathbb{R}^x$$

steigend, falls es bezüglich der Partialordnung $\leq$ auf $\mathbb{R}^n$ steigend ist, wobei

$$(x_1, \ldots, x_n) \leq (y_1, \ldots, y_n) \Leftrightarrow x_i \leq y_i \quad \forall i = 1, \ldots, n.$$

Definition 12.3 *Zufallsvariablen $Y_1, \ldots, Y_n$ heißen* NEGATIV ASSOZIIERT, *falls für jedes Paar*

$$A_1, A_2 \subseteq \{1, \ldots, n\} \quad \textit{mit} \quad A_1 \cap A_2 = \emptyset$$

und je zwei steigende Funktionen

$$f_1 : \mathbb{R}^{|A_1|} \to \mathbb{R}$$
$$f_2 : \mathbb{R}^{|A_2|} \to \mathbb{R}$$

gilt

$$Cov(f_1(Y_i|i \in A_1),\ f_2(Y_j|j \in A_2)) \leq 0.$$

Ein Vektor $Y = (Y_1, \ldots, Y_n)$ heißt negativ assoziiert, wenn seine Komponenten es sind.

Wir sammeln (ohne Beweis) ein paar Fakten über negative Assoziiertheit. Zunächst stellen wir fest, dass negative Assoziiertheit unter natürlichen Operationen erhalten bleibt.

Proposition 12.4 *Sind $Y_1, \ldots, Y_n$ negativ assoziierte Zufallsvariablen und ist $c \in \mathbb{R}$, so sind auch $cY_1, \ldots, cY_n$ negativ assoziiert.*

Proposition 12.5 *Eine Vereinigung unabhängiger Familien von negativ assoziierten Zufallsvariablen ist wieder eine Familie negativ assoziierter Zufallsvariablen.*

Für einen letzten Erhaltungssatz benötigen wir den Begriff der Permutationsverteilung.

Definition 12.6 *Sei $y = (y_1, \dots, y_n) \in \mathbb{R}^n$. Der Vektor*

$$Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$$

besitzt eine Permutationsverteilung bezüglich y, falls Y alle $n!$ Permutationen der Koordinaten von y mit gleicher Wahrscheinlichkeit $\frac{1}{n!}$ annimmt.

Satz 12.7 *Besitzt Y eine Permutationsverteilung bezüglich irgendeines y, so ist Y negativ assoziiert.*

Schließlich stellen wir fest, dass negative Assoziiertheit ein recht nützliches Konzept ist.

Proposition 12.8 *Seien $A_1, \dots, A_m \subseteq \{1, \dots, n\}$ paarweise disjunkt und $f_1, \dots, f_m$ steigend und nicht-negativ. Dann gilt für negativ assoziierte Zufallsvariablen $Y_1, \dots, Y_n$*

$$\mathbb{E} \prod_{i=1}^{m} f_i(Y_j | j \in A_i) \leq \prod_{i=1}^{m} \mathbb{E} f_i(Y_j | j \in A_i).$$

Mit Hilfe dieses Konzepts (und einiger technischer Arbeit) lässt sich nun das folgende Resultat beweisen.

Satz 12.9 *Gilt im Hopfield-Modell für $M = M(N)$*

$$M(N) \geq c \frac{N}{\log N}$$

mit $c > \frac{1}{2}$, dann gilt für jedes feste μ

$$\lim_{N \to \infty} \mathbb{P}[\forall\, i = 1, \dots, N : T_i \xi^\mu = \xi_i^\mu] = 0.$$

Bemerkung: Dieser Satz stellt somit fest, dass sich das Speicherverhalten bei $\frac{N}{2 \log N}$ drastisch ändert. Ähnliche Gegenpositionen zu Teil b) und c) des vorhergehenden (positiven) Speicherresultats lassen sich leider (noch) nicht beweisen.

Beweis: Wir können wieder nur ein Muster ξ^1 betrachten und annehmen, dass $\xi_i^\mu = 1$ für alle $i = 1, \dots, N$ gilt. Damit ist dann zu zeigen, dass

$$\mathbb{P}[\sum_{j=1}^{N} \sum_{\mu=2}^{M} \xi_i^\mu \xi_j^\mu \geq -N \quad \forall\, i = 1, \dots, N) \to 0$$

gilt. Die Hauptschwierigkeit, die uns im vorhergehenden Satz nicht weiter gestört hat, ist die fehlende Unabhängigkeit der Zufallsvariablen

$$X_i^\mu := \xi_i^\mu \sum_{j=1}^{N} \xi_j^\mu$$

für verschiedene i. Die negative Assoziiertheit wird uns hier einen Ausweg weisen. Wir definieren

$$W_\mu = \frac{1}{\sqrt{N}} \sum_{j=1}^N \xi_j^\mu.$$

Dann gilt

$$\begin{aligned} &\mathbb{P}[\sum_{j=1}^N \sum_{\mu=2}^M \xi_i^\mu \xi_j^\mu \geq -N \quad \forall\, i = 1, \ldots, n] \\ = &\sum_{\{w_\mu\}_{\mu=2,\ldots,M}} \mathbb{P}[\forall\, i : \sum_{\mu=2}^M \xi_i^\mu w_\mu \geq -\sqrt{N} | W_\mu = w_\mu \;\forall\, \mu] \times \mathbb{P}[W_\mu = w_\mu \;\forall\, \mu]. \end{aligned}$$

Hierbei erstreckt sich die Summe rechts über alle Werte w_μ, die die Zufallsvariablen W_μ annehmen können. Setze für $w = (w_\mu)_{\mu=2,\ldots,M}$

$$\mathbb{P}_w[\cdot] = \mathbb{P}[\cdot | W_\mu = w_\mu \;\forall\, \mu = 2, \ldots, M].$$

Dann gehorchen unter $\mathbb{P}_w$ die Zufallsvariablen ξ_i^μ einer Permutationsverteilung, denn da ihre Summe festgelegt ist und sie nur Werte ± 1 annehmen können, ist jede ihrer möglichen Permutationen gleichwahrscheinlich. Somit sind die $(\xi_i^\mu)_{i=1}^N$ unter $\mathbb{P}_w$ negativ assoziiert. Also auch alle

$$\xi_i^\mu,\ j = 1, \ldots, N,\ \mu = 2, \ldots, M.$$

Wendet man nun die Eingangsüberlegung auf die (steigende, nicht-negative) Funktion

$$\mathbb{1}_{(-\sqrt{N}, \infty)}$$

an, so erhalten wir

$$\begin{aligned} &\mathbb{P}_w[\sum_{\mu=2}^M \xi_i^\mu w_\mu \geq -\sqrt{N} \;\forall\, i = 1, \ldots, N] \\ \leq\ &\prod_{i=1}^N \mathbb{P}_w[\frac{1}{\sqrt{N}} \sum_{\mu=2}^M \xi_i^\mu w_\mu \geq -1] \\ =\ &(\mathbb{P}_w[\frac{1}{\sqrt{N}} \sum_{\mu=2}^M \xi_1^\mu w_\mu \geq -1])^N. \end{aligned}$$

Es bleibt also im wesentlichen

$$\mathbb{P}_w[\frac{1}{\sqrt{N}} \sum_{\mu=2}^M \xi_i^\mu w_\mu \geq -1]$$

zu kontrollieren. Nun sind die ξ_1^μ unter $\mathbb{P}_w$ immer noch unabhängig und, da $\sum \xi_j^\mu = w_\mu$ ist und die ξ_j^μ identisch verteilt sind, ist

$$\mathbb{P}_w[\xi_1^\mu = \pm 1] - \frac{1}{2}(1 \pm \frac{w_\mu}{\sqrt{N}}).$$

Damit berechnet sich dann

$$\mathbb{E}_w(\frac{1}{\sqrt{N}}\sum_{\mu=2}^{M}\xi_1^\mu w_\mu) = \frac{1}{\sqrt{N}}\sum_{\mu=2}^{M} w_\mu \mathbb{E}_w[\xi_1^\mu] = \frac{1}{\sqrt{N}}\sum_{\mu=2}^{M} w_\mu \frac{1}{\sqrt{N}} w_\mu = \frac{1}{N}\sum_{\mu=2}^{M} w_\mu^2.$$

Ähnlich sieht man, dass

$$\mathbb{V}_w(\frac{1}{\sqrt{N}}\sum_{\mu=2}^{M}\xi_1^\mu w_\mu) = \frac{1}{N}\sum_{\mu=2}^{M} w_\mu^2(1-\frac{w_\mu^2}{N}).$$

Beide dieser Größen sind mit großer Wahrscheinlichkeit (in den ξ_i^μ) nahe bei $\alpha := \frac{M(N)}{N}$. In der Tat gilt:

Lemma 12.10 *Für die oben definierten Zufallsvariablen W_μ gilt das folgende: Sei $\Omega_{\varepsilon,c,N} \subseteq \Omega$ die Menge aller Realisierungen von $\xi_i^\mu = \xi_i^\mu(w)$, für die gilt*

$$\frac{1}{M}\sum_{\mu=2}^{M} W_\mu^2 e[1-\varepsilon, 1+\varepsilon]$$

und

$$\sup_{\mu=2,\ldots,M} |W_\mu| \leq \sqrt{2c\log M}.$$

Dann gilt für $\varepsilon \ll 1$

$$\mathbb{P}[\Omega_{\varepsilon,c,N}] \geq 1 - 2e^{-\varepsilon^2 M/8} - M^{-c+1}.$$

Beweis: Der Beweis ist nicht weiter schwer. Man beachte, dass

$$\mathbb{E}W_\mu^2 = 1 + \mathbb{E}\frac{1}{N}\sum_{i\neq j}\xi_i^\mu \xi_j^\mu = 1$$

ist und die W_μ für verschiedene μ i.i.d. sind. Nach dem Erwartungswert verhält sich die Ratenfunktion aus den großen Abweichungen nahezu quadratisch mit Steigung $\frac{1}{8}$. Dies ergibt den einen Term. Der andere folgt aus einer Chebyschev-Ungleichung zusammen mit Standardabschätzungen für $\log x$. □

Nun schreiben wir

$$\mathbb{P}_w[\frac{1}{\sqrt{N}}\sum_{\mu=2}^{M}\xi_1^\mu w_\mu \geq -1+\alpha] = 1 - \mathbb{P}_w[\frac{1}{\sqrt{N}}\sum_{\mu=2}^{M}\xi_1^\mu w_\mu < -1].$$

Ist nun α klein, d. h. $M \ll N$, so sollte

$$\mathbb{P}_2[\frac{1}{\sqrt{N}}\sum_{\mu=2}^{M}\xi_1^\mu w_\mu < -1]$$

nahe bei 0 sein, das Ereignis ist untypisch. Es liegt daher nahe, Techniken aus den großen Abweichungen zu verwenden (in diesem Fall das "Tilten", die exponentielle Maßtransformation). Für $t \leq 0$ sei

$$\mathbb{P}_w^t(\cdot) := \frac{\mathbb{E}_w[1\!\!1_{\{0\}} e^{t\sum_\mu \xi_1^\mu w_\mu}]}{\mathbb{E}_w[e^{t\sum_\mu \xi_1^\mu w_\mu}]}.$$

Nun ist für jedes $\delta > 0$

$$\begin{aligned}
\mathbb{P}_w[\frac{1}{\sqrt{N}}\sum_{\mu=2}^M \xi_1^\mu w_\mu < -1] &\geq \mathbb{P}_w[-1-\delta < \frac{1}{\sqrt{N}}\sum_{\mu=2}^M \xi_1^\mu w_\mu < -1] \\
&= \mathbb{E}_w^t[1\!\!1_{\{-1-\delta<\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu<-1\}} e^{-t\sum_\mu \xi_1^\mu w_\mu}] \times \mathbb{E}_w[e^{t\sum_\mu \xi_1^\mu w_\mu}] \\
&\geq e^{-t(-1-\delta)}\mathbb{E}_w[e^{t\sum_\mu \xi_1^\mu w_\mu}]\mathbb{P}_w^t[-1-\delta < \frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu < -1].
\end{aligned}$$

Wir wählen $t = t^*$, so dass

$$\mathbb{E}_w^{t^*}[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu] = -1 - \frac{8}{2}.$$

Eine kleine Rechnung zeigt, dass

$$\mathbb{E}_w[e^{t/\sqrt{N}\xi_1^\mu w_\mu}] = \cosh(\frac{tw_\mu}{\sqrt{N}}) + \frac{w_\mu}{\sqrt{N}}\sinh(\frac{tw_\mu}{\sqrt{N}})$$

und

$$\mathbb{E}_w[\frac{1}{\sqrt{N}}\xi_1^\mu w_\mu] = \sinh(\frac{tw_\mu}{\sqrt{N}}) + \frac{w_\mu}{\sqrt{N}}\cosh(\frac{tw_\mu}{\sqrt{N}}).$$

Also

$$\mathbb{E}_w^t[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu] = \frac{1}{\sqrt{N}}\sum_\mu w_\mu \left[\frac{\frac{w_\mu}{\sqrt{N}} + \tanh(\frac{tw_\mu}{\sqrt{N}})}{1 + \frac{w_\mu}{\sqrt{N}}\tanh(\frac{tw_\mu}{\sqrt{N}})}\right].$$

Betrachtet man, dass auf $\Omega_{\varepsilon,c,N}$ die w_μ höchstens const. $\sqrt{\log N}$ sind, nimmt vorweg, dass t^* von Ordnung $\frac{1}{\alpha}$ sein wird, so machen wir höchstens einen Fehler der Ordnung $\frac{\log N}{\sqrt{M}}$, wenn wir den tanh() nur bis zum ersten Term entwickeln. Daher erhalten wir

$$\mathbb{E}_w^t[\frac{1}{\sqrt{N}}\sum_{\mu=2}^M \xi_1^\mu w_\mu] = \frac{1}{\sqrt{N}}\sum_{\mu=2}^M w_\mu^2(q+t).$$

Somit – nimmt man $c \geq \frac{\log N}{\sqrt{M}}$ vorweg –

$$t^* = -\frac{1}{\alpha}(1 + \frac{\delta}{2} + 0(\varepsilon)).$$

Mit den gleichen Techniken sieht man, dass

$$\begin{aligned}
& e^{-t1*(-1-\delta)}\mathbb{E}_w[e^{t^*\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu}] \\
= \; & e^{-t^*(-1-\gamma)}\pi\cosh\left(\frac{t^*w_\mu}{\sqrt{N}}\right)\left(1 + \frac{w_\mu t^*}{\sqrt{N}}\tanh\left(\frac{t^*w_\mu}{\sqrt{N}}\right)\right) \\
\geq \; & e^{-\frac{1}{2\alpha}(1+\sigma\delta+0(\varepsilon))}.
\end{aligned}$$

Bleibt $\mathbb{P}_w^t[-1-\delta < \frac{1}{\sqrt{N}} \sum_\mu \xi_1^\mu w_\mu < -1]$ abzuschätzen. Dies werden wir versuchen, mittels Chebyschev-Ungleichung zu kontrollieren. Hierfür schätzen wir die Varianz ab:

$$
\begin{aligned}
& E_w^{t^*}(\frac{1}{\sqrt{N}} \sum_\mu \xi_1^\mu w_\mu - \mathbb{E}_w^{t^*} \frac{1}{\sqrt{N}} \sum_\mu \xi_1^\mu w_\mu)^2 \\
= \quad & \frac{1}{N} \sum_\mu (\mathbb{E}_w^{t^*}(\xi_1^\mu w_\mu)^2 - (E_w^{t^*}(xi_1^\mu w_\mu))^2) \\
= \quad & \frac{1}{N} \sum_\mu \left[w_\mu^2 \cosh\left(\frac{tw_\mu}{\sqrt{N}}\right) + \frac{w_\mu}{\sqrt{N}} \sinh\left(\frac{tw_\mu}{\sqrt{N}}\right) + w_\mu^2 \left[\frac{w_\mu/\sqrt{N} + \tanh(\frac{t^* w_\mu}{\sqrt{N}})}{1 + \frac{w_\mu}{\sqrt{N}} \tanh(\frac{tw_\mu}{\sqrt{N}})}\right]^2 \right] \\
= \quad & \alpha(1 + 0(\varepsilon)).
\end{aligned}
$$

Verwendet man nun die gewöhnliche Chebyshev-Ungleichung, erhält man damit

$$
\mathbb{P}_w^t[-1-\delta < \frac{1}{\sqrt{N}} \sum_\mu \xi_1^\mu w_\mu < -1] \geq 1 - \frac{\alpha(1+0(\varepsilon)}{\gamma^2}.
$$

Wählt man nun

$$
\delta = 2\sqrt{\alpha}(\to 0 \quad \text{mit} \quad N \to \infty),
$$

so folgt

$$
\mathbb{P}_w^t[-1-\delta < \frac{1}{\sqrt{N}} \sum_\mu \xi_1^\mu w_\mu < -1] \geq \frac{1}{2}.
$$

Somit gilt auf $\Omega_{\varepsilon,c,N}$ in der Tat:

$$
\begin{aligned}
& \mathbb{P}_w[\forall\, i = 1, \ldots, N \frac{1}{\sqrt{N}} \sum_{\mu=2}^M \xi_i^\mu w_n \geq -1] \qquad (12.1) \\
\leq \quad & [1 - e^{-\frac{1}{2\alpha}(1+\sigma\delta+0(\varepsilon))/2}]^N \\
\leq \quad & e^{-N} e^{-\frac{1}{2\alpha}(1+\sigma\delta+0(\varepsilon))/2}.
\end{aligned}
$$

Ist nun $M \geq \frac{N}{c \log N}$ mit $c > \frac{1}{2}$, so ist

$$
\alpha \geq \frac{\log N}{2-\gamma} \quad \text{mit einem} \quad \gamma > 0
$$

(wobei wir $\alpha \to 0$ annehmen), so können wir ε und δ wählen, so dass für hinreichend großes N gilt

$$
Ne^{-\frac{1}{2\alpha}(1+\sigma\delta+0(\varepsilon))} > \frac{1}{N^{\delta/4}},
$$

so dass die rechte Seite von (12.1) sehr schnell gegen 0 konvergiert. Da auf der anderen Seite

$$
\mathbb{P}[\Omega_{\varepsilon,c,N}^c] \to 0
$$

gilt, folgt die Behauptung. □

Auf ähnlichen Argumenten (und gleichen Methoden) beruht der Beweis dessen, dass auch eine der Schranken beim Speichern mit "Fehler" scharf war. Dies soll hier nicht wiederholt werden, das entsprechende Resultat lautet

Satz 12.11 *Gilt*

$$M(N) \geq \frac{c(1-2r)^2 N}{\log N},$$

so gilt für jedes zufällig ausgewählte $\tilde{\xi}^\mu \in S_{rN}(\xi^\mu)$ *und jedes feste* μ

$$\mathbb{P}[T_i \tilde{\xi}^\mu = \xi_i^\mu \quad \forall\, i = 1, \ldots, N] \to 0$$

mit $N \to \infty$.

Diese Resultate ergeben ein ziemlich präzises Bild der Speicherkapazität des Hopfield-Modells. Interessanterweise stimmen diese Resultate nicht mit den Simulationen Hopfields (und anderer) überein und auch die Physiker kamen mit der Replica-Methode auf andere Ergebnisse: Alle erhielten eine Speicherkapazität von

$$M = \alpha_c N,$$

wobei $\alpha_c \approx 0,14$ war. Allerdings wurde ein Muster auch dann als gespeichert gewertet, wenn bei der Rekonstruktion kleinere Fehler auftraten (die stören im Prinzip nicht weiter, da αN Muster, die zufällig gewählt werden ($\alpha < 1$) typischerweise einen Hamming-Abstand von $\frac{N}{2}$ haben – Rekonstruktionen mit kleinen Störungen sind also in der Regel eindeutig den Originalmustern zuzuordnen.

Eine mathematische Analyse dieses Sachverhalts geht auf Newman zurück. Verbesserungen seiner Techniken erzielten Lonkianova, Talagrand und Tirozzi.

Satz 12.12 *Es gibt ein* $\alpha_c > 0$, *so dass für*

$$M \leq \alpha_c N,$$

$\varepsilon > 0$ *und* $0 < \delta < \frac{1}{2}$ *existieren, so dass*

$$\mathbb{P}[\bigcap_{\mu=1}^{M} h_N(\xi^\mu, \delta) > H_N(\xi^\mu) + \varepsilon N] \to 1$$

für $N \to \infty$. *Hierbei ist*

$$H_N(\sigma) = -\frac{1}{N} \sum_{i,j} \sum_{\mu=1}^{M} \sigma_i \sigma_j \xi_i^\mu \xi_j^\mu$$

der Standard-Hopfield-Hamiltonian und

$$h_N(\sigma) = \min_{\sigma' \in S_{\delta N}(\sigma)} H_N(\sigma').$$

Bemerkung: Newmans Wert für α_c war $\alpha_c \geq 0,056$. Dies wurde von Lonkianova auf $\alpha_c \geq 0,072$ und von Talagrand auf $\alpha_c \geq 0,08$ verbessert. Der derzeitige "Rekord" ist von Tirozzi: $\alpha_c \geq 0,11$. Die Simulationsschranke von $\alpha_c \geq 0,14$ zu erreichen ist noch nicht gelungen. Insbesondere ist noch größtenteils unklar, was jenseits dieser Schranke geschieht.

Das Theorem besagt dabei grob gesprochen das folgende: Startet man eine Rekonstruktion bei $T = 0$ in ξ^μ, so erhält man zwar nicht notwendig ξ^μ zurück, aber aufgrund des Energiewalls um ξ^μ eine Konfiguration, die nicht weiter als δN von ξ^μ entfernt ist.

Beweis: Wir betrachten das Gegenereignis:

$$\begin{aligned}
& 1 - \mathbb{P}[\bigcap_{\mu=1}^{M} \{h(\xi^\mu, \delta) > H(\xi^\mu) + \varepsilon N\}] \\
= \;& \mathbb{P}[\bigcup_{\mu=1}^{M} \bigcup_{\substack{I \subseteq \{1,\dots,N\} \\ (I)=\delta N}} \{H(\xi_I^\mu) \leq H(\xi^\mu) + \varepsilon N\}] \\
\leq \;& \sum_{\mu=1}^{M} \sum_{I:|I|=\delta N} \mathbb{P}[H(\xi_I^\mu) \leq H(\xi^\mu) + \varepsilon N].
\end{aligned}$$

Hierbei ist

$$(\xi_I^\mu)_i = \begin{cases} \xi_i^\mu, & \text{falls } i \notin I \\ -\xi_i^\mu, & \text{falls } i \in I \end{cases}.$$

Betrachten wir $H(\xi_I^\mu) - H(\xi^\mu)$:

$$\begin{aligned}
& -\frac{1}{N} \sum_{\nu=1}^{M} \sum_{i,j} (\xi_I^\mu)_i (\xi_I^\mu)_j \xi_i^\nu \xi_j^\nu \\
& +\frac{1}{N} \sum_{\nu=1}^{M} \sum_{i,j} \xi_i^\mu \xi_j^\mu \xi_i^\nu \xi_j^\nu \\
= \;& -\frac{1}{N} \sum_{i,j} ((\xi_I^\mu)_i \xi_i^\mu (\xi_I^\mu)_j \xi_j^\mu - \xi_i^\mu \xi_j^\mu \xi_j^\mu) \\
& -\frac{1}{N} \sum_{i,j} \sum_{\nu \neq \mu} (\xi_I^\mu)_i \xi_i^\nu (\xi_I^\mu)_j \xi_j^\nu - \xi_i^\mu \xi_i^\nu \xi_j^\mu \xi_j^\nu.
\end{aligned}$$

Wieder können wir der Einfachheit halber annehmen, dass $\xi_i^\mu \equiv 1 \quad \forall\, i = 1, \dots, N$ gilt. Ferner sei $I = \{1, \dots, \delta N\}$ (und δN eine ganze Zahl). Dann ist

$$\begin{aligned}
H_N(\xi_I^\mu) - H_N(\xi^\mu) \;&=\; -\frac{1}{N} \sum_{i,j} (\xi_I^\mu)_i (\xi_I^\mu)_j - 1 \\
& \quad -\frac{1}{N} \sum_{i,j} \sum_{\nu \neq \mu} (\xi_I^\mu)_i \xi_i^\nu (\xi_I^\mu)_j \xi_j^\nu - \xi_i^\nu \xi_j^\nu.
\end{aligned}$$

Wir sehen, dass ein Summand in der obigen Summe genau dann nicht 0 ist, wenn $i \in I$ und $j \in I^c$ oder $i \in I^c$ oder $j \in I$ ist. In diesem Fall ergibt der Summand gerade $-2\xi_i^\nu \xi_j^\nu$. Also

$$H_N(\xi_I^\mu) - H_N(\xi^\mu) = 4\delta(1-\delta)N + \frac{4}{N} \sum_{i \in I} \sum_{j \in I^c} \sum_{\nu \neq \mu} \xi_i^\nu \xi_j^\nu.$$

Sonst ist zu zeigen, dass

$$\mathbb{P}[\frac{4}{N}\sum_{i\in I}\sum_{j\in I^c}\sum_{\nu\neq\mu}\xi_i^\nu\xi_j^\nu \leq \varepsilon N - 4\delta(1-\delta)N]$$

für geeignetes ε und δ hinreichend schnell gegen 0 geht, denn es wird ja anschließend noch mit M und $\binom{N}{\delta N}$ multipliziert (für die Auswahl der μ und der Mengen I). Mit der üblichen exponentiellen Chebyshev-Ungleichung erhalten wir für alle $t \geq 0$

$$\begin{aligned}
&\mathbb{P}[\frac{4}{N}\sum_{i\in I}\sum_{j\in I^c}\sum_{\nu\neq\mu}\xi_i^\nu\xi_j^\nu \leq \varepsilon N - 4\delta(1-\delta)N] \\
\leq\ & e^{t(\varepsilon N-4\delta(1-\delta)N)}\mathbb{E}e^{-\frac{t4}{N}\sum_{i\in I}\sum_{j\in I^c}\sum_{\nu\neq\mu}\xi_i^\nu\xi_j^\nu} \\
=\ & e^{tN(\varepsilon-4\delta(1-\delta))}(\mathbb{E}\exp(-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1)^{m-1}.
\end{aligned}$$

Bedauerlicherweise kann man nun nicht der üblichen Chebyshev-Maschinerie folgen, denn $\xi_i^1\xi_j^1$ sind nicht unabhängig für $i,j \in I \times I^c$. Allerdings sind sie bedingt unabhängig, wenn man die $\xi_i^1,\ i \in I$ bzw. $\xi_j^1,\ j \in I^c$ fest lässt. Bezeichnen wir mit E_I und $\mathbb{E}_{I^c}$ Erwartungswerte bezüglich von Variablen mit Indizes in I bzw. I^c, so ist

$$\mathbb{E}e^{-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1} = \mathbb{E}_I\mathbb{E}_{I^c}e^{-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1} = \mathbb{E}_I\mathbb{E}_{I^c}.$$

Nun ist

$$\begin{aligned}
\mathbb{E}_{I^c}\exp(-\frac{4t}{N}\sum_{i\in I^c}\sum_{j\in I^c}\xi_i^1\xi_j^1) &= \prod_{j\in I^c}\cosh(\frac{4t}{N}\sum_{i\in I}\xi_i^1) \\
&= \prod_{j\in I^c}\cosh(\frac{4t}{\sqrt{N}}\frac{\sum_{i\in I}\xi_i^1}{\sqrt{N}}) \\
&\leq \prod_{j\in I^c}\exp(\frac{1}{2}(\frac{4t}{\sqrt{N}}\frac{\sum_{i\in I}\xi_i^1}{\sqrt{N}})^2) \\
&= \prod_{j\in I^c}(\mathbb{E}_{Z_j}\exp(\frac{4tz_j}{\sqrt{N}}\frac{\sum_{i\in I}\xi_i^1}{\sqrt{N}}).
\end{aligned}$$

Hierbei seien die z_j unabhängige Gaußsche Zufallsvariable, $\mathcal{N}(0,1)$-verteilt, und wir haben die bekannten Beziehungen

$$\cosh(x) \subseteq e^{\frac{1}{2}x^2}$$

und

$$e^{x^2/2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2+xz}dz$$

ausgenutzt. Also

$$\mathbb{E}\exp(-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1) \leq \mathbb{E}_{(z_j)_{j\in I^c}}\mathbb{E}_I\exp(\sum_{i\in I}\sum_{j\in I^c}4t\frac{z_j}{\sqrt{N}}\frac{\xi_i^1}{\sqrt{N}}).$$

Nun vertauschen wir die Rollen von I und I^c und erhalten für weitere i.i.d. $\mathcal{N}(0,1)$-verteilte Zufallsvariable

$$\begin{aligned}\mathbb{E}\exp(-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1) &\leq \mathbb{E}_{(z_j)_{j=1,\dots,N}}\exp(4t\sum_{i\in I}\frac{z_i}{\sqrt{N}}\sum_{j\in I^c}\frac{z_j}{\sqrt{N}}) \\ &= \mathbb{E}_z\exp(4t\sqrt{\delta(1-\delta)}\sum_{i\in I}\frac{z_i}{\sqrt{\delta N}}\sum_{j\in I^c}\frac{z_j}{\sqrt{(1-\delta)N}}).\end{aligned}$$

Nun sind $(\sum_{i\in I}\frac{z_i}{\sqrt{\delta N}},\ \sum_{j\in I^c}\frac{z_j}{\sqrt{(1-\delta)N}})$ in Verteilung gleich einem Vektor $(Z,\tilde{Z})$, wobei z und $\tilde{z}$ i.i.d. $\mathcal{N}(0,1)$-verteilt sind. Also

$$\mathbb{E}\exp(-\frac{4t}{N}\sum_{i\in I}\sum_{j\in I^c}\xi_i^1\xi_j^1) \leq \mathbb{E}_{z,\tilde{z}}e^{4t\sqrt{\delta(1-\delta^2)}Z\tilde{z}}.$$

Die rechte Seite existiert nur für hinreichend kleines t, lässt sich in dem Fall aber explizit berechnen. In der Tat ist für genügend kleines γ

$$\begin{aligned}\mathbb{E}_{z,\tilde{z}}e^{\gamma z\tilde{z}} &= \frac{1}{2\pi}\int\int e^{\gamma z\tilde{z}-\frac{1}{2}z^2-\frac{1}{2}\tilde{z}^2}dzd\tilde{z} \\ &= \frac{1}{2\pi}\int\int e^{-\frac{1}{2}(z^2-2\gamma z\tilde{z}+\gamma^2\tilde{z}^2)}dz\ e^{\frac{\gamma^2\tilde{z}^2}{2}}e^{-\tilde{z}^2/2}d\tilde{z} \\ &= \frac{1}{\sqrt{2\pi}}\int e^{-\frac{\tilde{z}^2}{2}(1-\gamma^2)}d\tilde{z} = \frac{1}{\sqrt{1-\gamma^2}}.\end{aligned}$$

Nach einer Variablentransformation

$$\tilde{t} = t\sqrt{\delta(1-\delta)}$$

ist also

$$\begin{aligned}&\mathbb{P}(\frac{1}{N}\sum_{i\in I}\sum_{j\in I^c}\sum_{\nu\neq\mu}\xi_i^\nu\xi_j^\nu \leq \varepsilon N - 4\delta(1-\delta)N] \\ \leq\ & \exp\left(\frac{tN}{\sqrt{\delta(1-\delta)}}^{(\varepsilon N-4\delta(1-\delta))}\right)\left(\frac{1}{\sqrt{1-16t}}\right)^M \\ =\ & \exp\left(\frac{tN}{\sqrt{\delta(1-\delta)}}^{(\varepsilon N-4\delta(1-\delta))} - \frac{M}{2}\log(1-16t)\right).\end{aligned}$$

Nun lässt sich t so wählen, dass dieser Ausdruck im wesentlichen durch

$$\exp(-\text{const}\,N\sqrt{\delta(1-\delta)})$$

gegeben ist. Nun ist

$$\begin{aligned}&\sum_{\mu=1}^M\sum_{I:|I|=\delta N}\mathbb{P}(H_N(\xi_I^\mu)-H_N(\xi^\mu)\leq\varepsilon N) \\ \leq\ & M\binom{N}{\delta N}\exp(-\text{const.}\,N\sqrt{\delta(1-\delta)}) \\ \cong\ & M\,e^{N(\delta\log\delta+(1-\gamma)\log(1-\delta))-\text{const.}\,N\sqrt{\delta(1-\gamma)}},\end{aligned}$$

wobei man für die Abschätzung der Binomialkoeffizienten die Stirlingsche Formel benutzt. Nun ist für $\delta \ll 1$

$$\delta \log \delta + (1-\delta)\log(1-\delta) \approx \delta \log \delta$$

und $\sqrt{\delta(1-\delta)} \approx \sqrt{\delta^1}$.

Nun überprüft man schnell, dass

$$\delta \log \delta|_{\delta=0} = 0 \quad \text{und} \quad \sqrt{\delta^1}|_{\delta=0} = 0$$

sowie

$$\frac{\partial}{\partial \delta}(\delta \log \delta) = 1 + \log \delta < \frac{\partial}{\partial \delta}\sqrt{\delta}$$

für eine Umgebung von $\delta = 0$, somit können wir ein hinreichend kleines δ finden, so dass es für dieses δ eine Konstante C gibt mit

$$\mathbb{P}(\bigcup_\mu \bigcup_I H_N(\xi_I^\mu) - H_N(\xi^\mu) \leq \varepsilon N) \leq Me^{-Nc}.$$

Dies konvergiert exponentiell schnell gegen 0. □

12.3 Die Thermodynamik des Hopfield-Modells

Die Thermodynamik des Hopfield-Modells kann wieder über verschiedene Zugänge studiert werden. Der einfachste ist die Betrachtung der freien Energie. Dies wurde z. B. von Vermet in seiner Dissertation durchgeführt. Eine andere ist die Analyse der Gibbs-Maße, die im Regime $M \ll N$ sehr einer Überlagerung von Gibbs-Maßen im Curie-Weiss-Modell ähnelt. Diese Analyse stammt i.w. aus einer Reihe von Arbeiten von Bovier und Gayroid.

Satz 12.13 *Sei* $\alpha = \lim \frac{M}{N}$. *Dann gilt:*

1. *Ist* $\beta < 1$, *so folgt*

$$\lim_{\substack{N\to\infty \\ \alpha\geq 0}} \bar{f}_{N,M,\beta} = \lim_{\substack{N\to\infty \\ \alpha\geq 0}} -\frac{1}{\beta N}\log \mathbb{E}Z_{N,M,\beta} = -\frac{\log 2}{\beta} + \frac{\alpha}{2\beta}\log(1-\beta) + \frac{\alpha}{2}.$$

2. *Ist* $\beta > 1$, *so ist*

$$\lim_{M,N\to\infty} \bar{f}_{N,M,\beta} = -\infty$$

und

$$\lim_{N\to\infty} \bar{f}_{N,M,\beta} = -\frac{\log 2}{\beta} + \frac{m}{\beta}\inf_{x\in\mathbb{R}}\{\frac{x^2}{2}\log\cosh(\sqrt{\beta}x)\}$$

endlich.

Beweis: Die zentrale Idee ist es, Gaußsche Hilfsvariablen einzuführen, um die quadratische Form in der Hamiltonfunktion zu linearisieren. Diese Idee haben wir schon im Falle der Hubbard-Stratonovich-Transformation kennengelernt. Nun ist

$$Z_{N,M,\beta} = \sum_{\sigma \in \{-1,+1\}^N} \prod_{\mu=1} e^{\frac{\beta}{2N} <\xi^\mu, \sigma>^2},$$

wobei $< \cdot, \cdot >$ das Standardskalarprodukt im $\mathbb{R}^{N'}$ bezeichnet. Wegen

$$e^{x^2/2a} = \sqrt{\frac{a}{2\pi}} \int_{\mathbb{R}} e^{-\frac{a}{2}t^2 + tx} dt$$

folgt

$$Z_{N,M,\beta} = (\frac{N}{2\pi})^{M/2} \int_{\mathbb{R}^M} \sum_{\sigma \in \{\pm 1\}^N} e^{-\frac{N}{2}\sum_{\mu=1}^M x_\mu^2 + \sqrt{\beta}\sum_{\mu=1}^M x_\mu \sum_{i=1}^N \sigma_i \xi_i^\mu} dx,$$

wobei $x = (x_1, \ldots, x_M)$ ist. Nun kann man (wie bei der Hubbard-Stratonovich-Transformierten) die Integration über die σ_i schnell ausführen und erhält mit Fubini

$$\mathbb{E}Z_{N,M,\beta} = 2^N [\sqrt{\frac{N}{2\pi}} \int_{\mathbb{R}} e^{-\frac{Nx^2}{2} + N \log\cosh(\sqrt{\beta}x)}]^M.$$

Die Laplacesche Methode sagt nun, dass sich dieses Integral asymptotisch wie das Maximum des Integranden verhält. Wir bekommen somit

$$\bar{f}_{N,M,\beta} = -\frac{\log 2}{\beta} - \frac{M}{2\beta N}\log\frac{N}{2\pi} - \frac{M}{\beta}\inf_{x\in\mathbb{R}}\{\frac{x^2}{2} - \log\cosh\sqrt{\beta}x\} + R_N$$

mit $R_N \to 0$ für $N \to \infty$. Nun kennen wir schon aus dem Studium des Curie-Weiss-Modells, dass für $\beta > 1$

$$\inf_{x\in\mathbb{R}}\{\frac{x^2}{2} - \log\cosh(\sqrt{\beta}x)\} = c_\beta < 0$$

gilt. Also gilt für jede Wahl von $M(N)$ mit

$$\lim_{N\to\infty} M(N) = +\infty,$$

dass

$$\lim_{\substack{N\to\infty \\ M\to\infty}} \bar{f}_{N,M,\beta} = +\infty.$$

Ist umgekehrt $M(N) \equiv M$ endlich, so überleben die beiden äußeren Terme

$$\lim_{N\to\infty} \bar{f}_{N,M,\beta} = -\frac{\log 2}{\beta} - \frac{M}{\beta} - \frac{M}{\beta}\inf_{x\in\mathbb{R}}\{\frac{x^2}{2} - \log\cosh\sqrt{\beta}x\}.$$

Ist umgekehrt $\beta < 1$, so gilt

$$\inf_{x\in\mathbb{R}}\{\frac{x^2}{2} - \log\cosh(\sqrt{\beta}x)\} = 0$$

und somit

??

$\square$

Dieses Resultat klärt das Verhalten der sogenannten "annealten" freien Energie. Zuerst ist das Verhalten der "gequenchten" freien Energie

$$\lim_{N\to\infty} f_{N,M,\beta} = \lim_{N\to\infty} -\frac{1}{\beta N} \log Z_{N,M,\beta}$$

bzw.

$$\lim_{N\to\infty} \mathbb{E}\, f_{N,M,\beta} = \mathbb{E} \lim_{N\to\infty} -\frac{1}{\beta N} \log Z_{N,M,\beta}$$

interessanter und wichtiger. Wir zitieren hier nur ein Resultat von Bovier, Gayrardöö, Picco bzw. vermet.

Satz 12.14 *Sei* $\lim_{N\to\infty} \frac{M(N)}{N} = \alpha \in [0,1)$ *und* $\delta \in (0,1)$*, so dass*

$$\delta - 4\sqrt{\alpha}(1-\delta) > 0.$$

Sei

$$\begin{aligned} \Phi_\delta : (0,\infty) \times \mathbb{R} &\to \mathbb{R} \\ (\beta, h) &\mapsto -\frac{1}{\beta} \log\cosh(\beta h) + \frac{1-\delta}{2} h^2. \end{aligned}$$

Definiere die Funktion ψ_1 *und* ψ_2 *mittels*

$$\psi_1(\beta,\alpha,\delta) = \min_{h\in\mathbb{R}} \Phi_\delta(\beta,h) - \frac{\log 2}{\beta} + \frac{\alpha}{\beta} \log(\delta - 4\sqrt{\alpha}(1-\delta))$$

und

$$\psi_2(\beta,\alpha) = \min_{h\in\mathbb{R}} \Phi_0(\beta,h) - \frac{\log 2}{\beta}.$$

Dann gilt für jedes $\beta > 0$*:*

1. *Für* $\mathbb{P}$*-fast alle* $(\xi_i^\mu)_{i,\mu}$ *gilt*

$$\psi_1(\beta,\alpha,\delta) \le \liminf_{\substack{N\to\infty \\ M/N\to\alpha}} f_{N,M,\beta} \le \limsup_{\substack{N\to\infty \\ M/N\to\alpha}} f_{N,M,\beta} \le \psi_2(\beta,\alpha).$$

2. *Es gilt*

$$\psi_1(\beta,\alpha,\delta) \le \liminf_{\substack{N\to\infty \\ M/N\to\alpha}} \mathbb{E}\, f_{N,M,\beta} \le \limsup_{\substack{N\to\infty \\ M/N\to\alpha}} \mathbb{E}\, f_{N,M,\beta} \le \psi_2(\beta,\alpha).$$

Aus diesm Resultat lässt sich durch $\alpha = 0$ und $\delta \downarrow 0$ folgern, dass die freie Energie im Hopfield-Modell für "kleine Ladyszahlen", d. h. $\alpha = 0$, gegen die freie Energie des Curie-Weiss-Modells konvergiert.

Eine weitere wichtige Eigenschaft des Hopfield-Modells und vieler anderer Spingläser ist die sogenannte Selbstmittlungseigenschaft bestimmter makroskopischer Größen. Diese Eigenschaft besagt im wesentlichen, dass eine solche makroskopische Größe fast sicher oder in Wahrscheinlichkeit gegen ihren Erwartungswert konvergiert – ist also im wesentlichen ein Gesetz der großen Zahlen. Diese Resultate haben eine lange Geschichte in der Spinglastheorie. Wir werden hier nur die Selbstmittlungseigenschaft der freien Energie im Hopfield-Modell und dies auch nicht unter den best-möglichen Bedingungen. Dafür lernen wir aber eine Methode kennen, die sich auch anderenorts beim Studium von Spingläsern bewährt hat.

Satz 12.15 *Sei $\beta > 0$ und*

$$M(N) \le \gamma(N)\sqrt{N} \quad \textit{mit} \quad \gamma(N) \underset{N\to\infty}{\longrightarrow} 0.$$

Dann gilt

$$\lim_{\substack{N\to\infty \\ M\to\infty}} \mathbb{E}[(f_{N,M,\beta} - \mathbb{E}\, f_{N,M,\beta})^2] = 0.$$

Beweis: Wähle $t \in [0,1]$ und betrachte anstelle des Original-Hamiltonians

$$H_N(\sigma) = -\frac{1}{N}\sum_{\mu=1}^{M}\sum_{i,j} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j$$

den mofidizierten Hamiltonian

$$\tilde{H}_N(\sigma, k, t) = -\frac{1}{N}\sum_{\mu=1}^{M}\sum_{\substack{i,j=1 \\ i,j\neq k}}^{N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j - \frac{t}{N}\sum_{\mu=1}^{M}\sum_{i=1}^{N} \xi_i^\mu \xi_k^\mu \sigma_i \sigma_k.$$

Hier ist $1 \le k \le N$. Offenbar ist

$$\tilde{H}_N(\sigma, k, 1) = H_N(\sigma).$$

Für die Hamiltonfunktion $\tilde{H_N}$ definieren wir die modifizierte Zustandssumme

$$\tilde{Z}_{N,M,\beta}(h,t) = \sum_{\sigma\in\{\pm 1\}^N} \exp(-\beta \tilde{H}_N(\sigma, k, t)).$$

Die entsprechende freie Energie ist

$$\tilde{f}_{N,M,\beta}(h,t) = -\frac{1}{\beta N}\log \tilde{Z}_{N,M,\beta}(k,t)$$

und das modifizierte Gibbs-Maß ist

$$\tilde{\mu}_{N,M,\beta}(\cdot, k, t) = \frac{\sum_2 e^{-\beta H_N(\sigma, kt)} \mathbb{1}_{(\cdot)}(\sigma)}{\tilde{Z}_{N,M,\beta}(k,t)}.$$

Nun führen wir die σ-Algebra

$$\mathcal{F}_k = \sigma(\xi_1^\mu, \ldots, \xi_k^\mu,\, \mu = 1, \ldots, M)$$

ein. Hierbei ist

$$\mathcal{F}_0 = \{\emptyset, \Omega\}.$$

Offenbar gilt

$$\begin{aligned}
f_{N,M,\beta} - \mathbb{E}\, f_{N,M,\beta} &= \sum_{k=1}^{N} [\mathbb{E}(f_{N,M,\beta}|\mathcal{F}_k) - \mathbb{E}(f_{N,M,\beta}|\mathcal{F}_{k-1})] \\
&= \sum_{k=1}^{N} [\mathbb{E}(\tilde{f}_{N,M,\beta}(k,1)|\mathcal{F}_k) - \mathcal{E}(\tilde{f}_{N,M,\beta}(k,1)|\mathcal{F}_{k-1})].
\end{aligned}$$

Man kann nachrechnen, dass die Summanden für verschiedene k orthogonal (im L^2-Sinne) sind. Nun gilt trivialerweise für $g \in e^1([0,1))$

$$g(1) = g(0) = \int_0^1 \frac{dg}{dt}(t)dt.$$

Für die Anwendung folgern wir zunächst, dass

$$\mathbb{E}(\tilde{f}_{N,M,\beta}(k,0)|\mathcal{F}_k) - \mathbb{E}(\tilde{f}_{N,M,\beta}(k,0)|\mathcal{F}_{k-1}) = 0,$$

weil der Anteil von $\tilde{f}_{N,M,\beta}(k,t)$, der messbar bezüglich $\mathcal{F}_k$ abr nicht bezüglich $\mathcal{F}_{k-1}$ ist, durch

$$-\frac{t}{N} \sum_{\mu=1}^{M} \sum_{i \neq k} \xi_i^\mu \xi_k^\mu \sigma_i \sigma_k$$

gegeben ist, also für $t = 0$ verschwindet. Also bleibt noch der Ableitungsterm zu kontrollieren; dies ergibt

$$\begin{aligned}
f_{N,M,\beta} \quad - \quad & \mathbb{E}\, f_{N,M,\beta} = \frac{1}{N} \sum_{k=1}^{N} \int_0^1 \sum_{\mu=1}^{M} \sum_{\substack{i=1 \\ i \neq k}}^{N} \\
& [\mathbb{E}(\xi_i^\mu, \xi_k^\mu \tilde{\mu}_{N,M,\beta}(\sigma_i\sigma_k, k, t)|\mathcal{F}_{k-1}) \\
- \quad & \xi_k^\mu\, \mathbb{E}(\xi_i^\mu\, \tilde{\mu}_{N,M,\beta}(\sigma_i\sigma_k, k, t)|\mathcal{F}_k]dt.
\end{aligned}$$

Da trivialerweise

$$\tilde{\mu}_{N,M,\beta}(\sigma_i\sigma_k, k, t) \leq 1,$$

folgt aus der Orthogonalität der Summanden

$$\mathbb{E}[(f_{N,M,\beta} - \mathbb{E}\, f_{N,M,\beta}^2] \leq \frac{4M^2}{N}.$$

Dies konvergiert nach Voraussetzung gegen 0. □

Wir haben schon bei der Analyse der freien Energie gesehen, dass das Hopfield-Modell für kleine M Ähnlichkeit mit dem Curie-Weiss-Modell aufweist. Um dies Ähnlichkeit noch weiter aufzuzeigen, führt man Schritte durch, die auch bei der Analyse des Curie-Weiss-Modells hilfreich waren, mit N ?? die Hubbard-Stratonovich-Transformation. Da die Beweise technisch entweder sehr ähnlich zu den Beweisen im Curie-Weiss-Fall sind oder sehr aufwendig, geben wir hier nur die Resultate. Dazu ist es zunächst wichtig zu bemerken, dass die Hamilton-Funktion im Hopfield-Modell eine Funktion des Overlaps

$$m_N(\sigma) = (m_N^\mu(\sigma))_{m=1}^M$$

mit

$$m_N^\mu(\sigma) = \frac{1}{N}\sum_{i=1}^N \sigma_i \xi_i^\mu$$

ist. In der Tat gilt ja

$$H_N(\sigma) = -N\|m_N\|_2^2,$$

wobei $\|\cdot\|_2^2$ das Quadrat der Euklidischen Norm in $\mathbb{R}^\mu$ ist. Daher ist die Verteilung von m_N unter $\mu_{N,\beta}$ für die Analyse des Modells relevant:

$$\nu_{N,\beta} := \mu_{N,\beta} \circ (m_N)^{-1}.$$

Weiter wollen wir auch die Verteilung des Overlaps mit einem äußeren Magnetfeld in eine bestimmte Richtung studieren. Diese sei o.B.d.A. ξ^1. Mathematisch drückt sich dies in der Hamiltonfunktion als

$$H_{N,n}(\sigma) = H_N(\sigma) - h\sum_{i=1}^N \sigma_i \xi_i^1$$

aus. Das zugehörige Gibbs-Maß ist

$$\mu_{N,\beta,h}(\sigma) = \frac{e^{-\beta H_{N,h}(\sigma)}}{Z_{N,\beta,h}}$$

mit

$$Z_{N,\beta,h} = \sum_{\sigma\in\{\pm 1\}} e^{-\beta H_{N,h}(\sigma)}.$$

Schließlich ist die Verteilung von m_N unter $\mu_{N,\beta,h}$

$$\nu_{N,\beta,h} = \mu_{N,\beta,h} \circ m_N^{-1}.$$

Die Analyse der $\nu_{N,\beta}$ bzw. $\nu_{N,\beta,h}$ verläuft wieder über die Hubbard-Stratonovich-Transformierte. Für diese erhält man z. B. für $\nu_{N,\beta}$:

Lemma 12.16 *Sei*

$$Q_{N,\beta} := \nu_{N,\beta} * \mathcal{N}(0, \frac{1}{\beta N}),$$

wobei in diesem Fall $\mathcal{N}(0, \frac{1}{\beta N})$ eine M-dimensionale Gauß-Verteilung mit Erwartungswertvektor 0 und Kovarianzmatrix

$$C = \frac{1}{\beta N} Id_M$$

ist. Dann ist $Q_{N,\beta}$ stetig bezüglich $\lambda\!\!\lambda^\mu$ mit Dichte

$$f_{N,\beta}(x) = \frac{\exp(-N\beta\Phi_{\beta,N}(x))}{\Xi_{N,\beta}}$$

mit

$$\Phi_{\beta,N}(x) = \frac{1}{2}\sum_{\mu=1}^{M} x_\mu^2 - \frac{1}{\beta N}\sum_{i=1}^{N} \log\cosh(\beta < \xi_i, x >),$$

wo $< \cdot, \cdot >$ das Skalarprodukt im $\mathbb{R}^M$ ist.

Beweis: Der Beweis funktioniert wie im Curie-Weiss-Modell. □

Für endliches, von M unabhängiges M konvergiert $\Phi_{N,\beta}(x)$ fast sicher gegen:

$$\Phi_\beta(x) := \frac{1}{2}\|x\|_2^2 - \frac{1}{\beta}\mathbb{E}\log\cosh(\beta < \xi_1, z >),$$

wächst aber M, so müssen die Fluktuationen von $\Phi_{N,\beta}(x)$ und $\Phi_\beta(x)$ kontrolliert werden. Danach kann man die Minima von $\Phi_\beta(\cdot)$ analysieren und versuchen, die Varadhan-Laplace-Methode anzuwenden. (Hierbei ist eine Schwierigkeit, dass untypischerweise in diesem Fall die Dimension des Raumes mit N wächst – diese Schwierigkeit ist jedoch nicht unüberwindbar.) Auf diese Weise lassen sich viele Resultate über das Hopfield-Modell erzielen (der interessierte Leser sei auf eine ganze Serie von Arbeiten von Bovier und Gayrord verwiesen, ebenso etliche Arbeiten von Talagrand, daneben Arbeiten von Gentz, Löwe u. a.), unter anderem auch das folgende Resultat, das sehr an den Phasenübergang im Curie-Weiss-Modell erinnert:

Satz 12.17 *(Bovier, Gayrord, Picco): Sei $M(N)$ dergestalt, dass $\frac{M(N)}{N} \to 0$ mit $N \to \infty$. Dann gilt für reelles h:*

$$\lim_{h\downarrow 0} \nu_{N,\beta,h} \Rightarrow \delta_{m^*(\beta)e_1},$$

wobei e_1 den ersten Einheitsvektor im $\mathbb{R}^M$ bezeichnet und $m^*(\beta)$ die größte Lösung der Curie-Weiss-Gleichung

$$m = \tanh(\beta m).$$

Der relativ längliche Beweis soll hier nicht geführt werden.

Auf ähnliche Weise lassen sich zentrale Grenzwertsätze u.v.m. für $M \ll N$ auf das Hopfield-Modell übertragen. Auch für $M = \alpha N$ für sehr kleines α sind noch Konvergenzaussagen erhältlich. Leider fehlen aber bis zum heutigen Zeitpunkt jedwede Ergebnisse für den interessanten Spinglasfall

$$M = \alpha N$$

und α “groß”, z. B. $\alpha = 1$.

13 Einige Ergebnisse im SK-Modell

Schon in Kapitel 10 haben wir angedeutet, dass das SK-Modell schwer zu analysieren ist (jedenfalls von einem mathematischen Standpunkt), dass einige erste wichtige Resultate in dieser Richtung aber sei neuestem zur Verfügung stehen. Diese Resultate von Talagrand beruhen u. a. auf Ideen, die wir auch im Kontext des Hopfield-Modells kennengelernt haben, die aber im einzelnen zu technisch sind, um hier im Detail vorgeführt zu werden. Wir wollen uns auf ein Resultat von Aizenman, Lebowitz und Ruelle konzentrieren, das die paramagnetische, d. h. die Hochtemperaturphase des SK-Modells analysiert und dort den zentralen Grenzwertsatz für die freie Energie beweist. Hierfür fasen wir $Z_{N,\beta}$ bzw.

$$f_N := \frac{1}{N} \log Z_{N,\beta}$$

als Zufallsvariablen in den J_{ij} auffassen (zur Erinnerung:

$$Z_{N,\beta} := \frac{1}{2^N} \sum_{\sigma \in \{-1,+1\}^N} e^{\beta/\sqrt{N} \sum_{i<j} \sigma_i \sigma_j J_{ij}}).$$

Man beginnt wieder mit dem üblichen

$$e^x = \cosh(x)(1 + \tanh(x)).$$

Wendet man dies auf die Exponentialfunktion in der Definition von $Z_{N,\beta}$ an, so ergibt sich

$$\begin{aligned} Z_{N,\beta} &= \mathbb{E}_\sigma \prod_{i<j} \cosh(\frac{\beta}{\sqrt{N}} J_{ij}\sigma_i\sigma_j)(1 + \tanh(\frac{\beta}{\sqrt{N}}\sigma_i\sigma_j J_{ij})) \\ &= \prod_{i<j} \cosh(\frac{\beta}{\sqrt{N}} J_{ij}) \mathbb{E}_\sigma \prod_{i<j} (1 + \sigma_i\sigma_j th(\frac{\beta}{\sqrt{N}} J_{ij}) \end{aligned}$$

aufgrund der Symmetrien von $\cosh(\cdot)$ und $\tanh(\cdot)$. Somit ist

$$\log Z_{N,\beta} = sum_{i<j} \log\cosh(\frac{\beta}{\sqrt{N}} J_{ij}) + \log \mathbb{E}_\sigma \prod_{i<j} (1 + \sigma_i\sigma_j th(\frac{\beta}{\sqrt{N}} J_{ij}).$$

Der erste Term lässt sich mittels

$$\log\cosh(x) = \frac{x^2}{2} + 0(x^4)$$

als

$$\frac{\beta^2}{2N} \sum_{i<j} J_{ij}^2 + 0(\sum_{i<j} \frac{\beta^4}{N^2} J_{ij}^4)$$

umschreiben. Der letzte Term verschwindet in L^2. Die erste Summe besteht aus $\binom{N}{2}$ unabhängigen Summanden mit Erwartungswert 1. Zieht man diesen ab, so ergibt sich

$$\frac{\beta^2}{2N} \sum_{i<j} J_{ij}^2 = \frac{\beta^2}{2N} \sum_{i<j} J_{ij}^2 - 1 + \frac{\beta^2 N}{4}.$$

Die Summe rechts konvergiert nach dem gewöhnlichen zentralen Grenzwertsatz gegen eine Normalverteilung mit Erwartungswert 0 und Varianz

$$\beta^4 \mathbb{E}(J_{ij}^2 - 1)^2 = \beta^4/8(\mathbb{E}J_{ij}^4 - 1).$$

Der Term $N\beta^2/4$ kann nun als

$$\log \mathbb{E}\, Z_{N,\beta}$$

aufgefasst werden, so dass wir eigentlich dabei sind, die Fluktuationen von

$$\log \frac{Z_{N,\beta}}{\mathbb{E}Z_{N,\beta}}$$

zu analysieren. Es bleibt die Auseinandersetzung mit

$$\log \mathbb{E}_\sigma \prod_{i=j} (1 + \sigma_i \sigma_j \tanh(\frac{\beta}{\sqrt{N}} J_{ij})).$$

Wir betrachten den Term ohne den Logarithmus und beachten, dass sich das Produkt ausführen lässt als

$$\mathbb{E}_\sigma \prod_{i<j} (1 + \sigma_i \sigma_j th(\frac{\beta}{\sqrt{N}} J_{ij})) = \mathbb{E}_\sigma \sum_\Gamma \bar{u}(\Gamma).$$

Dabei sind die Γ Graphen auf der Knotenmenge $\{1, \ldots, N\}$

$$\bar{u}(\Gamma) = \prod_{e \in \Gamma} \sigma_{i_1} \sigma_{i_2} \ldots \sigma_{i_{2n-1}} \sigma_{i_{2k}} \tanh(\frac{\beta}{\sqrt{N}} J_e),$$

e sind die Kanten von Γ und $i_1, i_2, \ldots, i_{2^k}$ die (nicht-notwendig verschiedenen) Knoten Da $\mathbb{E}\sigma_i^e = 0$ für ungerades ℓ die σ_i unabhängig sind, folgt

$$\mathbb{E}_\sigma \sum_\Gamma \bar{w}(\Gamma) = \sum_{\Gamma \partial \Gamma = \emptyset} w(\Gamma).$$

Dabei bedeutet $\partial \Gamma = \emptyset$, dass alle Knoten in Γ gerade häufig vorkommen und

$$w(\Gamma) = \prod_{e \in \Gamma} \tanh(\frac{\beta}{\sqrt{N}} J_e).$$

Offenbar spielen die Graphen Γ mit $\partial \Gamma = \emptyset$ eine zentrale Rolle. Man sieht schnell über Induktion, dass so ein Γ sich als

$$\Gamma = \gamma_1 \circ \ldots \circ \gamma_n$$

für ein n und Kreise $\gamma_1, \ldots, \gamma_n$ schreiben lässt. Es liegt daher nahe, die Summe $\sum_{\gamma \text{ Kreis}} w(\gamma)$ zu untersuchen. Wir wollen für sie eine Gaußsche Limes-Verteilung nachweisen. Hierbei sind allerdings lange Kreise aus technischen Gründen lästig. Daher betrachten wir

$$F_{\leq k} = \sum_{\gamma : |\gamma| \leq h} w(\gamma)$$

(hierbei ist $|\gamma|$ die Anzahl der Kanten eines Kreises γ), analysieren $F_{\leq k}$ und lassen dann $k \to \infty$ gehen (dieses Vorgehen muss natürlich gerechtfertigt werden). Wir wollen zeigen, dass $F_{\leq k}$ eine Normalverteilung zum Limes haben. Die Idee dabei ist die Momentenmethode, d. h. wir zeigen, dass die Momente von $F_{\leq k}$ gegen die Momente einer Zufallsvariablen $G \sim \mathcal{N}(0, \sigma^2)$ mit geeigneten $\sigma^2 > 0$ konvergieren. Dies ist ausreichend, da die Normalverteilung eindeutig durch ihre Momente bestimmt ist. Für diese Momente von G gilt:

$$\begin{aligned} \mathbb{E}\, G^n &= 0, && \text{falls } n \text{ ungerade ist} \\ \text{und} \quad \mathbb{E}\, G^n &= (n-1)!!(\sigma^2)^{n/2}, && \text{falls } n \text{ gerade ist.} \end{aligned}$$

Also berechnen wir

$$\|F_{\leq k}\|_2^2 = \sum_{j=3}^{k} \sum_{\gamma:|\gamma|=j} \|w(\gamma)|_2^2.$$

Da für $\gamma \neq \gamma'$ gilt

$$\mathbb{E}_J(w(\gamma)w(\gamma')) = 0$$

(da zumindest eine Kante von γ und γ' verschieden ist, $\mathbb{E}_J\, J_{ij} = 0$ und $\tanh(\cdot)$ ungerade ist), die $w(\gamma')$ also paarweise unkorreliert sind. Da

Anhang A

On teaching mathematics by V.I. Arnold

This is an extended text of the address at the discussion on teaching of mathematics in Palais de Découverte in Paris on 7 March 1997.

Mathematics is a part of physics. Physics is an experimental science, a part of natural science. Mathematics is the part of physics where experiments are cheap.

The Jacobi identity (which forces the heights of a triangle to cross at one point) is an experimental fact in the same way as that the Earth is round (that is, homeomorphic to a ball). But it can be discovered with less expense.

In the middle of the twentieth century it was attempted to divide physics and mathematics. The consequences turned out to be catastrophic. Whole generations of mathematicians grew up without knowing half of their science and, of course, in total ignorance of any other sciences. They first began teaching their ugly scholastic pseudo-mathematics to their students, then to schoolchildren (forgetting Hardy's warning that ugly mathematics has no permanent place under the Sun).

Since scholastic mathematics that is cut off from physics is fit neither for teaching nor for application in any other science, the result was the universal hate towards mathematicians - both on the part of the poor schoolchildren (some of whom in the meantime became ministers) and of the users.

The ugly building, built by undereducated mathematicians who were exhausted by their inferiority complex and who were unable to make themselves familiar with physics, reminds one of the rigorous axiomatic theory of odd numbers. Obviously, it is possible to create such a theory and make pupils admire the perfection and internal consistency of the resulting structure (in which, for example, the sum of an odd number of terms and the product of any number of factors are defined). From this sectarian point of view, even numbers could either be declared a heresy or, with passage of time, be introduced into the theory supplemented with a few "ideal" objects (in order to comply with the needs of physics and the real world).

Unfortunately, it was an ugly twisted construction of mathematics like the one above which predominated in the teaching of mathematics for decades. Having originated in France, this pervertedness quickly spread to teaching of foundations of mathematics, first to university students, then to school pupils of all lines (first in France, then in other countries, including Russia).

To the question "what is 2 + 3" a French primary school pupil replied: "3 + 2, since addition is commutative". He did not know what the sum was equal to and could not even understand what he was asked about!

Another French pupil (quite rational, in my opinion) defined mathematics as follows: "there is a square, but that still has to be proved".

Judging by my teaching experience in France, the university students' idea of mathematics (even of those taught mathematics at the École Normale Supérieure - I feel sorry most of all for these obviously intelligent but deformed kids) is as poor as that of this pupil.

For example, these students have never seen a paraboloid and a question on the form of the surface given by the equation $xy = z^2$ puts the mathematicians studying at ENS into a stupor. Drawing a curve given by parametric equations (like $x = t^3 - 3t,\ y = t^4 - 2t^2$) on a plane is a totally impossible problem for students (and, probably, even for most French professors of mathematics).

Beginning with l'Hospital's first textbook on calculus ("calculus for understanding of curved lines") and roughly until Goursat's textbook, the ability to solve such problems was considered to be (along with the knowledge of the times table) a necessary part of the craft of every mathematician.

Mentally challenged zealots of "abstract mathematics" threw all the geometry (through which connection with physics and reality most often takes place in mathematics) out of teaching. Calculus textbooks by Goursat, Hermite, Picard were recently dumped by the student library of the Universities Paris 6 and 7 (Jussieu) as obsolete and, therefore, harmful (they were only rescued by my intervention).

ENS students who have sat through courses on differential and algebraic geometry (read by respected mathematicians) turned out be acquainted neither with the Riemann surface of an elliptic curve $y^2 = x^3 + ax + b$ nor, in fact, with the topological classification of surfaces (not even mentioning elliptic integrals of first kind and the group property of an elliptic curve, that is, the Euler-Abel addition theorem). They were only taught Hodge structures and Jacobi varieties!

How could this happen in France, which gave the world Lagrange and Laplace, Cauchy and Poincaré, Leray and Thom? It seems to me that a reasonable explanation was given by I.G. Petrovskii, who taught me in 1966: genuine mathematicians do not gang up, but the weak need gangs in order to survive. They can unite on various grounds (it could be super-abstractness, anti-Semitism or "applied and industrial" problems), but the essence is always a solution of the social problem – survival in conditions of more literate surroundings.

By the way, I shall remind you of a warning of L. Pasteur: there never have been and never will be any "applied sciences", there are only applications of sciences (quite useful ones!).

In those times I was treating Petrovskii's words with some doubt, but now I am being more and more convinced of how right he was. A considerable part of the super-abstract activity comes down simply to industrialising shameless grabbing of discoveries from discoverers and then systematically assigning them to epigons-generalizers. Similarly to the fact that America does not carry Columbus's name, mathematical results are almost never called by the names of their discoverers.

In order to avoid being misquoted, I have to note that my own achievements were for some unknown reason never expropriated in this way, although it always happened to both my teachers (Kolmogorov, Petrovskii, Pontryagin, Rokhlin) and my pupils. Prof. M. Berry once formulated the following two principles:

The Arnold Principle. If a notion bears a personal name, then this name is not the name of the discoverer.

The Berry Principle. The Arnold Principle is applicable to itself.

Let's return, however, to teaching of mathematics in France.

When I was a first-year student at the Faculty of Mechanics and Mathematics of the Moscow State University, the lectures on calculus were read by the set-theoretic topologist L.A. Tumarkin, who conscientiously retold the old classical calculus course of French type in the Goursat version. He told us that integrals of rational functions along an algebraic curve can be taken if the corresponding Riemann surface is a sphere and, generally speaking, cannot be taken if its genus is higher, and that for the sphericity it is enough to have a sufficiently large number of double points on the curve of a given degree (which forces the curve to be unicursal: it is possible to draw its real points on the projective plane with one stroke of a pen).

These facts capture the imagination so much that (even given without any proofs) they give a better and more correct idea of modern mathematics than whole volumes of the Bourbaki treatise. Indeed, here we find out about the existence of a wonderful connection between things which seem to be completely different: on the one hand, the existence of an explicit expression for the integrals and the topology of the corresponding Riemann surface and, on the other hand, between the number of double points and genus of the corresponding Riemann surface, which also exhibits itself in the real domain as the unicursality.

Jacobi noted, as mathematics' most fascinating property, that in it one and the same function controls both the presentations of a whole number as a sum of four squares and the real movement of a pendulum.

These discoveries of connections between heterogeneous mathematical objects can be compared with the discovery of the connection between electricity and magnetism in physics or with the discovery of the similarity between the east coast of America and the west coast of Africa in geology.

The emotional significance of such discoveries for teaching is difficult to overestimate. It is they who teach us to search and find such wonderful phenomena of harmony of the Universe.

The de-geometrisation of mathematical education and the divorce from physics sever these ties. For example, not only students but also modern algebro-geometers on the whole do not know about the Jacobi fact mentioned here: an elliptic integral of first kind expresses the time of motion along an elliptic phase curve in the corresponding Hamiltonian system.

Rephrasing the famous words on the electron and atom, it can be said that a hypocycloid is as inexhaustible as an ideal in a polynomial ring. But teaching ideals to students who have never seen a hypocycloid is as ridiculous as teaching addition of fractions to children who have never cut (at least mentally) a cake or an apple into equal parts. No wonder that the children will prefer to add a numerator to a numerator and a denominator to a denominator.

From my French friends I heard that the tendency towards super-abstract generalizations is their traditional national trait. I do not entirely disagree that this might be a question of a hereditary disease, but I would like to underline the fact that I borrowed the cake-and-apple example from Poincaré.

The scheme of construction of a mathematical theory is exactly the same as that in any other natural science. First we consider some objects and make some observations in special cases. Then we try and find the limits of application of our observations, look for counter-examples which would prevent unjustified extension of our observations onto a too wide range of events (example: the number of partitions of consecutive odd numbers 1, 3, 5, 7, 9 into an odd number of natural summands gives the sequence 1, 2, 4, 8, 16, but then comes 29).

As a result we formulate the empirical discovery that we made (for example, the Fermat conjecture or Poincaré conjecture) as clearly as possible. After this there comes the difficult period of checking as to how reliable are the conclusions .

At this point a special technique has been developed in mathematics. This technique, when applied to the real world, is sometimes useful, but can sometimes also lead to self-deception. This technique is called modelling. When constructing a model, the following idealisation is made: certain facts which are only known with a certain degree of probability or with a certain degree of accuracy, are considered to be "absolutely" correct and are accepted as "axioms". The sense of this "absoluteness" lies precisely in the fact that we allow ourselves to use these "facts" according to the rules of formal logic, in the process declaring as "theorems" all that we can derive from them.

It is obvious that in any real-life activity it is impossible to wholly rely on such deductions. The reason is at least that the parameters of the studied phenomena are never known absolutely exactly and a small change in parameters (for example, the initial conditions of a process) can totally change the result. Say, for this reason a reliable long-term weather forecast is impossible and will remain impossible, no matter how much we develop computers and devices which record initial conditions.

In exactly the same way a small change in axioms (of which we cannot be completely sure) is capable, generally speaking, of leading to completely different conclusions than those that are obtained from theorems which have been deduced from the accepted axioms. The longer and fancier is the chain of deductions ("proofs"), the less reliable is the final result.

Complex models are rarely useful (unless for those writing their dissertations).

The mathematical technique of modelling consists of ignoring this trouble and speaking about your deductive model in such a way as if it coincided with reality. The fact that this path, which is obviously incorrect from the point of view of natural science, often leads to useful results in physics is called "the inconceivable effectiveness of mathematics in natural sciences" (or "the Wigner principle").

Here we can add a remark by I.M. Gel'fand: there exists yet another phenomenon which is comparable in its inconceivability with the inconceivable effectiveness of mathematics in physics noted by Wigner - this is the equally inconceivable ineffectiveness of mathematics in biology.

"The subtle poison of mathematical education" (in F. Klein's words) for a physicist consists precisely in that the absolutised model separates from the reality and is no longer compared with it. Here is a simple example: mathematics teaches us that the solution of the Malthus equation $dx/dt = x$ is uniquely defined by the initial conditions (that is that the corresponding integral curves in the (t, x)-plane do not intersect each other). This conclusion of the mathematical model bears little relevance to the reality. A computer experiment shows that all these integral curves have common points on the negative t-semi-axis. Indeed, say, curves with the initial conditions $x(0) = 0$ and $x(0) = 1$ practically intersect at $t = -10$ and at $t = -100$ you cannot fit in an atom between them. Properties of the space at such small distances are not described at all by Euclidean geometry. Application of the uniqueness theorem in this situation obviously exceeds the accuracy of the model. This has to be respected in practical application of the model, otherwise one might find oneself faced with serious troubles.

I would like to note, however, that the same uniqueness theorem explains why the closing stage of mooring of a ship to the quay is carried out manually: on steering, if the velocity of approach would have been defined as a smooth (linear) function of the distance, the process of mooring would have required an infinitely long period of time. An alternative is an impact with the quay (which is damped by suitable non-ideally elastic bodies). By the way, this problem had to be seriously confronted on landing the first descending apparata on the Moon and Mars and also on docking with space stations - here the uniqueness theorem is working against us.

Unfortunately, neither such examples, nor discussing the danger of fetishising theorems are to be met in modern mathematical textbooks, even in the better ones. I even got the impression that scholastic mathematicians (who have little knowledge of physics) believe in the principal difference of the axiomatic mathematics from modelling which is common in natural science and which always requires the subsequent control of deductions by an experiment.

Not even mentioning the relative character of initial axioms, one cannot forget about the inevitability of logical mistakes in long arguments (say, in the form of a computer breakdown caused by cosmic rays or quantum oscillations). Every working mathematician knows that if one does not control oneself (best of all by examples), then after some ten pages half of all the signs in formulae will be wrong and twos will find their way from denominators into numerators.

The technology of combatting such errors is the same external control by experiments or observations as in any experimental science and it should be taught from the very beginning to all juniors in schools.

Attempts to create "pure" deductive-axiomatic mathematics have led to the rejection of the scheme used in physics (observation - model - investigation of the model - conclusions

- testing by observations) and its substitution by the scheme: definition - theorem - proof. It is impossible to understand an unmotivated definition but this does not stop the criminal algebraists-axiomatisators. For example, they would readily define the product of natural numbers by means of the long multiplication rule. With this the commutativity of multiplication becomes difficult to prove but it is still possible to deduce it as a theorem from the axioms. It is then possible to force poor students to learn this theorem and its proof (with the aim of raising the standing of both the science and the persons teaching it). It is obvious that such definitions and such proofs can only harm the teaching and practical work.

It is only possible to understand the commutativity of multiplication by counting and re-counting soldiers by ranks and files or by calculating the area of a rectangle in the two ways. Any attempt to do without this interference by physics and reality into mathematics is sectarianism and isolationism which destroy the image of mathematics as a useful human activity in the eyes of all sensible people.

I shall open a few more such secrets (in the interest of poor students).

The determinant of a matrix is an (oriented) volume of the parallelepiped whose edges are its columns. If the students are told this secret (which is carefully hidden in the purified algebraic education), then the whole theory of determinants becomes a clear chapter of the theory of poly-linear forms. If determinants are defined otherwise, then any sensible person will forever hate all the determinants, Jacobians and the implicit function theorem.

What is a group? Algebraists teach that this is supposedly a set with two operations that satisfy a load of easily-forgettable axioms. This definition provokes a natural protest: why would any sensible person need such pairs of operations? "Oh, curse this maths" – concludes the student (who, possibly, becomes the Minister for Science in the future).

We get a totally different situation if we start off not with the group but with the concept of a transformation (a one-to-one mapping of a set onto itself) as it was historically. A collection of transformations of a set is called a group if along with any two transformations it contains the result of their consecutive application and an inverse transformation along with every transformation.

This is all the definition there is. The so-called "axioms" are in fact just (obvious) properties of groups of transformations. What axiomatisators call "abstract groups" are just groups of transformations of various sets considered up to isomorphisms (which are one-to-one mappings preserving the operations). As Cayley proved, there are no "more abstract" groups in the world. So why do the algebraists keep on tormenting students with the abstract definition?

By the way, in the 1960s I taught group theory to Moscow schoolchildren. Avoiding all the axiomatics and staying as close as possible to physics, in half a year I got to the Abel theorem on the unsolvability of a general equation of degree five in radicals (having on the way taught the pupils complex numbers, Riemann surfaces, fundamental groups and monodromy groups of algebraic functions). This course was later published by one of the audience, V. Alekseev, as the book The Abel theorem in problems.

What is a smooth manifold? In a recent American book I read that Poincaré was not acquainted with this (introduced by himself) notion and that the "modern" definition was only given by Veblen in the late 1920s: a manifold is a topological space which satisfies a long series of axioms.

For what sins must students try and find their way through all these twists and turns? Actually, in Poincaré's Analysis Situs there is an absolutely clear definition of a smooth manifold which is much more useful than the "abstract" one.

A smooth k-dimensional submanifold of the Euclidean space R^N is its subset which in a neighbourhood of its every point is a graph of a smooth mapping of R^k into $R^{(}N - k)$ (where R^k and $R^{(}N-k)$ are coordinate subspaces). This is a straightforward generalization of most common smooth curves on the plane (say, of the circle $x^2 + y^2 = 1$) or curves and surfaces in the three-dimensional space.

Between smooth manifolds smooth mappings are naturally defined. Diffeomorphisms are mappings which are smooth, together with their inverses.

An "abstract" smooth manifold is a smooth submanifold of a Euclidean space considered up to a diffeomorphism. There are no "more abstract" finite-dimensional smooth manifolds in the world (Whitney's theorem). Why do we keep on tormenting students with the abstract definition? Would it not be better to prove them the theorem about the explicit classification of closed two-dimensional manifolds (surfaces)?

It is this wonderful theorem (which states, for example, that any compact connected oriented surface is a sphere with a number of handles) that gives a correct impression of what modern mathematics is and not the super-abstract generalizations of naive submanifolds of a Euclidean space which in fact do not give anything new and are presented as achievements by the axiomatisators.

The theorem of classification of surfaces is a top-class mathematical achievement, comparable with the discovery of America or X-rays. This is a genuine discovery of mathematical natural science and it is even difficult to say whether the fact itself is more attributable to physics or to mathematics. In its significance for both the applications and the development of correct Weltanschauung it by far surpasses such "achievements" of mathematics as the proof of Fermat's last theorem or the proof of the fact that any sufficiently large whole number can be represented as a sum of three prime numbers.

For the sake of publicity modern mathematicians sometimes present such sporting achievements as the last word in their science. Understandably this not only does not contribute to the society's appreciation of mathematics but, on the contrary, causes a healthy distrust of the necessity of wasting energy on (rock-climbing-type) exercises with these exotic questions needed and wanted by no one.

The theorem of classification of surfaces should have been included in high school mathematics courses (probably, without the proof) but for some reason is not included even in university mathematics courses (from which in France, by the way, all the geometry has been banished over the last few decades).

The return of mathematical teaching at all levels from the scholastic chatter to presenting the important domain of natural science is an espessially hot problem for France. I was astonished that all the best and most important in methodical approach mathematical books are almost unknown to students here (and, seems to me, have not been translated into French). Among these are Numbers and figures by Rademacher and Töplitz, Geometry and the imagination by Hilbert and Cohn-Vossen, What is mathematics? by Courant and Robbins, How to solve it and Mathematics and plausible reasoning by Polya, Development of mathematics in the 19th century by F. Klein.

I remember well what a strong impression the calculus course by Hermite (which does exist in a Russian translation!) made on me in my school years.

Riemann surfaces appeared in it, I think, in one of the first lectures (all the analysis was, of course, complex, as it should be). Asymptotics of integrals were investigated by means of path deformations on Riemann surfaces under the motion of branching points (nowadays, we would have called this the Picard-Lefschetz theory; Picard, by the way, was Hermite's son-in-law - mathematical abilities are often transferred by sons-in-law: the dynasty Hadamard - P. Levy - L. Schwarz - U. Frisch is yet another famous example in the Paris Academy of Sciences).

The "obsolete" course by Hermite of one hundred years ago (probably, now thrown away from student libraries of French universities) was much more modern than those most boring calculus textbooks with which students are nowadays tormented.

If mathematicians do not come to their senses, then the consumers who preserved a need in a modern, in the best meaning of the word, mathematical theory as well as the immunity (characteristic of any sensible person) to the useless axiomatic chatter will in the end turn down the services of the undereducated scholastics in both the schools and the universities.

A teacher of mathematics, who has not got to grips with at least some of the volumes of the course by Landau and Lifshitz, will then become a relict like the one nowadays who does not know the difference between an open and a closed set.

V.I. Arnold

Literatur

[1] Ben Arans, Bogachev, Molchanov: (2004)

[2] Borgs, Kotecky

[3] Georgii,

[4] Resnick,