

修士論文 (1991年度)

連想記憶モデルにおける平衡状態・非平衡状態の性質^{*}

東工大・理 尾関 智子

(1992年3月31日受理)

神経生理学的知見に基づくホップフィールドモデルは、脳の自己連想記憶を説明するモデルである。このモデルは、主にスピングラスとの形式的な類似性のために、平衡統計力学の観点から活発に研究がなされてきた。それにより、神経回路網の平衡状態における性質が様々な観点から明らかにされてきた。例えば、記憶容量や準安定状態の数などが挙げられる。しかしながら、平衡状態の理論を用いたのでは、回復されるスピードや回路網の初期状態に依存した振る舞いなど、記憶回復過程における神経回路網の動的な性質を記述する事が出来ない。回路網の動的な振る舞いを直接扱うのは困難であり、どの問題にも適用できる一般的な方法と言うのは開発されていない。この論文の目的は、以下に述べるような二つの方法によって、非平衡状態における回路網の性質を調べることである。一つめの方法は、信号・ノイズ解析によって回路網のミクロな状態方程式から、マクロな状態変数の時間発展の方程式を導き出した、甘利・馬被の理論である。甘利・馬被の理論は、神経細胞への入力信号のうちのノイズの部分がガウス分布に従うという仮定のもとで、絶対零度における回路網の動的な振る舞いをうまく記述した。我々はその仮定の妥当性を検討するために、モンテカルロシミュレーションを行った。それにより記憶回復が成功する場合には、甘利・馬被の理論における仮定が正しいことがわかった。そこで甘利・馬被の理論を有限温度に拡張し、回路網の動的及び静的な性質を得た。そうして得られた平衡状態における相図は、平衡状態の平均場理論によって求められたものと非常によく似た振る舞いを示すことがわかった。二つめの方法は、時間依存したギンツブルグ・ランダウ方程式(TDGL)によるものである。TDGL方程式は、自由エネルギーの谷への最急降下を記述した方程式である。しかしながら今回用いたモデルにおいては、TDGL方程式はいくつかの問題点を抱えている。つまり平衡状態の平均場理論によって求められた回路網の平衡状態が、自由エネルギーの鞍点(不安定平衡点)になっているのである。そこでTDGL方程式を修正して回路網の初期状態によって決定される動的な振る舞いを記述した。これらの方法により導かれた時間発展の方程式により回復過程のダイナミクスを論じる。

^{*} 本稿は、英文で投稿されましたが、編集部より依頼して、日本語で再投稿していただきました。(編集部)

尾関 智子

1 はじめに

人間の脳の中にある神経回路網は、相互に結合したたくさんの神経細胞の集合体である。一個の神経細胞は、細胞体や樹状突起、軸索、シナプスなどからなる。個々の神経細胞は、他の神経細胞から入力信号を受け取り、その入力の加重和がそれぞれの神経細胞に固有な閾値を越えると自らも他の神経細胞へとパルス信号を出力する。この様に個々の神経細胞の働きは、単純であるが、それらが集まって神経回路網を形成すると思考や記憶とその回復、分類などのような複雑な働きをする。これは相互結合した単純な要素の集合からなる物理系でよく観測される協力現象である。

ホップフィールド [1] が神経回路網の連想記憶モデルとスピングラスの類似性を指摘して以来、数多くの研究が物理学者によってなされてきた。連想記憶とは、次のような脳の働きを指す。例えば不完全な文字が示されたとしても、人はそこから正しい文字を想像することができる。ホップフィールドモデルにおいては、記憶させるパターンは神経回路網の引き込み領域として埋め込まれる。記憶させたパターンの内の一つに近い状態から出発させると、神経回路網はその記憶させたパターンに近づいていく。(この時パターンは記憶回復されたという。) この論文の目的は、連想記憶のホップフィールドモデルにおける記憶回復過程の動的な振舞いを記述するところにある。

この章においては、まずホップフィールドモデルを紹介し、その後甘利と馬被 [2] によって提案された、回復過程の動的振る舞いを記述する理論を紹介する。最後にアミットら [3] による神経回路網の統計力学を示す。

1.1 ホップフィールドモデル

ホップフィールドによって提案されたモデルは、次のとおりである。神経回路網は N 個の相互結合した神経細胞からなり、個々の神経細胞は二つの状態をとる。(マッカラック・ピッツモデル [8])

$$S_i^t = \pm 1 \quad (1)$$

ここで S_i^t は、時刻 t における i 番目の神経細胞の状態である。 i 番目の神経細胞から j 番目の神経細胞への入力の重みを J_{ij} で表すことにすると、時刻 t における i 番目の神経細胞への入力の加重和は次のようになる。

$$h_i^t = \sum_{j \neq i} J_{ij} S_j^t \quad (2)$$

それぞれの神経細胞がランダムかつ非同期的に状態更新をすると仮定する。つまり δt の時間間隔に神経細胞が一つだけ次の規則に従ってその状態を変えるものとする。

$$S_i^{t+\delta t} = \text{sgn}(h_i^t) \quad (3)$$

以上のような神経回路網に P 個のランダムパターン $\{\xi^\mu\}$ を記憶させる場合を考える。ここで ξ_i^μ は、等しい確率で 1 または -1 をとる。するとパターンどうしは N が無限大の極限で直交条件を満たす。

$$\frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu = \delta_{\mu\nu} \quad (4)$$

このようなパターンを埋め込むには、ヘブの学習則に従って J_{ij} を次のように決定すればよい。

$$J_{ij} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu \quad (5)$$

結合のマトリックス J_{ij} をこの様に決定すると、 $\{\xi^\mu\}$ を安定な平衡状態として ($\{\xi_i^\mu\}$ はダイナミクス (3) に対して安定である) 神経回路網に埋め込むことができる。

ホップフィールドは神経回路網のエネルギーを次のように定義し、

$$E = -\frac{1}{2} \sum_{j \neq i} J_{ij} S_i S_j \quad (6)$$

神経回路網の協同現象（記憶とその回復）を説明した。

1. 記憶させたパターンは神経回路網の自由エネルギー (6) の極値になっている。
2. 初期状態から出発して神経回路網の状態は最初エネルギーを減少させるように変化し、エネルギーの極値に到達すると状態変化は起こらない。つまりパターンが回復される。

1.1.1 絶対記憶容量

パターン比 ($\alpha = P/N$) の絶対容量の評価、つまり厳密な意味でどのくらいのパターンを回路網の平衡状態として埋め込むことができるかは、興味深い問題である。ここで絶対記憶容量を計算してみる [2]。

神経回路網の状態が ν 番目の記憶させたパターンと一致しているとき、 i 番目の神経細胞への入力加重和は、次のようになる。

$$\sum_{j \neq i} J_{ij} \xi_j^\nu = \frac{1}{N} \sum_{j \neq i} \sum_\mu \xi_i^\mu \xi_j^\mu \xi_j^\nu = \xi_i^\nu + \frac{1}{N} \sum_{j \neq i} \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu \quad (7)$$

第二項はノイズであり、中心極限定理により、平均 0、分散 α のガウス分布に従うランダム変数になっていることが分かる。 i 番目の神経細胞が $\xi_i^\nu = 1$ という状態にあるとき、その神経細胞

尾関 智子

が次のステップで間違ふ（次のステップでの出力が-1になる）確率 p は、ノイズの項が-1より小さくなる確率に等しい。

$$p = \operatorname{erf}(1/\sqrt{\alpha}) \quad (8)$$

$$\operatorname{erf}(u) = \int_{-\infty}^{-u} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt \quad (9)$$

したがって神経回路網全体が間違わない確率は、

$$p_{\text{no error}} = (1 - p)^N \quad (10)$$

となる。 $p_{\text{no error}} = 1$ が成り立つような α の範囲を求める。 $N \rightarrow \infty$ の極限で (8) 式より

$$p \simeq \sqrt{\alpha} \exp(-\frac{1}{2\alpha}) \quad (11)$$

となる。さらに

$$\begin{aligned} \log[p_{\text{no error}}] &= N \log[1 - p] \simeq -Np \simeq -N\sqrt{\alpha} \exp(-\frac{1}{2\alpha}) \\ &= -\exp[\log N + \frac{1}{2} \log(\alpha) - \frac{1}{2\alpha}] \end{aligned} \quad (12)$$

となる。 $p_{\text{no error}}$ が 1 に等しくなるためには、次の式が成り立たなければならない。

$$\log N + \frac{1}{2} \log(\alpha) - \frac{1}{2\alpha} < 0. \quad (13)$$

ここでまず上式の不等式を等式と考えて、 $1/\alpha$ の近似的な臨界値を求める。

$$\frac{1}{\alpha} \simeq 2 \log N - \log \log N \quad (14)$$

従って $p_{\text{no error}} = 1$ が成立するためには、

$$\alpha < \alpha_0 = \frac{1}{2 \log N - \log \log N} \quad (15)$$

とならなければならない。これより、絶対記憶容量は $N \rightarrow \infty$ の極限で、ゼロとなってしまうことが分かる。しかしながら、ホップフィールドはシミュレーションに基づいて、 $\alpha < \alpha_c \sim 0.15$ であるとき、回路網は少ないエラーで連想記憶の役割を果たすと結論している。これはスピンガラスとの類似性から次のように説明することができる。記憶させたパターンは、厳密な意味では神経回路網のエネルギーの極値になっていない。しかし記憶させたパターンのまわりには、高いエネルギーバリアによって囲まれた多くの準安定状態が存在するため記憶させたパターンに近い回復が可能になる。ところが記憶させたパターンとは無関係な意味の無い状態もたくさんあり、その数は α が増加するにつれて無限に増加する。 α が α_c を超えると、エネルギーバリアが無くなって記憶回復が不可能になる。

1.2 甘利・馬被の理論

動的な回復過程を詳細に記述するためには、非平衡状態を直接動的に扱わなければならない。系の動的振る舞いを記述するのに微視的な状態変化の式 (3) を用いたのでは N 個の方程式が必要となり不便である。そこで巨視的な状態変数、方向余弦（またはオーバーラップ）を定義する。

$$m_t^\mu = \frac{1}{N} \sum_i \xi_i^\mu S_i^t \quad (16)$$

m_t^μ は -1 から 1 までの値を取り得る。神経回路網が μ 番目のパターン $\{\xi^\mu\}$ と一致する状態にあるとき、 m_t^μ の絶対値は 1 となり、神経回路網が μ 番目のパターンと無関係な状態にあるとき、 m_t^μ は 0 となる。方向余弦の時間依存した振る舞いを記述する方程式は、神経回路網がどのように記憶させたパターンに近付くかを説明する。この節では、そのような方程式を導く甘利・馬被の理論を紹介する。

1.2.1 甘利・馬被の理論 I

甘利と馬被 [2] は、神経細胞の状態更新が同期的である場合のダイナミクスを論じた。以下では全ての神経細胞が離散的時刻 ($t = 1, 2, \dots$) にいっせいにその状態を変える場合について考える。ここで簡単のため、次のような形をした一番目の記憶されたパターン ξ^1 が回復される過程を考える。

$$\xi^1 = (1, 1, \dots, 1) \quad (17)$$

このようににおいても一般性は失われない。すると神経回路網の状態と記憶させたパターン ξ^1 との方向余弦 (16) は、

$$m_t^1 = m_t = \frac{1}{N} \sum_i S_i^t \quad (18)$$

となる。

初期状態がランダムに選ばれる時、微視的な状態変換の方程式 (3) は次のようになる。

$$\begin{aligned} S_i^1 &= \text{sgn}(h_i^0) \\ h_i^0 &= \sum_j J_{ij} S_j^0 = \frac{1}{N} \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu S_j^0 = m_0 + N_i^0 \end{aligned} \quad (19)$$

ノイズの項は、

$$N_i^0 = \frac{1}{N} \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu S_j^0 \quad (20)$$

尾関 智子

それぞれ独立かつ同じ分布に従う多数のランダムな変数 $\xi_i^\mu \xi_j^\mu S_j^0$ の和であり、漸近的に平均が0分散が r のガウス分布に従う。

$$r = \frac{1}{N^2}(P-1)(N-1) \simeq \frac{P}{N} = \alpha \quad (21)$$

したがって次式が導かれる。

$$S_i^1 = \text{sgn}(m_0 + \sqrt{\alpha}\varepsilon_i) \quad (22)$$

ここで ε_i は、平均が0で分散が1のガウス分布に従うランダム変数である。

$$m_1 = \frac{1}{N} \sum_i S_i^1 = \frac{1}{N} \sum_i \text{sgn}(m_0 + \sqrt{\alpha}\varepsilon_i) \quad (23)$$

$N \rightarrow \infty$ の極限で、さらに右辺はアンサンブル平均に収束する。

$$\begin{aligned} m_1 &= E[\text{sgn}(m_0 + \sqrt{\alpha}\varepsilon_i)] \\ &= 1 \times \text{Prob}[m_0 + \sqrt{\alpha}\varepsilon_i > 0] - 1 \times \{1 - \text{Prob}[m_0 + \sqrt{\alpha}\varepsilon_i > 0]\} \\ &= 2\text{Prob}[m_0 + \sqrt{\alpha}\varepsilon_i > 0] - 1 \end{aligned} \quad (24)$$

よって $t=0$ から $t=1$ へ方向余弦の変換式は

$$m_1 = F(m_0/\sqrt{\alpha}) \quad (25)$$

$$F(s) = \int_{-s}^s \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) du. \quad (26)$$

となる。次式も成り立つように見えるが

$$m_{t+1} = F(m_t/\sqrt{\alpha}) \quad (27)$$

これは $t=0$ の時にしか成立しない。 $t \neq 0$ のときには、 S_j^t は ξ^μ に依存していて、 $\xi_i^\mu \xi_j^\mu S_j^t$ はもはや独立なランダム変数ではないからである。従って N_i^t がガウス分布に従うと考えることはできない。

ガウス分布に従うという仮定(20)は成り立たないが、甘利と馬被は N_i^t が平均が0で分散 σ_t^2 のガウス分布に従うランダム変数であると仮定した。

$$N_i^t = \frac{1}{N} \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu S_j^t \quad (28)$$

そして一ステップ前の神経細胞の状態 S^{t-1} の ξ 依存性を考慮し、時間依存する変数 m_t と σ_t に対する巨視的状态変換の方程式を導いた。

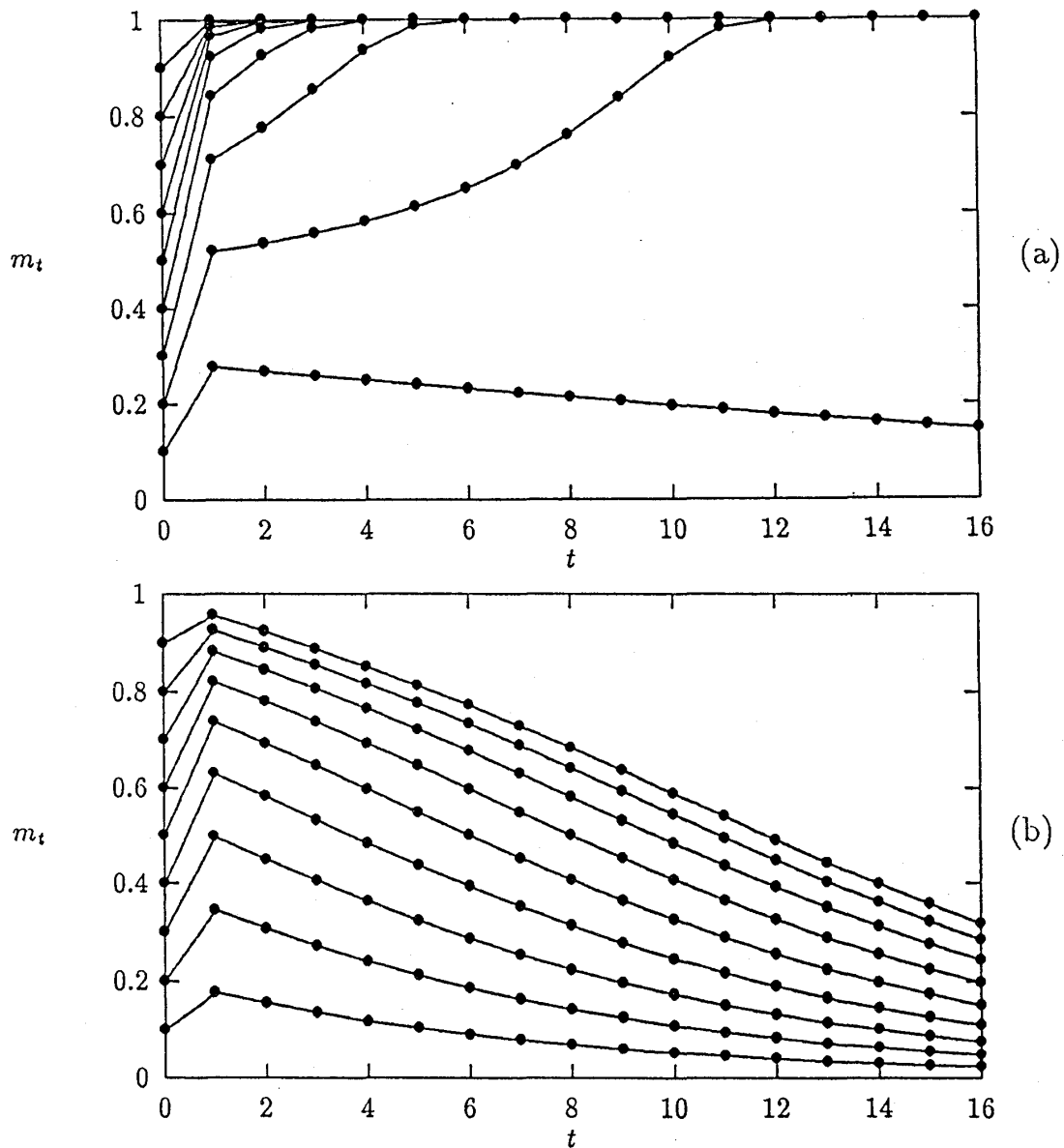


図 1: 甘利と馬被の理論によって求められた方向余弦の振舞い; (a) $\alpha = 0.08$ 、方向余弦の初期値の閾値現象が見られる。(b) $\alpha = 0.2$ 、どんなに記憶させたパターンに近い初期状態から出発しても記憶回復できない。

$$m_{t+1} = F(m_t/\sigma_t) \quad (29)$$

$$\sigma_{t+1}^2 = \alpha + 4[p(\bar{m}_t)]^2 + 4\alpha\bar{m}_t p(\bar{m}_t)m_{t+1} \quad (30)$$

$$\bar{m}_t = m_t/\sigma_t \quad (31)$$

$$p(u) = (1/\sqrt{2\pi}) \exp(-u^2/2) \quad (32)$$

巨視的状态変数に関する以上の方程式は付録と同様の方法によって求められる。

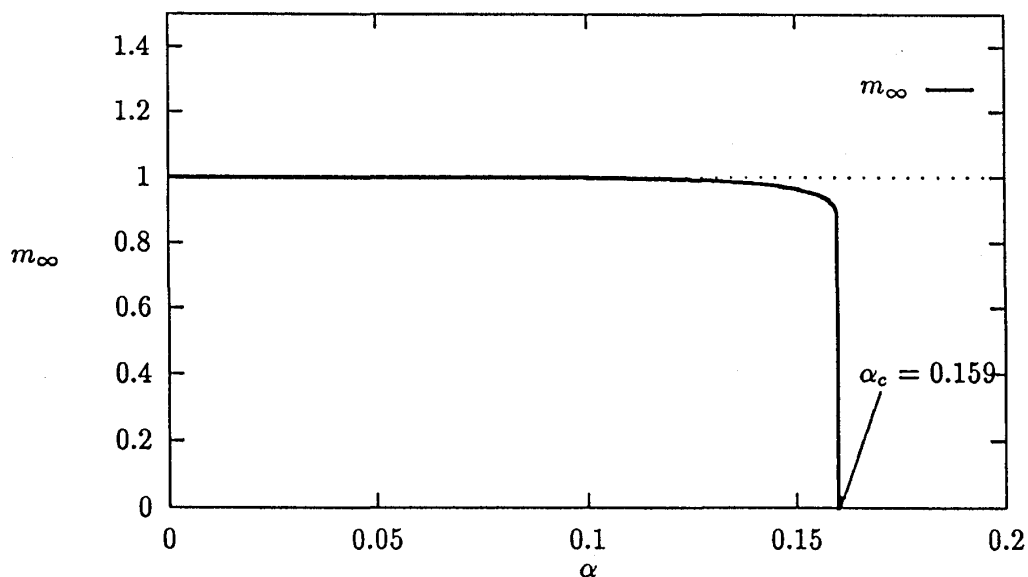


図 2: 甘利と馬被によって求められたパターン比に対する平衡状態の方向余弦; $\alpha < \alpha_c$ の時、神経回路網は連想記憶の働きをするが、 α がいったん α_c を超えてしまうと途端に連想記憶モデルとして働かなくなる。

以上のような式から求められた典型的な例が図 1 に示されている。 α が 0.08 と 0.2 の場合について、方向余弦の動的な振舞いを式 (29) と (30) から導いた。図から次のような現象を見ることができる。

1. $\alpha = 0.08$ のとき $t \rightarrow \infty$ の極限で、回復が成功しているように見えても、 m_t は厳密には 1 にならない ($m = 0.9996 \dots$)。これより記憶させたパターンそのものは神経回路網の平衡状態ではなく、厳密な意味での回復は得られないことが分かる。
2. $\alpha = 0.08$ のとき方向余弦の初期値に依存する閾値現象がみられる。臨界値 m_c が存在して、初期状態が $m_0 > m_c$ を満たすとき m_t は時間と共に増加してついには 1 に収束する (パターンが回復される)。 $m_0 < m_c$ である時は、はじめは記憶させたパターンに近づくがその後 m_t は次第に減少しついには 0 となる。
3. $\alpha = 0.2$ のとき、どんなに記憶させたパターンに近い状態から出発してもパターンは回復されない。この条件のもとでも、はじめはパターンに近づくがその後時間の経過と共にパターンから離れていってしまう。

α の臨界値が存在して α の値によって上記の項目 2 と 3 のような異なった現象が見られる。図 2 に見られるように、この違いは α_c のところで突然起こる。さらにこの図からパターン比の臨界

値 α_c はおよそ0.1597であるといえる。この値はホップフィールドフィールドがシミュレーションによって求めた値と非常に近い。

注目すべき事は、方向余弦の動的な振る舞いは強い履歴依存性を示すという事である。違う過去を持った二つの系がある時点で同じ方向余弦を持ったとしても、次の瞬間においては全く違った振る舞いを示す。それゆえ神経回路網の状態は、方向余弦の値のみによって記述することはできないことが分かる。

2章のシミュレーションの結果と比較すると、甘利と馬被の理論は方向余弦の動的振る舞いをうまく記述していることが分かる。神経回路網の動的過程は履歴依存性を示すが、甘利と馬被は神経細胞の一ステップ前の状態 S^{t-1} の ξ 依存性を考慮しただけで、系の振る舞いを記述することに成功した。2章では、この方法がうまく働いているのかを詳細に吟味することにする。

1.2.2 甘利・馬被の理論 II

甘利と馬被は、さらにノイズ N_i^t の平均値をも考慮に入れることによって、より正確に方向余弦の振る舞いを記述しようと試みた。

ここでは、ノイズが平均 $S_i^{t-1}b_t$ 、分散 σ_t^2 のガウス分布に従うものと仮定する。すると N_i^t は平均0、分散1のガウス分布に従うランダム変数を用いて次のように表すことができる。

$$N_i^t = S_i^{t-1}b_t + \sigma_t u$$

ゆえに方向余弦は

$$\begin{aligned} m_{t+1} &= E[\text{sgn}(m_t + S_i^{t-1}b_t + \sigma_t u)] \\ &= 2\text{Prob}[m_t + S_i^{t-1}b_t + \sigma_t u > 0] - 1 \\ &= F\left(\frac{m_t + S_i^{t-1}b_t}{\sigma_t}\right) \end{aligned} \quad (33)$$

となる。ここで $F(s)$ は、式(26)で表される。さらに S_i^{t-1} が ± 1 になる確率は $(1 \pm m_{t-1})/2$ となるので、

$$m_{t+1} = \langle F(m_t/\sigma_t) \rangle \quad (34)$$

$$\langle f(m_t/\sigma_t) \rangle = [(1 + m_{t-1})f\left(\frac{m_t + b_t}{\sigma_t}\right) + (1 - m_{t-1})f\left(\frac{m_t - b_t}{\sigma_t}\right)]/2 \quad (35)$$

である。つぎに b_t についての式を導く。

$$S_i^t b_{t+1} = E[N_i^{t+1}] = \frac{1}{N} \sum_{\mu \neq i} \sum_{j \neq i} E[\xi_i^\mu \xi_j^\mu S_j^{t+1}] \quad (36)$$

S_j^{t+1} の ξ^μ 依存性は次式で表される。

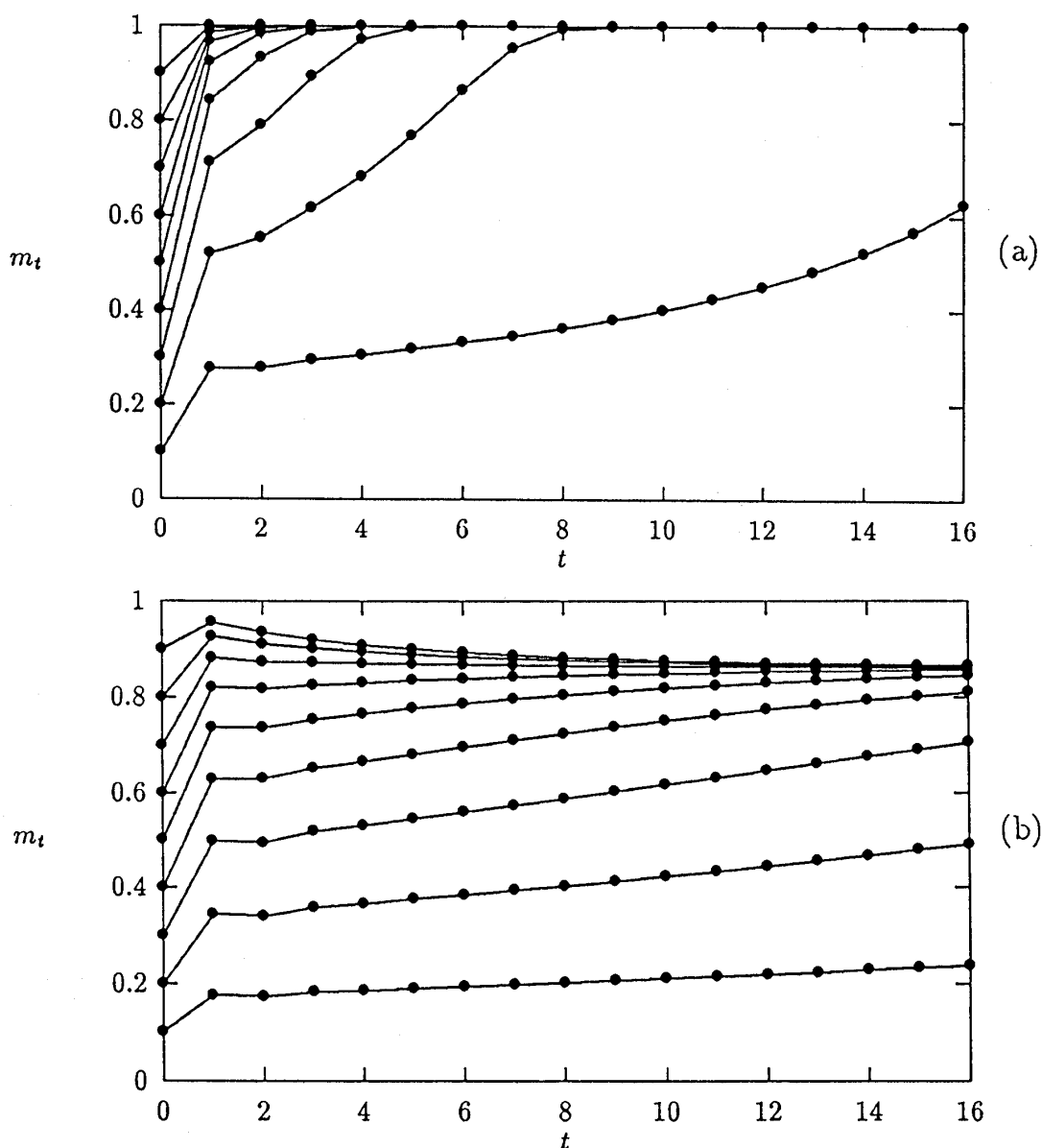


図 3: ノイズの平均 $E[N_i^t]$ を考慮した場合の甘利・馬被理論による方向余弦の振舞い; (a) $\alpha = 0.08$ (b) $\alpha = 0.2$ α が小さい場合でも方向余弦の初期値に依存した閾値現象は見られない。

$$\xi_i^\mu \xi_j^\mu S_j^{t+1} = \xi_i^\mu \xi_j^\mu \text{sgn}(m_t + N^{-1} \xi_i^\mu \xi_j^\mu S_i^t + Q_j^t) \quad (37)$$

Q_j^t もガウス分布に従うランダム変数である。ここで $\xi_i^\mu \xi_j^\mu$ についての条件付き期待値をとると、

$$E[\xi_i^\mu \xi_j^\mu S_j^{t+1}] = \frac{2S_i^t}{N\sigma_t} p\left(\frac{m_t + S_j^{t-1}b_t}{\sigma_t}\right) \quad (38)$$

となる。さらに S_j^{t-1} について平均をとると次式が得られる。

$$b_{t+1} = \frac{2r}{\sigma_t} \langle p(m_t/\sigma_t) \rangle \quad (39)$$

σ_i^2 の方程式については付録と同様の方法で得られる。

$$\sigma_{i+1}^2 = \alpha + 4 \langle p(m_t/\sigma_t) \rangle^2 + 4\alpha m_t/\sigma_t \langle p(m_t/\sigma_t) \rangle m_{t+1} \quad (40)$$

式(34)と(39)、(40)から得た方向余弦の振舞いは、図3に示されている。ノイズをゼロとして議論した節1.2.1では、 α が小さい場合に起こる方向余弦の初期値に依存した閾値現象が見られたが、ここでは見られない。ノイズの平均がゼロでないとした議論は、節1.2.1の改良となっていることが期待されるが、予想に反して、2章を見れば分かるようにシミュレーションとは異なった結果が得られてしまう。従って甘利・馬被の理論には何か重要な限界が含まれていると結論せざるを得ない。まず第一に考えられるのは、ノイズがガウシアンであるという仮定のなかに何か本質的な問題が含まれているということである。従ってノイズの項がガウス分布であるという仮定が実際に満たされているかどうかを確かめることが重要である。

1.3 ニューラルネットワークの平均場理論

アミットら[3]は、ホップフィールドモデルに対する平均場理論を詳細に論じた。次のようなハミルトニアンから出発して、

$$H = -\frac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j \quad (41)$$

確率的なノイズ（温度）が存在するようなネットワークの相図を描いた。 s 個の記憶させたパターン（ $\mu = 1, \dots, s \leq P$ ）を安定にするためにハミルトニアンに次のような項を加え、

$$H_h = -\sum_{\nu=1}^s h^\nu \sum_i \xi_i^\nu S_i \quad (42)$$

レプリカ法[5]を用いて自由エネルギーを計算した。

$$\ln Z = \lim_{n \rightarrow 0} (Z^n - 1)/n \quad (43)$$

一スピンあたりの自由エネルギーは

$$f = \lim_{n \rightarrow 0} \lim_{N \rightarrow \infty} \frac{-1}{\beta n N} (\langle \langle Z^n \rangle \rangle - 1) \quad (44)$$

で与えられる。ここで $\langle \langle \dots \rangle \rangle$ は ξ についての平均をとることを意味する。神経回路網は無限レンジの相互作用をしているので、 f は熱力学的極限で平均場理論によって厳密に解ける。

$$\begin{aligned} f = & \frac{1}{2} \alpha + (\alpha/2\beta n) \text{Tr} \ln[(1-\beta)\mathbf{I} - \beta \mathbf{q}] \\ & + (1/2n) \sum_{\nu\rho} (m_\rho^\nu)^2 + (\alpha\beta/2n) \sum_{\rho \neq \alpha} r_{\rho\sigma} q_{\rho\sigma} \\ & - \frac{1}{n\beta} \langle \langle \ln \text{Tr}_S \exp[\beta\{(\alpha\beta/2) \sum_{\rho \neq \sigma} r_{\rho\sigma} S^\rho S^\sigma + \sum_{\nu\rho} (m_\rho^\nu + h^\nu) \xi^\nu S^\rho\}] \rangle \rangle_\xi \end{aligned} \quad (45)$$

尾関 智子

ここで ρ と σ はレプリカ番号である。

m_ρ^μ の物理的な意味は、記憶させたパターンのうちのひとつと神経回路網との方向余弦である。

$$m_\rho^\mu = \frac{1}{N} \left\langle \left\langle \sum_i \xi_i^\mu < S_i^\rho > \right\rangle \right\rangle \quad (46)$$

もう一つの巨視的な状態変数はエドワーズ・アンダーソンオーダーパラメーターである。

$$q_{\rho\sigma} = \left\langle \left\langle \frac{1}{N} \sum_i < S_i^\rho > < S_i^\sigma > \right\rangle \right\rangle \quad (47)$$

残りの一つは $p-s$ 個のパターンと神経回路網の方向余弦の二乗平均である。

$$r_{\rho\sigma} = \alpha^{-1} \sum_{\mu=s+1}^{\alpha N} < m_\rho^\mu m_\sigma^\mu > \quad (48)$$

レプリカ対称性の破れは、結果には大きな影響を与えないことが分かっている。そこでレプリカ対称性を仮定してさらに議論を進める。

$$m_\rho^\mu = m^\mu \quad (49)$$

$$q_{\rho\sigma} = q, \rho \neq \sigma \quad (50)$$

$$r_{\rho\sigma} = r, \rho \neq \sigma \quad (51)$$

すると自由エネルギーは次のようになる。

$$\begin{aligned} f = & \frac{1}{2}\alpha + \frac{1}{2} \sum_\mu (m^\mu)^2 + \frac{\alpha}{2\beta} [\ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q}] \\ & + (\alpha\beta r/2)(1 - q) - \beta^{-1} \int \frac{dz}{\sqrt{2\pi}} \exp(-z^2/2) \\ & \times \left\langle \left\langle \ln 2 \cosh \beta [\sqrt{\alpha r} z + \sum_\nu (m^\nu + h^\nu \xi^\nu)] \right\rangle \right\rangle_\xi \end{aligned} \quad (52)$$

自由エネルギー f 、式(52)を m^ν と q 、 r について変分をとると平衡状態の状態方程式が得られる。

$$m^\nu = \left\langle \left\langle \xi^\nu \tanh \beta [\sqrt{\alpha r} z + \sum_\nu (m^\nu + h^\nu \xi^\nu)] \right\rangle \right\rangle \quad (53)$$

$$q = \left\langle \left\langle \tanh^2 \beta [\sqrt{\alpha r} z + \sum_\nu (m^\nu + h^\nu \xi^\nu)] \right\rangle \right\rangle \quad (54)$$

$$r = q/(1 - \beta + \beta q)^2 \quad (55)$$

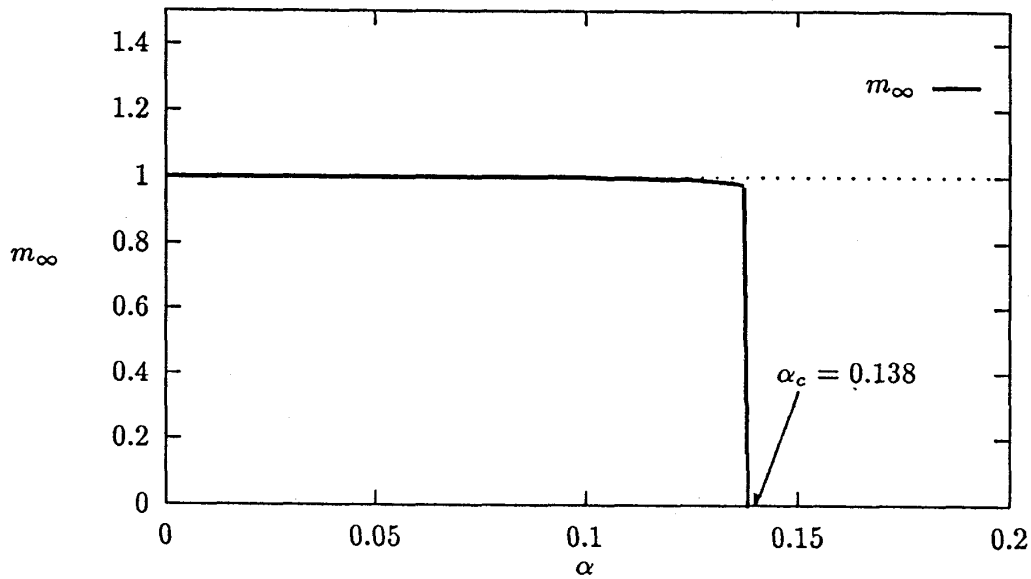


図 4: 平均場理論によって求められたパターン比と平衡状態の方向余弦との関係; スピングラス相から回復相への転移は一次相転移であることが分かる。 $T = 0$ におけるパターン比の臨界値 α_c は余利と馬被によって求められた値 $\alpha_c = 0.159$ と近い。

1.3.1 絶対零度における平衡状態の性質

μ 番目のパターンが記憶回復されている状態は、 $m^\nu = m\delta_{\mu\nu}$ で表される。 $T \rightarrow 0$ の極限で式 (53)-(55) は以下になる。

$$m = \text{erf}(m/\sqrt{2\alpha r}) \quad (56)$$

$$C \equiv \beta(1 - q) \rightarrow \sqrt{2/\pi\alpha r} \exp(-m^2/2\alpha r) \quad (57)$$

$$r = (1 - C)^{-2} \quad (58)$$

式 (56) は常に $m = 0$ の解を持つ。これがスピングラス解と呼ばれるもので、どんなパターンとも巨視的な方向余弦を持たない状態を示す。 α が $\alpha_c = 0.138$ より大きいとき、式 (56) は $m = 0$ の解のみを持つ。 α が α_c より小さくなると、突然 $m \neq 0$ の強磁性解と呼ばれる解が現れる。図 4 に示されるように α_c を越えると方向余弦は急激に 1 に近づく。方向余弦の平衡状態における振る舞いは甘利と馬被によるもの (図 2) と定性的に一致している。さらにパターン比の臨界値 $\alpha_c = 0.138$ も甘利と馬被の $\alpha_c = 0.159$ という値に近い。全く異なったアプローチから似た結果が得られるのは注目すべきことである。

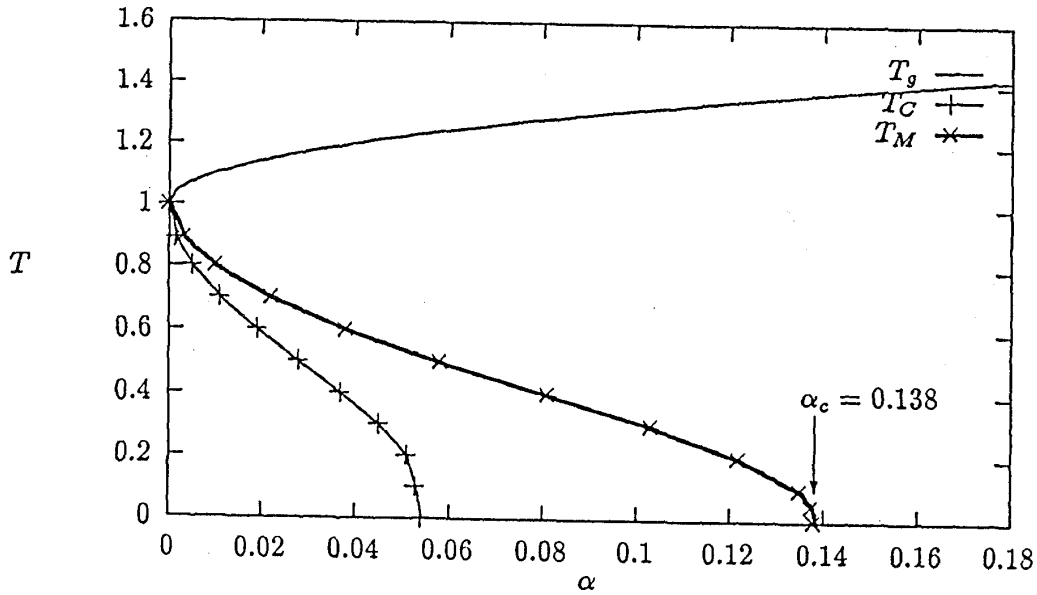


図 5: 平均場理論による相図; T_g は常磁性相から S G 相への転移温度である。 T_M は S G 相から回復相への一次転移の温度である。 T_C より温度が低いときは、回復相が神経回路網の自由エネルギーの最小値になっている。

1.3.2 有限温度における平衡状態

スピングラス相 無秩序な常磁性相から S G (スピングラス) 相への相転移は 2 次である。式 (54) と (55) を q と r のべきで展開すると、転移温度 T_g を以下のような手順で求めることができる。

$$q \simeq \beta^2 \alpha \simeq \frac{\beta^2 \alpha q}{(1 - \beta)^2} + O(q^2) \quad (59)$$

これより

$$T_g = 1 + \sqrt{\alpha} \quad (60)$$

となる。

回復状態 回復状態とは本質的には記憶させたパターンと巨視的な方向余弦を持ち、連想記憶モデルとして働くような状態である。回復状態は、次式で表される。

$$m^\mu = m \delta^{\mu\nu} \quad (61)$$

すると式 (53) と (54) は、次のようになる。

$$m = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(m + \sqrt{\alpha} r z) \quad (62)$$

$$q = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(m + \sqrt{\alpha r} z) \quad (63)$$

$$r = q/(1 - \beta + \beta q)^2 \quad (64)$$

相図を求めるために式 (62) - (64) を数値的に解くと図 5 が得られる。 T_M は回復相とスピングラス相の間の転移温度である。 T_C より温度が低いときは、回復相が神経回路網の自由エネルギーの最小値になっている。

2 ノイズの分布

1 章で見てきたように甘利と馬被 [2] によって提案された理論は、神経回路網の状態の ξ 依存性を考慮することによって、方向余弦の動的な振る舞いをうまく記述することに成功した（ノイズの平均 N_i^2 をゼロとした場合）。仮定の基本は、一つの神経細胞への入力の中のノイズの項がガウス分布に従うということである。この章では、シミュレーションを行い、ガウス分布に従うという仮定が成り立っているかを吟味する。

2.1 シミュレーション

2.1.1 キュミュラント展開

x を連続的な値をとり得る確率変数とし、 $f(x)$ を確率密度とする。 x の n 次のモーメントは次の様に定義される。

$$\langle x^n \rangle = \int dx x^n f(x) \quad (65)$$

ここで $\langle x \rangle$ は x の平均値であり、分散は $\langle x^2 \rangle - \langle x \rangle^2$ で定義される。特性関数 $\phi_x(k)$ を導入する。

$$\phi_x(k) \equiv \langle e^{ikx} \rangle = \int dx e^{ikx} f(x) = \sum_{n=0}^{\infty} \frac{(ik)^n \langle x^n \rangle}{n!} \quad (66)$$

確率密度 $f(x)$ は特性関数のフーリエ変換で与えられる。従って全てのモーメントが分かると確率密度を知ることができる。

特性関数をキュミュラント展開で表した方が便利である。

$$\phi_x(k) = \exp\left[\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n(x)\right] \quad (67)$$

$C_n(x)$ は n 番目のキュミュラントである。キュミュラントとモーメントの関係は次のように表される。

$$C_1(x) = \langle x \rangle$$

尾関 智子

$$\begin{aligned}
C_2(x) &= \langle x^2 \rangle - \langle x \rangle^2 \\
C_3(x) &= \langle x^3 \rangle - 3\langle x \rangle \langle x^2 \rangle + 2\langle x \rangle^3 \\
C_4(x) &= \langle x^4 \rangle - 3\langle x^2 \rangle^2 - 4\langle x \rangle \langle x^3 \rangle \\
&\quad + 12\langle x \rangle^2 \langle x^2 \rangle - 6\langle x \rangle^4
\end{aligned} \tag{68}$$

ここでちょうど $C_1(x)$ は平均、 $C_2(x)$ は分散である。

確率変数 x がガウス分布に従うならば、3 次以上の項はゼロになる。従って式 (28) のノイズの項 N_i^t がガウシアンであるかどうかを調べるには、キュミュラントを計算すればよい。

2.1.2 シミュレーションのアルゴリズム

神経回路網のサイズ N が 5000、7000、9000 に対して、 α がそれぞれ 0.08 と 0.20 の場合についてシミュレーションを行った。まず αN 個のランダムパターンを作り、それらのパターンをヘブルール (5) によって神経回路網に埋め込む。一番目のパターンと有限の方向余弦を持った初期状態から出発させて、全ての神経細胞を同期的に状態更新させる。そして状態更新のたびに N_i^t のキュミュラントを計算する。

2.2 結果と議論

図 6 に $N = 9000$, $\alpha = 0.08$ と 0.2 の場合の方向余弦の振る舞いが示されている。甘利と馬被の理論は方向余弦の振る舞いを定性的によく記述していることが分かる。

ノイズの項のキュミュラントは図 7 に示されている。この図から興味深い現象が見られる。(a) は $\alpha = 0.08$ で $m_0 = 0.5$ の場合である。ダイヤモンドの印の付いた実線は方向余弦の振る舞いを示す。この場合記憶させたパターンが回復されていて、さらに 3 次と 4 次のキュミュラントがほとんどゼロになっている。即ちノイズの項がガウス分布に従っているといえる。しかしながら (b) や (c) のようにパターンが回復されない場合には、3 次と 4 次のキュミュラントがゼロになっていない。従ってノイズの項はガウシアンではない。

さらに図 8-10 には m と σ^2 の動的な振る舞いを描いてある。ここではダイヤモンドの印の付いた実線は甘利と馬被によってノイズの平均がゼロであると仮定して求められた巨視的変数の振る舞いを表わしている。その他の印はシミュレーションの結果を表わしている。図 7 にあるように、パターンが回復されるときは甘利と馬被の理論は動的な振る舞いをうまく記述しているといえる。しかしパターンが回復されない場合には回復過程を甘利と馬被の理論で記述することはできない。

以上の結果からパターンの回復が成功する過程においてのみ、ノイズの分布がガウス分布になっていて甘利と馬被の理論が適用できることが分かる。

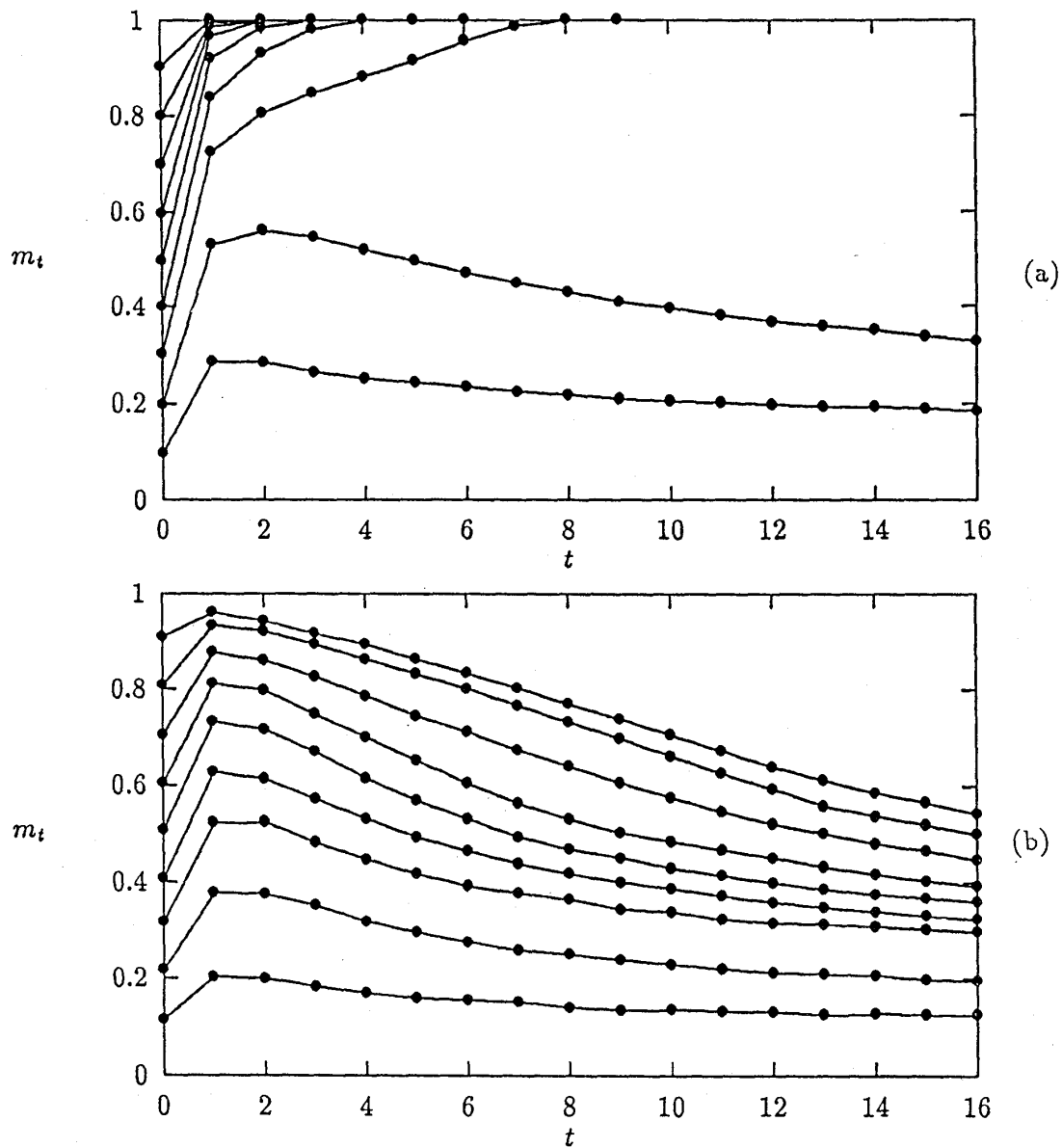


図 6: シミュレーションによる方向余弦の振る舞い; (a) $N = 9000, \alpha = 0.08$ 記憶させたパターンと近い初期状態から出発させるとパターンを回復することができる。しかしながら方向余弦の初期値が小さいとパターンは回復されない。(b) $N = 9000, \alpha = 0.2$ どんな初期状態から出発してもパターンは回復されない。甘利と馬被の理論は回復過程をうまく記述していることが分かる。

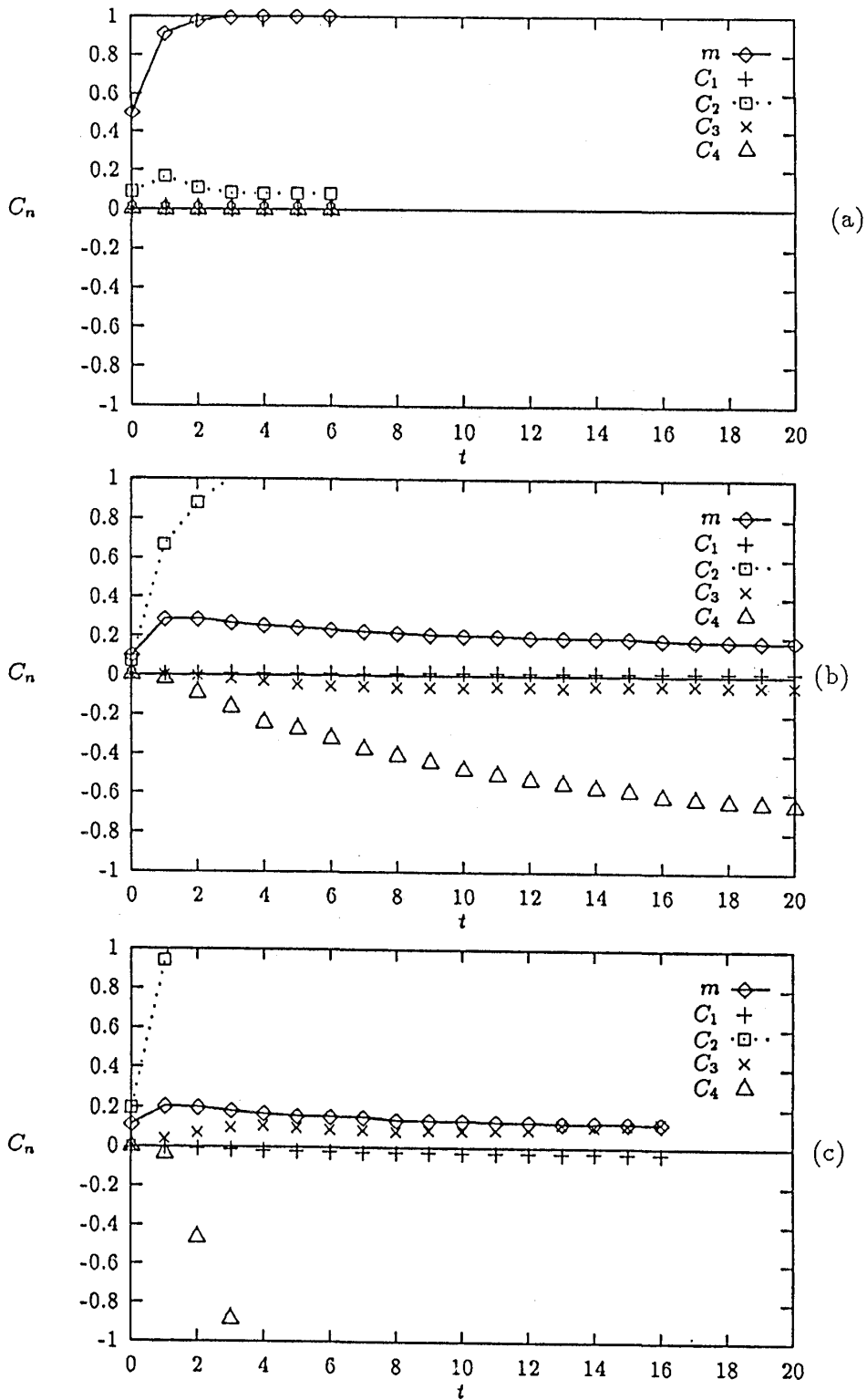


図 7: シミュレーションによるノイズのキュミュラント; $N = 9000$ (a) $\alpha = 0.08, m_0 = 0.5$ (b) $\alpha = 0.08, m_0 = 0.1$ (c) $\alpha = 0.20, m_0 = 0.1$ C_i ($i = 1, \dots, 4$) はノイズの i 番目のキュミュラントを表わしている。記憶回復が成功する場合のみ (a)、3 次と 4 次のキュミュラント C_3, C_4 がゼロになっている。

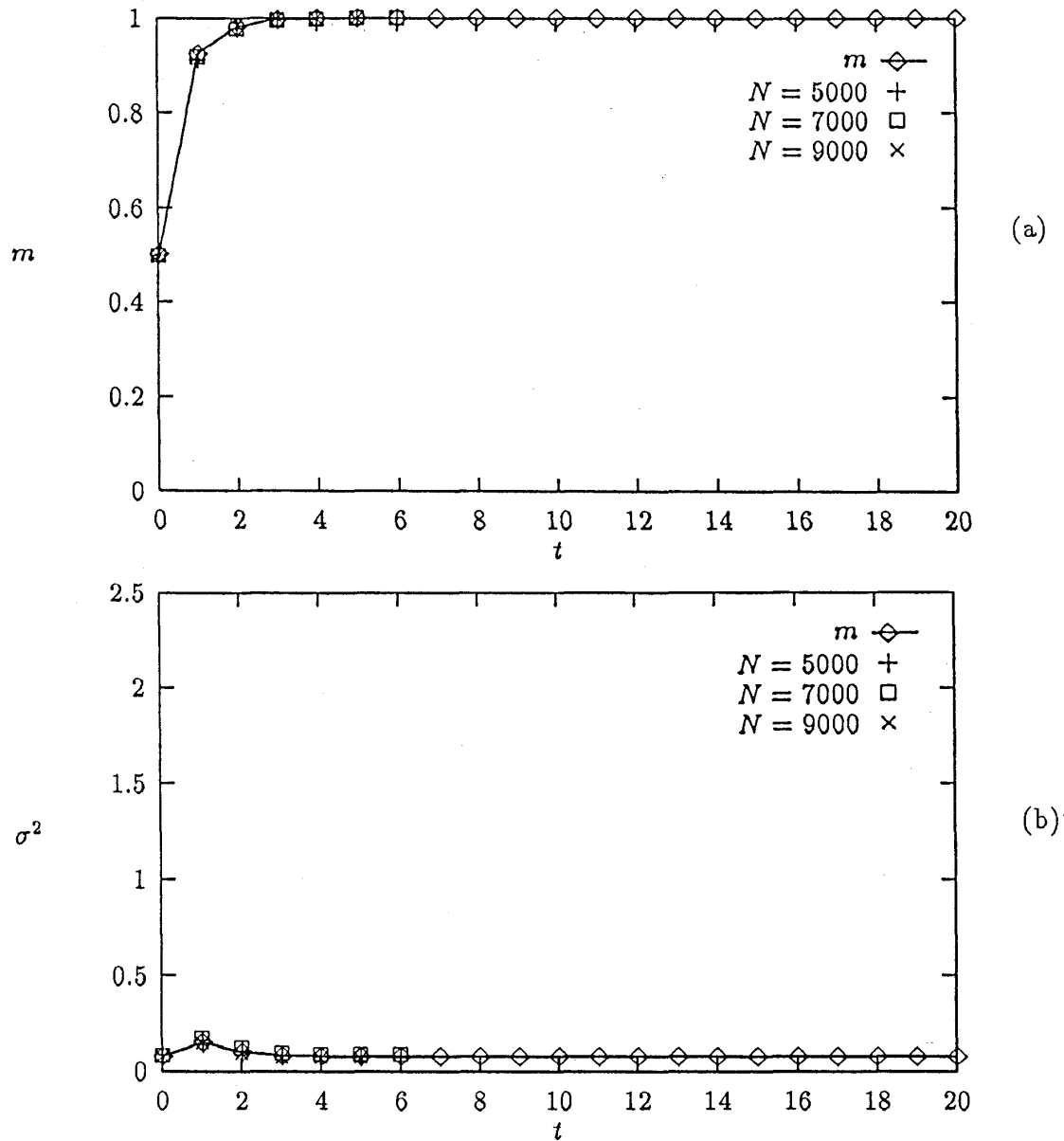


図 8: 巨視的状态変数の動的振る舞い; $\alpha = 0.08, m_0 = 0.5$ ダイヤモンドの印の付いた実線は甘利と馬被によってノイズの平均がゼロであると仮定して求められた巨視的変数 m_t と σ_t^2 の振る舞いを表わしている。その他の印はシミュレーションの結果を表わしている。パターンが回復されるときは甘利と馬被の理論は動的な振る舞いをうまく記述しているといえる。

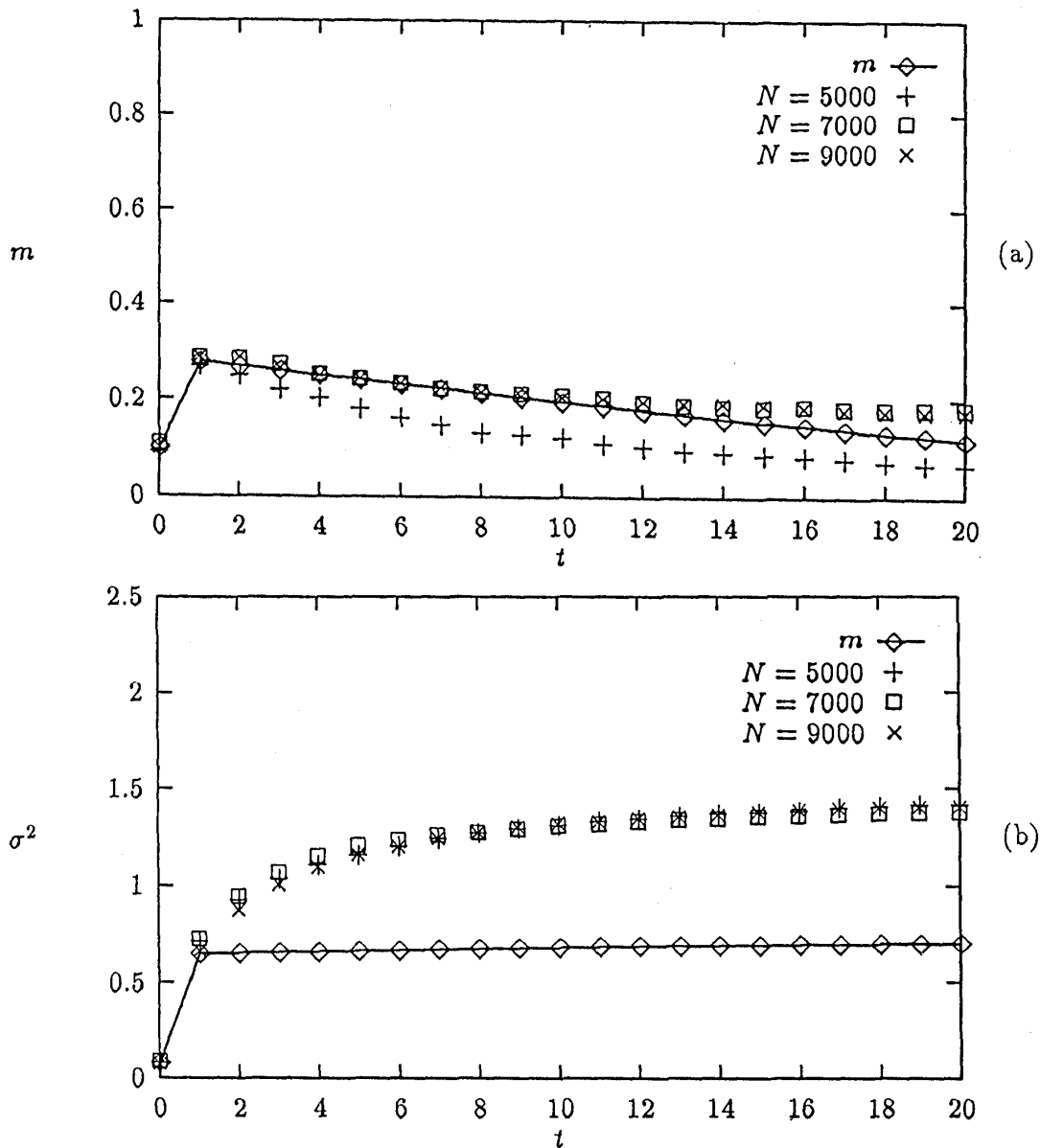


図 9: 巨視的状态変数の動的振る舞い; $\alpha = 0.08, m_0 = 0.1$ ダイヤモンドの印の付いた実線は甘利と馬被によってノイズの平均がゼロであると仮定して求められた巨視的変数 m_t と σ_t^2 の振る舞いを表わしている。その他の印はシミュレーションの結果を表わしている。甘利と馬被の理論によって求められたものはシミュレーションによる結果と定性的に一致するが、定量的には一致しない。

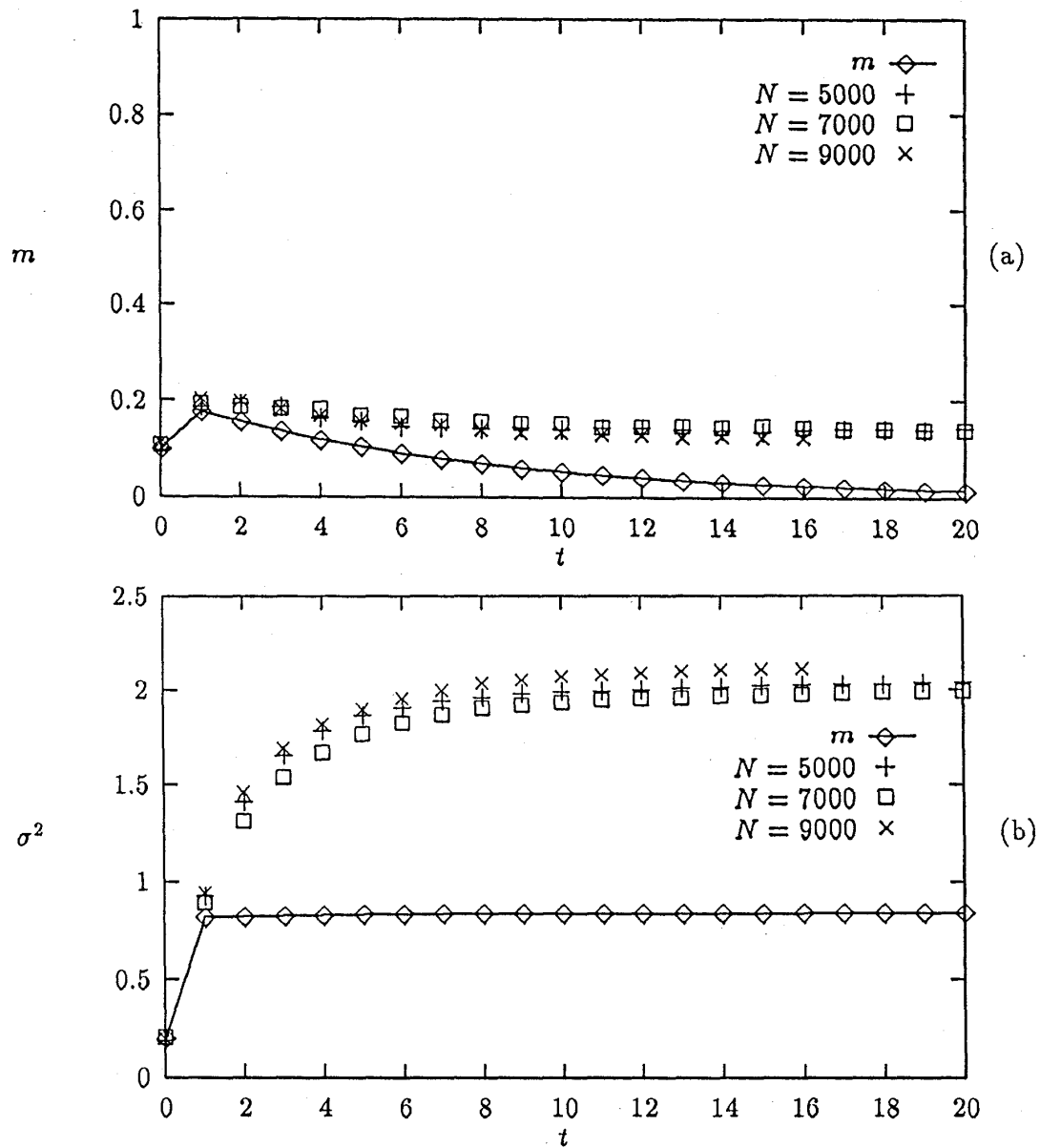


図 10: 巨視的状态変数の動的振る舞い; $\alpha = 0.20, m_0 = 0.1$ ダイヤモンドの印の付いた実線は甘利と馬被によってノイズの平均がゼロであると仮定して求められた巨視的変数 m_t と σ_t^2 の振る舞いを表わしている。その他の印はシミュレーションの結果を表わしている。甘利と馬被の理論は記憶回復が成功しないときには厳密な意味で動的な振る舞いを記述しているとはいえない。

尾関 智子

3 有限温度における甘利・馬被の理論

この章では、神経細胞が確率的な動作を行う場合について考える。それぞれの神経細胞がある一定の確率で状態更新をする。この章では甘利と馬被の理論を有限温度（確率的ダイナミクス）に拡張して、相図を描く。そして得られた結果をアミットらの結果 [3] やフォンタナリらの結果 [6] と比較する。

3.1 有限温度への拡張

個々の神経細胞の状態更新が確率的な過程を考える。温度の概念を熱的ノイズとして導入する。時刻 t における神経細胞 i への総入力を h_i^t とおくと、 S_i^{t+1} が 1 になる確率を次のように定義する。

$$\text{Prob}[S_i^{t+1} = 1] = \frac{1}{1 + \exp(-2\beta h_i^t)} \quad (69)$$

ここで β は $1/T$ であり、パラメータ T は温度（熱的ノイズ）と呼ばれるものである。すると S_i^{t+1} の期待値は

$$\begin{aligned} E[S_i^{t+1}] &= 1 \times \frac{1}{1 + \exp(-2\beta h_i^t)} + (-1) \times \left\{ 1 - \frac{1}{1 + \exp(-2\beta h_i^t)} \right\} \\ &= \tanh[\beta h_i^t] \end{aligned} \quad (70)$$

となる。ここで h_i^t は次のように表わせる。

$$\begin{aligned} h_i^t &= \frac{1}{N} \sum_{j \neq i} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} S_j^t \\ &= m_t + N_i^t \end{aligned} \quad (71)$$

甘利と馬被の理論に従って、 N_i^t が平均 0、分散 σ_i^2 のガウス分布に従う変数であると仮定する。

$$h_i^t = m_t + \sigma_t \varepsilon_i \quad (72)$$

ε_i は平均 0、分散 1 のガウス分布に従う確率変数である。以下で巨視的状态変数の方程式を求める。

$$\begin{aligned} m'_{t+1} &= \frac{1}{N} \sum_i S_i^{t+1} = E[S_i^{t+1}] \\ &= \tanh \beta(m_t + \sigma_t \varepsilon_i) \end{aligned} \quad (73)$$

ここでさらに ε_i についても平均をとらなければならない。よって

$$m_{t+1} = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(m_t + \sigma_t z) \quad (74)$$

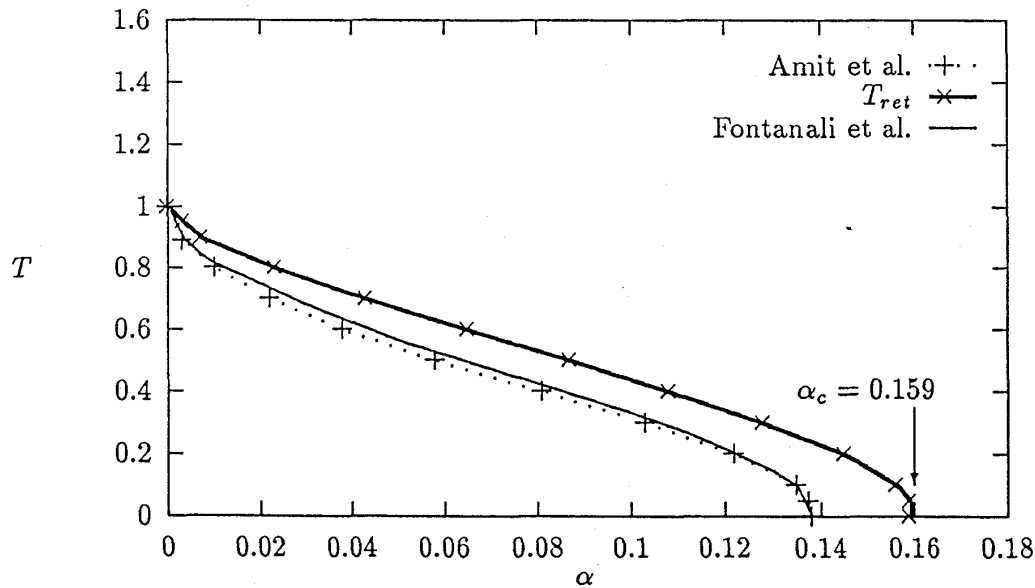


図 11: 平衡状態の相図; 太線は拡張された甘利・馬被の理論から得た平衡状態の相図である。お互いに近づいている点線と実線は、それぞれ非同期的または同期的なダイナミクスを平衡統計力学で記述することによって得られた結果である。

が得られる。

ノイズの分散 σ_t に関する方程式も甘利と馬被の理論と同様に求めることができる。導出の詳細は付録に示されている。

$$\sigma_{t+1}^2 = \alpha + 2\alpha\beta m_{t+1}m_t h(\beta m_t, \beta\sigma_t) + \beta^2 \sigma_t^2 h(\beta m_t, \beta\sigma_t)^2 \quad (75)$$

$$h(\beta m_t, \beta\sigma_t) = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \text{sech}^2 \beta(m_t + \sigma_t z) \quad (76)$$

3.2 相図

$t \rightarrow \infty$ の極限で式 (74) と (75) を数值的に解いて、平衡状態に於ける相図を得た。 T_{ret} は回復相とスピングラス相の間の一次転移の温度である。 T_{ret} 以下の温度で、神経回路網は連想記憶の役目を果たす。

驚くべきことに、図 5 に見られるように拡張された甘利・馬被の理論によって求めた相図はアミットらのものと非常によく似ている。定量的な違いは見られるものの定性的には非常によく似通っている。

尾関 智子

同期的ダイナミクスの平衡状態の性質は平衡統計力学によっても求められる。3.3.3 節で拡張された甘利・馬被の理論と平衡統計力学を比較検討する。

3.3 議論

3.3.1 拡張された甘利・馬被理論の平衡状態の性質

$t \rightarrow \infty$ の極限で、神経回路網は平衡状態に達すると考えられる。そこで m_t と σ_t を m と σ で表すことにすると、式 (74) と (75)、(76) は次のようになる。

$$m = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(m + \sigma z) \quad (77)$$

$$\begin{aligned} h(\beta m, \beta \sigma) &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \operatorname{sech}^2 \beta(m + \sigma z) \\ &= 1 - \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(m + \sigma z) \\ &\equiv 1 - \bar{q} \end{aligned} \quad (78)$$

$$\sigma^2 = \alpha + 2\alpha\beta m^2(1 - \bar{q}) + \beta^2 \sigma^2(1 - \bar{q})^2 \quad (79)$$

これらの結果を平均場理論による結果 (62)(63)(64) と比較するために、 σ を $\sqrt{\alpha r}$ と考える。すると $\bar{q}$ は q に、式 (79) は次のようになる。

$$\alpha r = \alpha + 2\alpha\beta m^2(1 - q) + \alpha\beta^2 r(1 - q)^2 \quad (80)$$

この式を r について解くと、

$$r = \frac{1 + 2\beta m^2(1 - q)}{1 - \beta^2(1 - q)^2} \neq \frac{q}{(1 - \beta + \beta q)^2} \quad (81)$$

となる。これより拡張された甘利と馬被の理論による方程式は、平均場理論によるものとは異なっていることが分かる。その結果として拡張された甘利・馬被理論は、常磁性相を持たない。

常磁性相は $m = q = 0$ で記述される。そこで式 (77)(78)(81) に $m = q = 0$ を代入する。

$$0 = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(\sqrt{\alpha r} z) \quad (82)$$

$$0 = \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(\sqrt{\alpha r} z) \quad (83)$$

$$r = \frac{1}{1 - \beta^2} \quad (84)$$

r はゼロにはなり得ないので、式 (83) の右辺はゼロより大きくなる。すると式 (83) は成り立たず、式 (77)(78)(81) は常磁性解を持たない。

3.3.2 同期的ダイナミクスの自由エネルギー

神経回路網が I という状態にある時、 i 番目の神経細胞への総入力を h_i^I で表すことにする。そのとき i 番目の神経細胞が次のタイムステップで S_i になる確率は次のように定義される。

$$\Pr(S_i) = \frac{\exp(\beta h_i^I S_i)}{\exp(\beta h_i^I) + \exp(-\beta h_i^I)} \quad (85)$$

すると神経回路網の状態が I から J へと変わる遷移確率は個々の神経細胞の遷移確率の積で表される。

$$W(J/I) = \prod_{i=1}^N \Pr(S_i^J) = \frac{\exp(\sum_i \beta h_i^I S_i(J))}{\prod_i [\exp(\beta h_i^I) + \exp(-\beta h_i^I)]} \quad (86)$$

詳細つりあいを決定する $W(I/J)$ と $W(J/I)$ の比は

$$\frac{W(I/J)}{W(J/I)} = \frac{\exp(\sum_i \beta h_i^J S_i(I))}{\exp(\sum_i \beta h_i^I S_i(J))} \times \frac{\prod_i [\exp(\beta h_i^I) + \exp(-\beta h_i^I)]}{\prod_i [\exp(\beta h_i^J) + \exp(-\beta h_i^J)]} \quad (87)$$

となる。右辺の初めの部分は次のように変形できる。

$$\begin{aligned} & \exp(\sum_i \beta h_i^J S_i(I) - \sum_i \beta h_i^I S_i(J)) \\ &= \exp(\beta \sum_{ij} J_{ij} S_j(J) S_i(I) - \beta \sum_{ij} J_{ij} S_j(I) S_i(J)) \\ &= \exp(\beta \sum_{ij} (J_{ij} - J_{ji}) S_j(J) S_i(I)) \end{aligned} \quad (88)$$

シナプス結合は対称的であるので、この部分は 1 になる。その結果として

$$\frac{W(I/J)}{W(J/I)} = \frac{F(I)}{F(J)} \quad (89)$$

が得られる。ここで

$$F(I) = \prod_i [2 \cosh \beta h_i^I] = \exp[\sum_i \ln 2 \cosh(\beta \sum_{j \neq i} J_{ij} S_j^I)] \quad (90)$$

である。詳細つりあいが満たされていると仮定すると、 $t \rightarrow \infty$ の極限で、漸近的な確率分布は次のようになる。

$$\Pr(\{S\}) \propto \exp[\sum_i \ln 2 \cosh(\beta \sum_{j \neq i} J_{ij} S_j)] \quad (91)$$

ここから有効エネルギーを読みとることができ、

$$\tilde{E} = -\frac{1}{\beta} \sum_i \ln 2 \cosh(\beta \sum_{j \neq i} J_{ij} S_j) \quad (92)$$

尾関 智子

これを使って、分配関数と自由エネルギーを表すことができる。

$$Z = \text{Tr}_S \text{Pr}(\{S\}) \quad (93)$$

$$f = -\frac{1}{\beta N} \ln Z \quad (94)$$

ここで Tr_S は全ての神経回路網の状態についての和を表している。

3.3.3 同期的なダイナミクスの平均場理論

フォンタナリら [6] は式 (92) を使って同期的ダイナミクスの自由エネルギーを計算し、相図を決定した。トレース Tr_S をとるために、次のような恒等式を用いた。

$$\begin{aligned} \text{Tr}_S \text{Pr}(\{S\}) &= \text{Tr}_S \exp\left[\sum_i \ln 2 \cosh\left(\beta \sum_{j \neq i} J_{ij} S_j\right)\right] \\ &= \text{Tr}_{S, \sigma} \exp\left[\beta \sigma_i \sum_{j \neq i} J_{ij} S_j\right] \end{aligned} \quad (95)$$

ここで $\sigma_i = \pm 1$ ($i = 1, \dots, N$) である。アミットら [3] の方法に従ってレプリカ法で自由エネルギーを計算し、アミットらが求めたものと非常によく似た相図を得ている。(図 11)

このことから分かるように、同期的ダイナミクスであるか非同期であるかは、拡張された甘利・馬被の理論による相図と平均場理論によるものとの量的な違いの原因とはなり得ない。2章で議論したように、甘利と馬被の理論は記憶回復が成功しない場合に有効ではない。したがって T_{ret} の境界領域で甘利・馬被の理論が成り立っているかは疑わしい。さらに平均場理論もいくつかの仮定を含んでいる。最も注意すべきことは、レプリカ法を用いること自体、数学的な厳密性を欠くということである。結局、平衡状態を記述するのに厳密な意味でどちらの理論を用いるのが適当であるかを決定することは容易ではない。しかしながら、最も重要な点は全く異なったアプローチから似通った一次転移の温度を得たということである。

4 TDGL 方程式による方法

これまで有限温度に拡張された甘利・馬被の理論によって平衡統計力学により得られた平衡状態の性質と非常によく似た結果が得られたことを述べてきた。特に平衡状態を動的な方程式から記述できたということは特筆すべきことである。そこで自然に生まれてくる問題として次のようなことが上げられる。平衡状態の平均場理論を基にしてそこから非平衡状態を記述できないかということである。この章では TDGL 方程式を用いて平均場理論からダイナミクスを記述する。

以下では一つのパターン (ξ^1) を回復する過程を考える。

$$m^\mu = m\delta_{\mu 1}$$

このとき自由エネルギーの式 (52) は次のようになる。

$$\begin{aligned} f = & \frac{1}{2}\alpha + \frac{1}{2}m^2 + \frac{\alpha}{2\beta}[\ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q}] \\ & + (\alpha\beta r/2)(1 - q) - \beta^{-1} \int (dz/\sqrt{2\pi}) \exp(-z^2/2) \\ & \times \ln 2 \cosh \beta(\sqrt{\alpha r}z + m) \end{aligned} \quad (96)$$

4.1 TDGL方程式

TDGL方程式 (Time-Dependent Ginzburg-Landau 方程式) は、現象論的方程式である。自由エネルギー (96) は、秩序変数の関数であり、TDGL方程式は、神経回路網の状態の自由エネルギーの谷での最急降下を記述する。

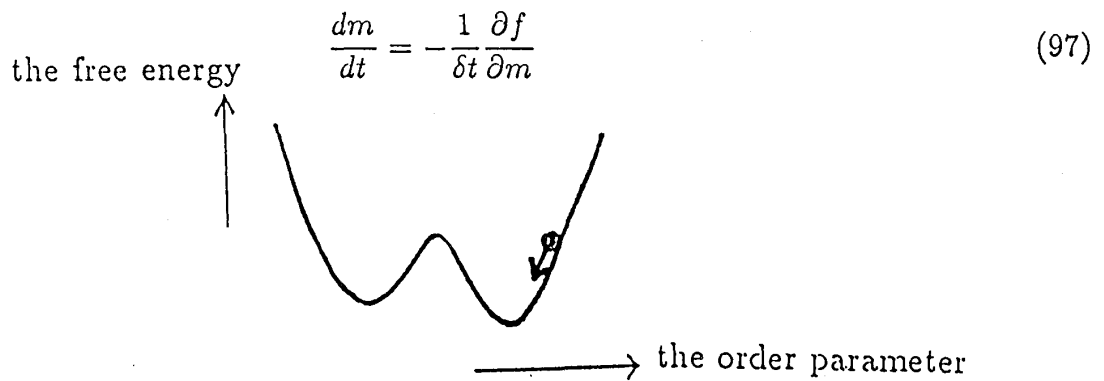
$$\frac{dm}{dt} = -\frac{1}{\delta t} \frac{\partial f}{\partial m} \quad (97)$$


図 12: TDGL方程式

この神経回路モデルの巨視的状态は三つの秩序変数 (m, r, q) で記述されている。従って以下のような三つの方程式を解くことができれば、秩序変数の動的な振る舞いを知ることができる。

$$\frac{dm}{dt} = -\frac{1}{\delta t} \frac{\partial f}{\partial m} = -\frac{1}{\delta t} \left\{ m - \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh \beta(m + \sqrt{\alpha r}z) \right\} \quad (98)$$

$$\frac{dr}{dt} = -\frac{1}{\delta t} \frac{\partial f}{\partial r} = -\frac{\alpha\beta}{2\delta t} \left\{ -q + \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tanh^2 \beta(m + \sqrt{\alpha r}z) \right\} \quad (99)$$

$$\frac{dq}{dt} = -\frac{1}{\delta t} \frac{\partial f}{\partial q} = -\frac{\alpha\beta}{2\delta t} \left\{ \frac{q}{(1 - \beta + \beta q)^2} - r \right\} \quad (100)$$

TDGL方程式は現象論的な性質を持つが、無限レンジの相互作用を持つイジングモデルではTDGL方程式が厳密な意味で正しいことが分かっている [8]。

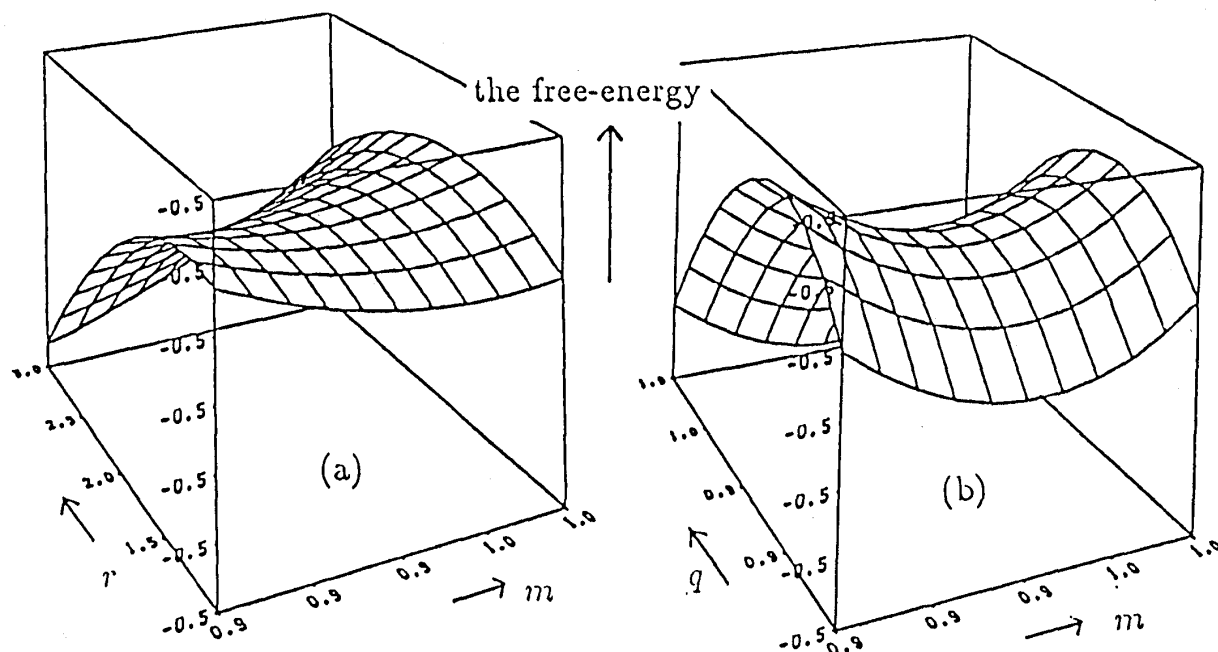


図 13: 自由エネルギーの描像; (a) 変数の内の一つ q を q_e に固定して、自由エネルギーを m と r の関数として (m_e, r_e, q_e) のまわりで描いた。(b) ここでは r を r_e に固定している。平衡解 (m_e, r_e, q_e) は自由エネルギーの鞍点にあることが分かる。

この方法はホップフィールドモデルの場合にはいくつかの問題点を含んでいる。平衡状態 (m_e, r_e, q_e) に関する方程式 (62)(63)(64) は不安定になっている。つまり平衡状態が自由エネルギーの極小ではなく、鞍点になっている。図 13 は、 (m_e, r_e, q_e) のまわりの自由エネルギーの様子を、一つの変数だけ平衡解に固定して残りの変数の関数として、記述したものである。この図を見ると分かるように、自由エネルギーは変数 m のゆらぎに対しては安定であるが、変数 r と q に関してはそうになっていない。従って（自由エネルギーの谷へと近づく様子を記述する）TDGL 方程式を用いたのでは、回復過程の動的振る舞いを記述することができないということは明らかである。それゆえ TDGL 方程式は、神経回路網の状態が平衡解 (m_e, r_e, q_e) へと近づくように修正されなければならない。

4.2 修正された TDGL 方程式

自由エネルギーは m の揺らぎに関しては安定であるので、 m に関する方程式はそのままよい。この章では、 r と q についての方程式を修正する。

まず、次のような行列を定義する。

$$M = \begin{bmatrix} \partial^2 f / \partial r^2 & \partial^2 f / \partial r \partial q \\ \partial^2 f / \partial r \partial q & \partial^2 f / \partial q^2 \end{bmatrix} \quad (101)$$

この行列の固有値の内の一つは、負である。

(r_e, q_e) のところに原点、行列 M の固有値が負の方向に x 軸、もう一つの固有ベクトルの方向に y 軸を置く。そして変数 r と q から x と y に変数変換を行い、神経回路網の状態が y の方向の揺らぎに対して安定になるように、 y についての微分方程式の符号を変える。

$$\frac{dm}{dt} = -\frac{\partial f(m, x, y)}{\partial m} \quad (102)$$

$$\frac{dx}{dt} = -\frac{\partial f(m, x, y)}{\partial x} \quad (103)$$

$$\frac{dy}{dt} = +\frac{\partial f(m, x, y)}{\partial y} \quad (104)$$

こうしてこれらの連立微分方程式を解き、変数を r と q に戻せば秩序変数の動的振る舞いを調べることができる。

4.3 結果と議論

4.3.1 秩序変数の時間発展の様子

ホップフィールドモデルのTDGL方程式は三つの秩序変数を含む。したがって、微分方程式(102)(103)(104)を解くとき、初期状態として三つの秩序変数の様々な組み合わせを選ばなければならない。

まず始めに神経回路網がランダムな初期状態から出発する場合を考える。つまり回路網の初期状態が一番目の記憶させたパターン ξ^1 と有限の方向余弦を持つが、それ以外のパターンとは無関係な初期状態を考える。この場合に初期値 r_0 を評価する。回路網の状態が一番目のパターン以外とは無関係なので、 m_0^μ ($\mu \neq 1$) の平均はゼロになる。

$$\langle m_0^\mu \rangle = \frac{1}{N} \langle \sum_i \xi_i^\mu S_i^0 \rangle = 0 \quad (105)$$

さらに分散は

$$\langle m_0^{\mu 2} \rangle = \frac{1}{N^2} \langle \sum_i \sum_j \xi_i^\mu \xi_j^\mu S_i^0 S_j^0 \rangle \quad (106)$$

であり、 $i \neq j$ の項からの寄与はゼロである。従って分散は次のようになる。

$$\langle m_0^{\mu 2} \rangle = \frac{1}{N^2} \times N = \frac{1}{N} \quad (107)$$

ランダムな状態とパターンとの方向余弦は $O(1/\sqrt{N})$ のオーダーを持つ。よって、 r の定義より、 r_0 は1になる。

$$r_0 = \frac{1}{\alpha} \sum_{\mu \neq 1} \langle m_0^{\mu 2} \rangle \simeq \frac{1}{\alpha} \cdot P \frac{1}{N} = 1 \quad (108)$$

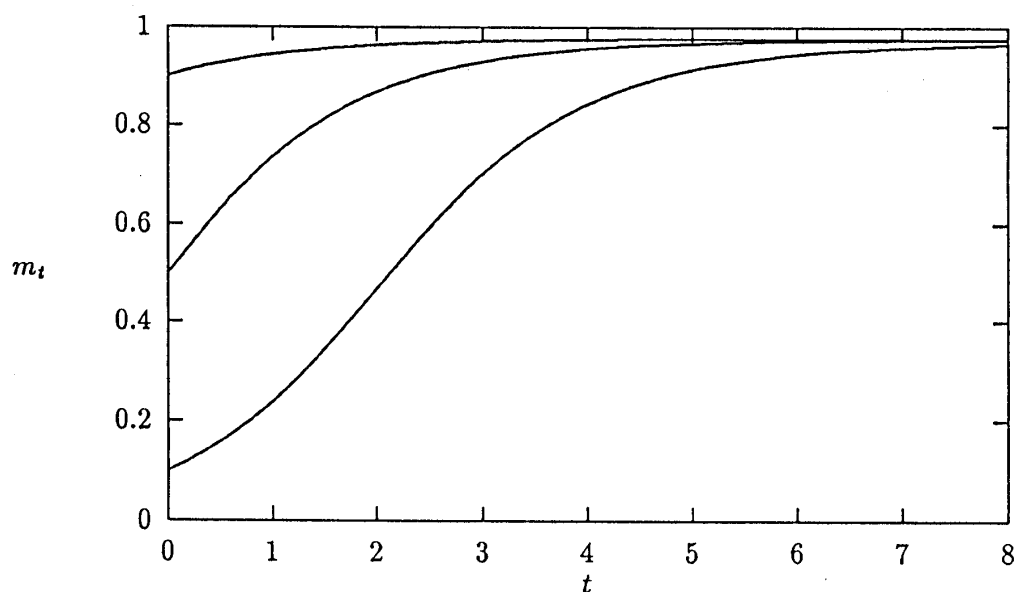


図 14: TDGL方程式によって導かれた方向余弦の動的振る舞い; $\alpha = 0.1, T = 0.3$ (回復相)
 $r_0 = 1, q_0 = 0.8$

様々な初期値の組み合わせ (m_0, q_0) に対して、式(102)-(104)を解いた。方向余弦 m の振る舞いは q の初期値に依存しないようであったので、 q を q_0 に固定し m_0 の値はいろいろと変えて方向余弦の時間依存した振る舞いを調べた。図 14に見られるように、 m の初期値に依存した閾値現象は見られなかった。つまりいかなる初期状態から出発しても回復相においては回復が成功してしまう。

次に初期値 r_0 をいろいろに変えて調べる。 $(m_0$ と q_0 は固定しておく。) 初期値 r_0 に依存した振る舞いが図 15に見られる。

1. 例え m_0 が小さくても、 r_0 が小さいならば記憶させたパターンは回復される。
2. 逆に r_0 が大きいと、パターンと有限の方向余弦を持ったどんな初期状態から出発してもパターンは回復されない。

これらの現象は r の定義から直観的に理解することができる。変数 r は回復されないパターンからのノイズであると考えることができる。したがってノイズが小さければ記憶させたパターンは回復される。しかしノイズが大きくなるとどんなにパターンに近い初期状態から出発させても回復は成功しない。

4.3.2 $m - r$ 平面

m と r の時間発展の様子を $m - r$ 平面に軌跡として描いた。(図 16) この図から初期状態に依存した回路網の動的な振る舞いが一目で分かる。 $o(m_e, r_e)$ は回復解、 $\bullet(m_e, r_e)$ はスピングラス解を意味する。これらの解は動的過程のアトラクターになっていて、回路網の状態はこの解のうちの一つに流れ込む。図の中には $dm/dt = 0$ と $dr/dt = 0$ の線も描いてある。これらの線の下領域では dm/dt と dr/dt の符号はそれぞれ正になっていて、 m と r は時間とともに増加する。さらに回路網の状態の流れを矢印で表わした。これより回復領域が分かる。一般に回復領域を求めるのは難しいが、修正された TDGL 方程式によって回復領域を容易に求めることができた。(図 17)

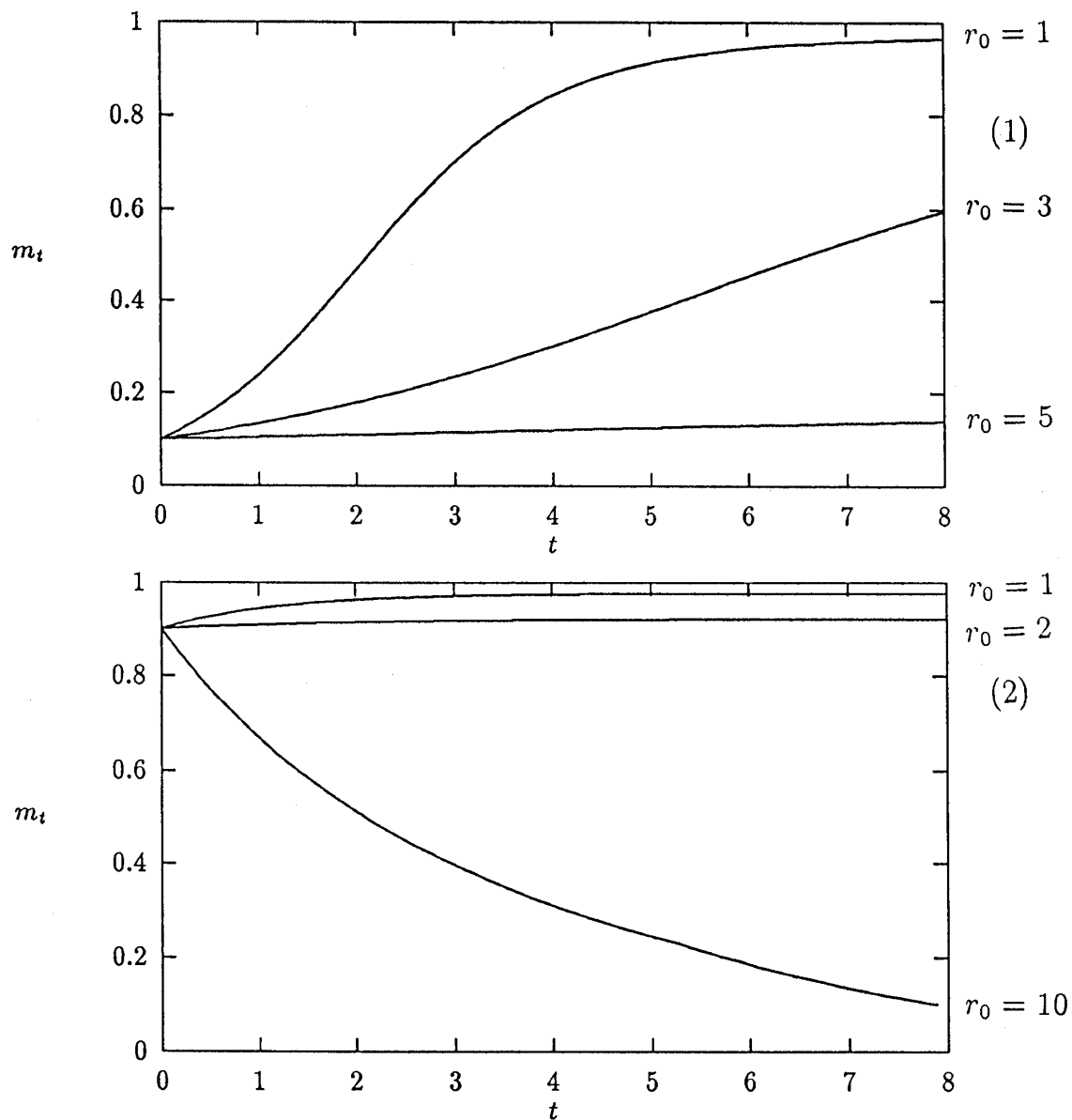


図 15: 方向余弦 m の初期値 r_0 への依存性; $\alpha = 0.1, T = 0.3$ (回復相) (1) $m_0 = 0.1, q_0 = 0.8$ たとえば m_0 が小さくてもパターンは回復されない。(2) $m_0 = 0.9, q_0 = 0.8$ r_0 が大きいと、どんな状態から出発してもパターンは回復されない。

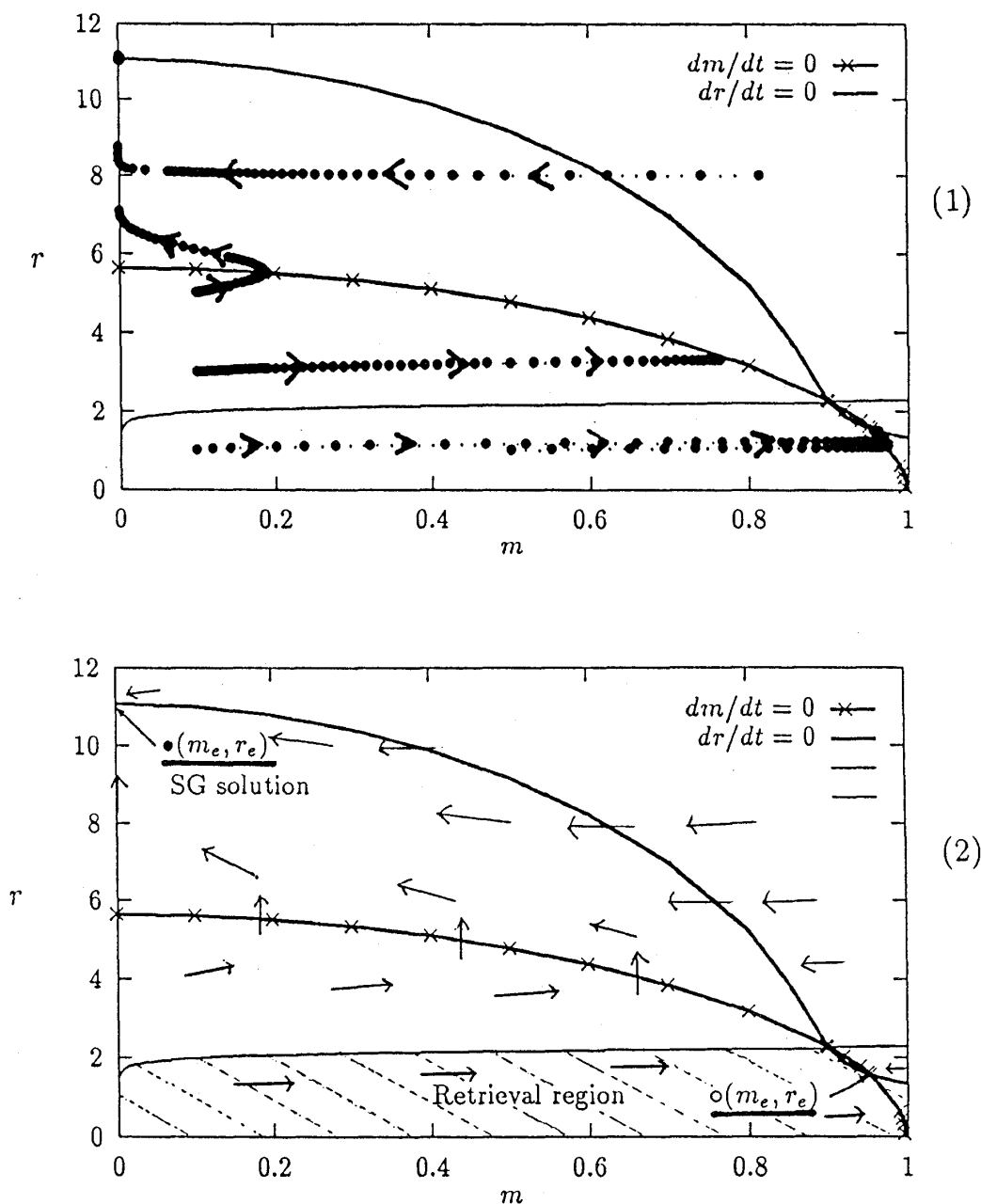


図 16: $m-r$ 平面における m と r の時間発展; $\alpha = 0.1, T = 0.3$ (回復相) (1) 初期状態に依存した回路網の動的振る舞い (2) $o(m_e, r_e)$ は回復解、 $\bullet(m_e, r_e)$ はスピングラス解を意味する。これらの解はダイナミクスのアトラクターになっていて、回路網の状態はこの解のうちの一つに流れ込む。さらに回路網の状態の流れを矢印で表わした。

尾関 智子

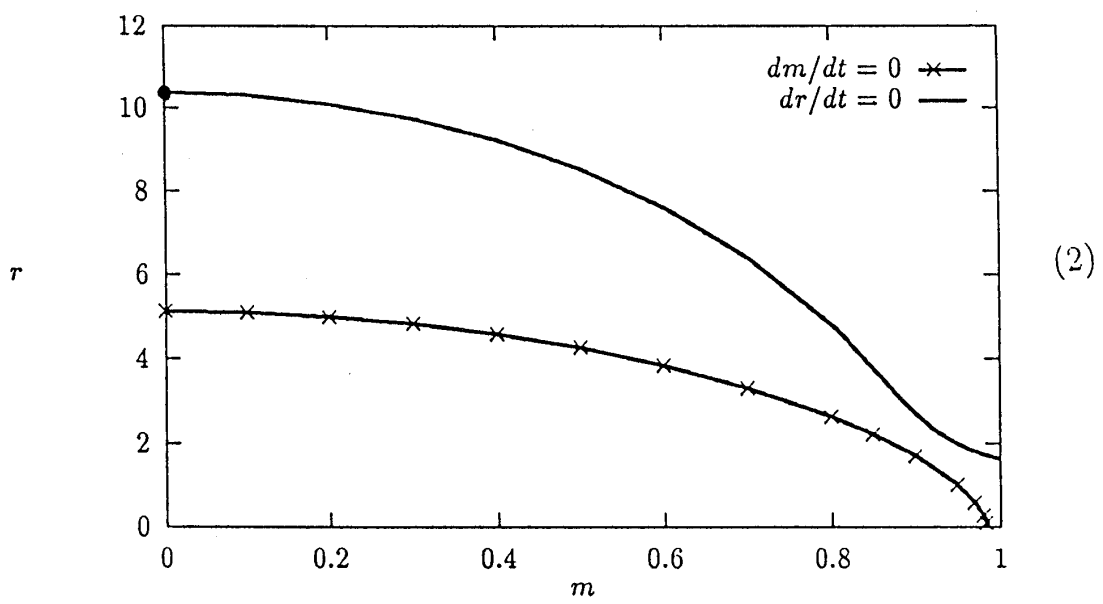
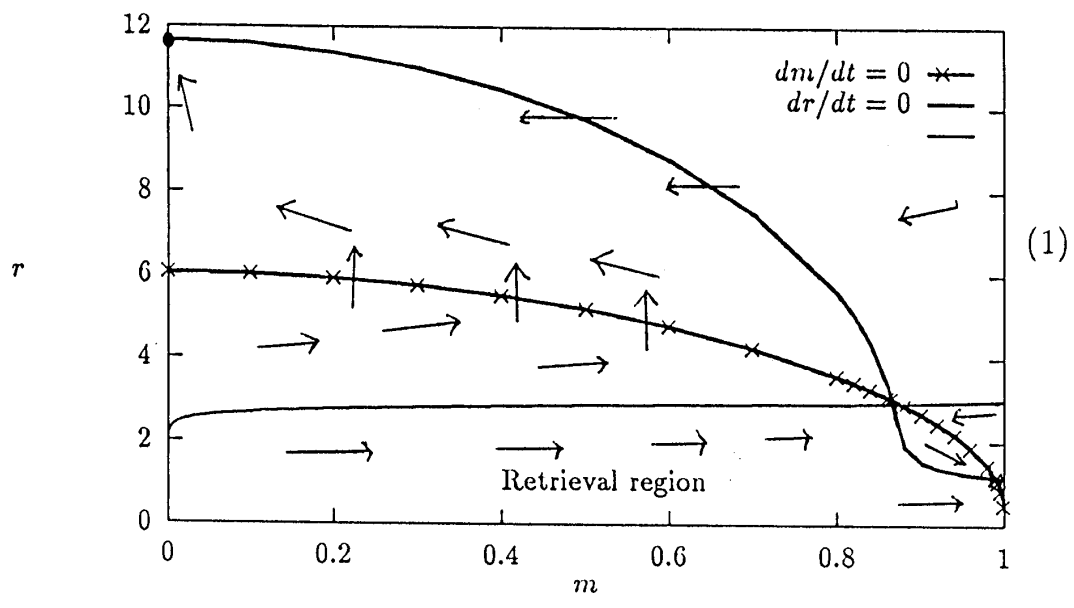


図 17: $m - r$ 平面上の回復領域; (1) $\alpha = 0.1, T = 0.2$ (回復相) (2) $\alpha = 0.1, T = 0.4$ (スピングラス相)

5 議論

ホップフィールドモデルの動的及び静的な性質を二つの方法によって論じた。一つの方法は甘利・馬被の理論であり、もう一つはTDGL方程式による方法である。

モンテカルロシミュレーションを行ない、ノイズの項がガウス分布に従うという甘利・馬被の仮定は記憶回復が成功する場合にのみ有効であることが分かった。実は σ_i^2/α は r に等しいので、図16と比較するため、 m と σ_i^2/α の時間発展をプロットした。(図18) この図を見ると甘利・馬被の理論は絶対零度での回復過程をうまく記述していることがよりいっそう明らかである。さらに付け加えておくと本来ゼロでないノイズの平均 $E[N_i^2]$ まで考慮に入れた甘利・馬被の理論はダイナミクスをうまく記述できないように見えたが、実は2ステップめまではガードナーら[4]の求めた厳密解と一致している。3章では甘利・馬被の理論を有限温度に拡張して平衡平均場理論によるものと一致する結果を得た。拡張された甘利・馬被の理論の利点は次のとおりである。

1. 回復過程のダイナミクスを直接扱うことができる。
2. 仮定が単純明快であるのでコンピューターシミュレーションなどによりその妥当性の検討が可能である。

有限温度に拡張された甘利・馬被の理論から求めた動的な振る舞いを図19に示した。回復相では方向余弦の初期値による閾値現象が有限温度でも現われている。この結果はTDGL方程式によるもの(図14)と一致しない。しかし本来TDGL方程式は回路網がほとんど平衡状態に達している時の振る舞いを記述するものであることを考えれば、初期状態に依存した振る舞いに差違が見られてもさほど深刻な問題ではない。

4章では、TDGL方程式を導入し、 $m-r$ 平面を描くことによって回復領域を決めることができた。TDGL方程式による方法には限界があるかも知れない。TDGL方程式は平衡状態での自由エネルギーを使っているので、厳密な意味で非平衡状態を記述することができるとはいえない。しかしながら動的な過程においても正しいことがいえれば、 $m-r$ 平面から回復領域を容易に求めることができるので有効な手段だといえる。リーデルら[9]の方法との関係も明らかにする必要がある。

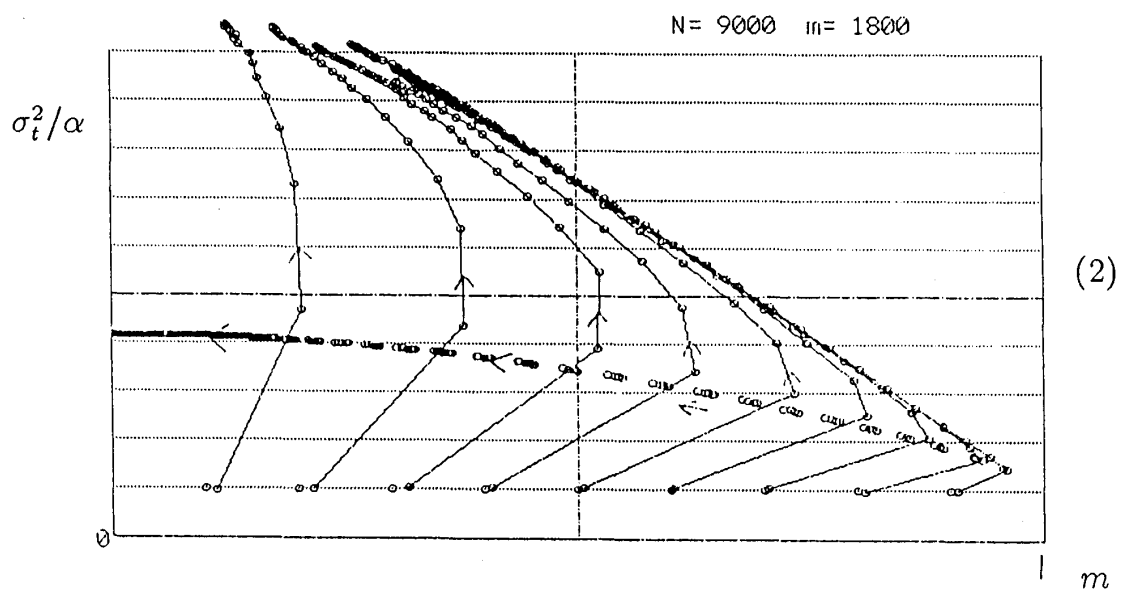
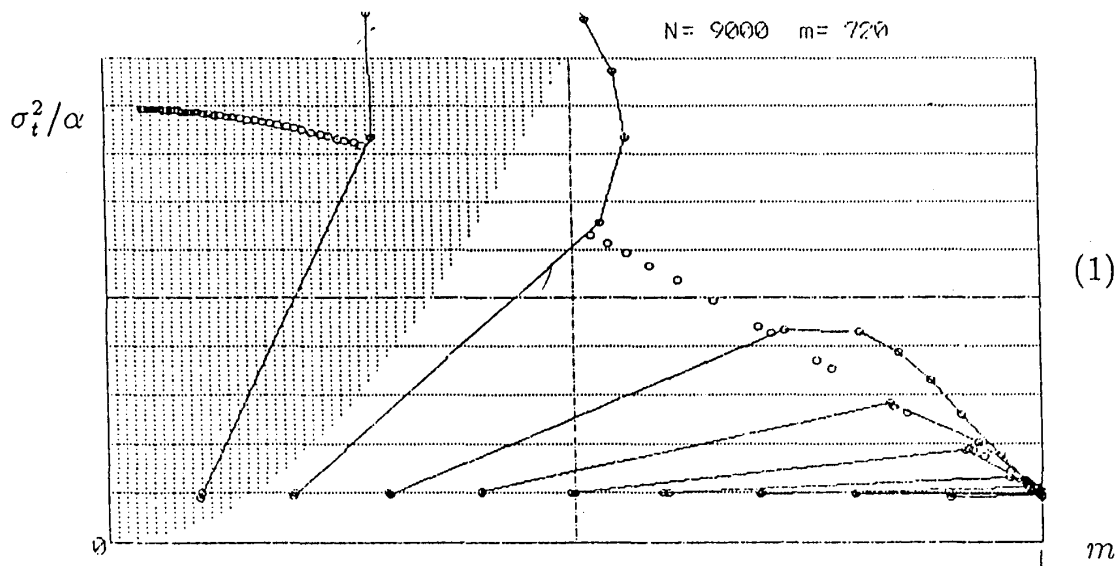


図 18: m と σ_t^2/α の時間発展; 印のついた線はシミュレーションの結果を表わしていて ($N = 9000$)、丸い印は甘利・馬被の理論によるものである。(1) $\alpha = 0.08$ 影の部分は回復できない領域である。(2) $r = 0.2$

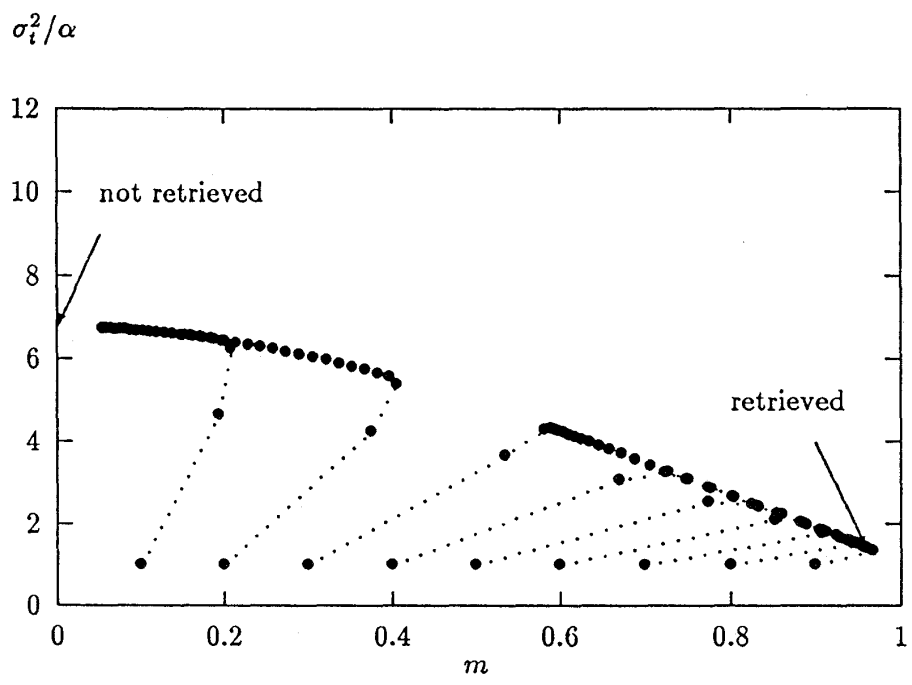


図 19: 拡張された甘利・馬被理論による動的振る舞い; $\alpha = 0.1, T = 0.3$ 閾値現象が有限温度においてもみられる。

尾関 智子

謝辞

西森秀稔助教授の絶え間ないご指導と励ましに感謝いたします。また、図 16 に基づく解析を示唆して下さった中村統太氏に感謝いたします。シミュレーションをするにあたって、伊藤伸泰氏と金田康正氏の開発された高速乱数発生ルーチン RNDTIK を利用させていただきました。この研究は一部文部省の科研費重点領域研究により援助を受けました。

付録

ここでは有限温度における σ_i^2 の時間発展方程式を求める。甘利・馬被の理論に従いノイズがガウス分布に従うという仮定の下で、 S^t の ξ^μ への依存性を考慮しなければならない。

$E[N_i^t] = 0$ が仮定されているので、分散は

$$\begin{aligned}\sigma_i^2 &= E[(N_i^t)^2] \\ &= \frac{1}{N^2} \sum_{\mu} \sum_{\mu'} \sum_j \sum_{j'} E[\xi_i^{\mu} \xi_j^{\mu} S_j^t \xi_i^{\mu'} \xi_{j'}^{\mu'} S_{j'}^t]\end{aligned}$$

となる。これらの項を四つのタイプに分ける。

$$1. \mu = \mu', j = j'$$

$$v_1 = 1; (P-1)(N-1)v_1$$

$$2. \mu \neq \mu', j = j'$$

$$v_2 = E[\xi_i^{\mu} \xi_j^{\mu} \xi_i^{\mu'} \xi_j^{\mu'}] = 0$$

$$3. \mu = \mu', j \neq j'$$

$$v_3 = E[\xi_j^{\mu} \xi_{j'}^{\mu} S_j^t S_{j'}^t]; (P-1)(N-1)(N-2)v_3$$

$$4. \mu \neq \mu', j \neq j'$$

$$v_4 = E[\xi_i^{\mu} \xi_j^{\mu} S_j^t \xi_i^{\mu'} \xi_{j'}^{\mu'} S_{j'}^t]; (P-1)(P-2)(N-1)(N-2)v_4$$

N と P が共に十分に大きいとき、

$$\sigma_i^2 = \alpha + \alpha N v_3 + \alpha^2 N^2 v_4$$

となる。

まず、 v_3 を評価する。 $S_j^t, S_{j'}^t$ の ξ^μ 依存性は次式で与えられる。

$$E[S_j^t] = \tanh(m_{t-1} + Q + \xi_j^{\mu} R + (1/N) \xi_j^{\mu} \xi_{j'}^{\mu} S_{j'}^{t-1})$$

$$E[S_j^t] = \tanh(m_{t-1} + Q' + \xi_j^\mu R + (1/N)\xi_j^\mu \xi_j^\mu S_j^{t-1})$$

ここで Q, Q', R は、

$$Q = \frac{1}{N} \sum_{\beta \neq 1, \mu} \sum_{k \neq j} \xi_j^\beta \xi_k^\beta S_k^{t-1}$$

$$Q' = \frac{1}{N} \sum_{\beta \neq 1, \mu} \sum_{k \neq j} \xi_{j'}^\beta \xi_k^\beta S_k^{t-1}$$

$$R = \frac{1}{N} \sum_{k \neq j, j'} \xi_k^\mu S_k^{t-1}$$

であり、仮定により相互に独立したガウス分布に従うランダム変数である。 v_3 は条件付き期待値を使って求める。

$$v_3 = EE[\xi_j^\mu \xi_{j'}^\mu S_j^t S_{j'}^t]$$

$\xi_j^\mu, \xi_{j'}^\mu$ が与えられたときの $\xi_j^\mu \xi_{j'}^\mu S_j^t S_{j'}^t$ の期待値を求めた後、 $\xi_j^\mu, \xi_{j'}^\mu$ について平均をとる。

$$v_3 = \frac{1}{4}(Y_{11} + Y_{-11} + Y_{1-1} + Y_{-1-1})$$

ここで

$$Y_{pq} = E[\xi_j^\mu \xi_{j'}^\mu S_j^t S_{j'}^t | \xi_j^\mu = p, \xi_{j'}^\mu = q]$$

である。各項は次のような式で表される。

$$Y_{11} = E[\tanh \beta(m_{t-1} + Q + R + (1/N)S_{j'}^{t-1}) \tanh \beta(m_{t-1} + Q' + R + (1/N)S_j^{t-1})]$$

$$Y_{1-1} = -E[\tanh \beta(m_{t-1} + Q + R - (1/N)S_{j'}^{t-1}) \tanh \beta(m_{t-1} + Q' - R - (1/N)S_j^{t-1})]$$

Q, Q', R は平均ゼロ、分散 1 のガウス分布に従うランダム変数を使って表すことができる。

$$Q + R = \sigma_{t-1}[u + (2P)^{-1}v]$$

$$Q' + R = \sigma_{t-1}[v + (2P)^{-1}u]$$

これより次式が得られる。

$$\begin{aligned} Y_{11} + Y_{1-1} &= \int_{-\infty}^{\infty} du dv P(u, v) \tanh \beta \sigma_{t-1}(\bar{m} + u + \tilde{v} + \tilde{x}') \tanh \beta \sigma_{t-1}(\bar{m} + v + \tilde{u} + \tilde{x}) \\ &\quad - \int_{-\infty}^{\infty} du dv P(u, v) \tanh \beta \sigma_{t-1}(\bar{m} + u - \tilde{v} - \tilde{x}') \tanh \beta \sigma_{t-1}(\bar{m} + v - \tilde{u} - \tilde{x}) \end{aligned}$$

$$\bar{m} = m_{t-1}/\sigma_{t-1}$$

尾関 智子

$$\tilde{x}' = S_{j'}^{t-1}/N\sigma_{t-1}, \tilde{x} = S_j^{t-1}/N\sigma_{t-1}$$

$$\tilde{u} = (2P)^{-1}u, \tilde{v} = (2P)^{-1}v$$

$$P(u, v) = P(u)P(v) = (2\pi)^{-1} \exp[-\frac{1}{2}(u^2 + v^2)]$$

$\tilde{v} + \tilde{x}'$ と $\tilde{u} + \tilde{x}$ は微小量なので、一次近似は次のようになる。

$$\begin{aligned} Y_{11} + Y_{1-1} &= \int_{-\infty}^{\infty} du dv P(u, v) \tanh \beta \sigma_{t-1}(\bar{m} + v) \\ &\times [\tanh \beta \sigma_{t-1}(\bar{m} + u + \tilde{v} + \tilde{x}') - \tanh \beta \sigma_{t-1}(\bar{m} + u - \tilde{v} - \tilde{x}')] \\ &+ \int_{-\infty}^{\infty} du dv P(u, v) \tanh \beta \sigma_{t-1}(\bar{m} + u) \\ &\times [\tanh \beta \sigma_{t-1}(\bar{m} + v + \tilde{u} + \tilde{x}) - \tanh \beta \sigma_{t-1}(\bar{m} + v - \tilde{u} - \tilde{x})] \end{aligned}$$

次に

$$-\bar{m} - \tilde{v} - \tilde{x}' = f(v) = f$$

$$-\bar{m} + \tilde{v} + \tilde{x}' = g(v) = g,$$

と表すと

$$\begin{aligned} &\tanh \beta \sigma_{t-1}(u - f) - \tanh \beta \sigma_{t-1}(u - g) \\ &= \tanh \beta \sigma_{t-1}(u - f) - \tanh \beta \sigma_{t-1}(u - f - (g - f)) \\ &\simeq (g - f) \beta \sigma_{t-1} \operatorname{sech}^2 \beta \sigma_{t-1}(u - f) \end{aligned}$$

となる。式 (109) の右辺の前半部分は、

$$\frac{2\beta}{N} [S_{j'}^{t-1} f(\beta m_{t-1}, \beta \sigma_{t-1}) + \frac{\sigma_{t-1}}{2\alpha} g(\beta m_{t-1}, \beta \sigma_{t-1})] \times h(\beta m_{t-1}, \beta \sigma_{t-1})$$

となる。ここで

$$\begin{aligned} f(\beta m_t, \beta \sigma_t) &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) \tanh \beta(m_t + \sigma_t z) \\ g(\beta m_t, \beta \sigma_t) &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) z \tanh \beta(m_t + \sigma_t z) \\ h(\beta m_t, \beta \sigma_t) &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) \operatorname{sech}^2 \beta(m_t + \sigma_t z) \end{aligned}$$

である。残りの部分は同様に計算でき、次のような結果を得る。

$$\begin{aligned} &Y_{11} + Y_{1-1} + Y_{-11} + Y_{-1-1} \\ &= \frac{2}{N} [2\beta(S_j^{t-1} + S_{j'}^{t-1})f(\beta m_{t-1}, \beta \sigma_{t-1}) + \frac{2\beta\sigma_{t-1}}{\alpha} g(\beta m_{t-1}, \beta \sigma_{t-1})] \\ &\quad \times h(\beta m_{t-1}, \beta \sigma_{t-1}) \end{aligned}$$

次式が成り立つので、

$$E[S_j^{t-1} + S_{j'}^{t-1}] = 2E[S_j^{t-1}] = 2m_{t-1}$$

最終的に次式を得る。

$$\alpha N v_3 = [2\alpha\beta m_t m_{t-1} + \beta\sigma_{t-1}g(\beta m_{t-1}, \beta\sigma_{t-1})] \times h(\beta m_{t-1}, \beta\sigma_{t-1})$$

v_4 も同様に計算できる。しかしながら実際は $E[N_i^t]$ はゼロではない。従って σ_i^2 を求めるには $E[(N_i^t)^2]$ から $E[N_i^t]^2$ を差し引かなければならない。ところが $E[N_i^t]^2$ は、 $N^2\alpha^2 v_4$ に等しいので、結局 σ_i^2 は次のようになる。

$$\begin{aligned}\sigma_{i+1}^2 &= \alpha + \alpha N v_3 \\ &= \alpha + 2\alpha\beta m_{t+1} m_t h(\beta m_t, \beta\sigma_t) + \beta^2 \sigma_i^2 h(\beta m_t, \beta\sigma_t)^2\end{aligned}\tag{109}$$

参考文献

- [1] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities" *Proc. Natl. Acad. Sci. U.S.A.* **79** (1982) 2554
- [2] S. Amari and K. Maginu, "Statistical Neurodynamics of Associative Memory" *Neural Networks* **1** (1988) 63
- [3] D. J. Amit, H. Gutfreund and H. Sompolinsky, "Statistical Mechanics of Neural Networks near Saturation" *Ann. Phys.* **173** (1987) 30
- [4] E. Gardner, B. Derrida and P. Mottishaw, "Zero Temperature Parallel Dynamics for Infinite Range Spin Glasses and Neural Networks" *J. Physique A* **48** (1987) 741
- [5] D. Sherrington and S. Kirkpatrick, "Solvable Model of a Spin-Glass" *Phys. Rev. Lett.* **35** (1975) 1792
- [6] José F. Fontanari and R. Köberle, "Information Storage and Retrieval in Synchronous Neural Networks" *Phys. Rev. A* **36** (1987) 2475
- [7] E. Domany, J. L. van Hemmen and K. Schulten, "Models of Neural Networks" (Springer-Verlag, 1991)
- [8] D. J. Amit, "Modeling Brain Function" (Cambridge University Press, 1989)
- [9] U. Riedel, R. Kühn and J. L. Van Hemmen, *Phys. Rev. A* **38** (1988) 1105