

神経回路モデルによる脳の記憶機構の研究

森 田 昌 彦

①

目次

神経回路モデルによる
脳の記憶機構の研究

森田 昌彦

目次

第1章 序論	1
1.1 記憶へのアプローチ	1
1.1.1 はじめに	1
1.1.2 記憶研究の戦略	2
1.1.3 本研究の目的と方法	4
1.2 研究の背景	5
1.2.1 神経回路網の力学に関する問題点	6
1.2.2 生理学的知見	7
1.3 論文の構成	8
第2章 連想記憶の力学	9
2.1 連想記憶モデル	10
2.1.1 自己相関モデル	10
2.1.2 想起ダイナミクス	11
2.1.3 その他のモデル	12
2.2 力学的性質	13
2.2.1 記憶容量	13
2.2.2 平衡点の分布	14
2.2.3 想起過程	15
2.3 考察	19
2.3.1 想起過程の直観的解釈	19
2.3.2 モデルの問題点と改良法	23

第3章 ダイナミクスを改良した神経回路網	26
3.1 部分反転法	27
3.1.1 基本的考え方	27
3.1.2 アルゴリズム	28
3.1.3 数値実験	31
3.1.4 理論的解析	36
3.2 連続型モデル	40
3.2.1 非単調ダイナミクス	41
3.2.2 数値実験と考察	42
3.3 構造をもつパターンの連想記憶	47
3.3.1 自己相関モデルと情報表現	47
3.3.2 クラスタを成すパターンの想起	48
3.3.3 考察	52
3.4 非単調ダイナミクスを利用した学習	55
3.4.1 相関学習と直交学習	55
3.4.2 学習則	56
3.4.3 数値実験	58
第4章 側頭葉短期記憶系の力学とそのモデル	60
4.1 背景	61
4.1.1 TE野の短期記憶ニューロン	61
4.1.2 系の力学的性質	62
4.1.3 従来の力学系との比較	62
4.2 モデルの構成	64
4.2.1 ダイナミクスの改良	65
4.2.2 モデルの構造	65
4.3 モデルの動作	68
4.3.1 スパースなパターンの連想記憶	68
4.3.2 複数のパターンの保持	74
4.4 考察	77

第5章 結論	80
5.1 本研究のまとめ	80
5.2 今後の課題と展望	81
謝辞	84
参考文献	85

序論

1.1 記憶へのアプローチ

1.1.1 はじめに

脳は、現在の科学にとって最も興味深い対象の一つであり、それを研究することの意義は言うまでもない。中でも「記憶」は、脳における情報処理の中心を成す機能である^[1]。脳の高次機能の中で記憶が関係しないものはないと言ってよいし、記憶の仕組みが明らかにならない限り、脳の高度な情報処理機能の解明もあり得ない。しかし、記憶のメカニズムは、脳研究における最大の謎とも言われる^[2]ように、その大部分が未解明のままである。

もっと素朴に考えても、人間の記憶というのは不思議なものである。なぜ我々は、過去の出来事をありありと思い浮かべることができるのだろうか。また、どうやって膨大な知識をうまく蓄えているのだろうか。こうした疑問がこの研究の出発点である。

以前、著者は同じ動機から、海馬を中心とする記憶システムのモデル化の研究を行い、ある程度の成果を得た^[3]。しかし、そこでのアプローチをそのまま進めても記憶機構の解明へは至らないという点で、それは必ずしも満足のいくものではなかった。思うに、その原因は、モデルを構築するための土台がしっかりしていないことにあった。特に、神経回路網の力学に関する理解が不足していたことによる部分が大きい。

本論文は、このような反省から行ったその後の研究の成果をまとめたものである。その中には、一見現実の脳とは関係がないように思える部分もあるかもしれないが、それは記憶の謎に迫るための基礎となる重要な部分であることを強調しておきたい。

なお、本論文では、特に重要で関係の深い事柄を除き、自らの研究成果のみを記した。生理学および心理学に関する基礎的事項に関しては、文献[1]~[3]などを参考にされたい。

1.1.2 記憶研究の戦略

脳の記憶機構を解明するために、どのような方法を取ればよいのだろうか。そして今、何が必要なのであろうか。

脳の研究は、これまで生理学や神経心理学を中心として進められてきた。特に近年の神経生理学の発展は、数多くの重要な知見をもたらした。

しかし、こうしたアプローチの限界が次第に見えてきたことも事実である。その限界とは、要するに、実験的な方法だけでは情報処理の原理の究明がほとんど不可能であることに起因している。そして、少なくとも記憶の研究に関する限り、そのような限界はすぐ近くまで来ており、従来の研究戦略はもはや通用しなくなりつつある。

一方において、最近では工学や物理学といった分野の研究者も、脳に強い関心を示すようになってきた。特に、人工的な神経回路網に関する研究は、数学的な興味や工学的応用の可能性から、一種のブームにさえなっている。しかし、これらの研究の多くは、現実の脳とは無関係に進められている。中には脳のモデル化の研究と称するものもあるが、ほとんどが都合の良い実験結果だけを引用するのみで、生理学などに対して有用な情報を提供し得ないでいる。つまり、神経生理学を中心とする実験と、数学的または工学的な理論の間には、交流がほとんどないのである。そしてこのことが、脳研究の飛躍を妨げている。

このような状況をもたらした原因の大部分が、研究の背景や立場の相違と、それによる相互理解の不足にあることは確かである。しかし、それがすべてではない。というのは、互いに理解を深めようという試みは、長い間続けられてきたからである。むしろ、積極的に理論と実験とをつなごうとする研究がなかったことに、大きな責任があると思われる。

そもそも、分野も方法論も全く異なる研究間の溝が、そう簡単に埋まらないのは当然である。それができるためには、少なくとも理論家と実験家の双方が理解できる交流の場が必要である。そして、その場とは、適切なモデルのことだと言つてよい。

理論と実験の間だけでなく、研究のレベル間にも溝が存在する^[3]。特に、生理学のレベルと心理学のレベルの間には広大なギャップがあり、これもまた記憶研究にとって大きな障害になっている。

このギャップを埋めるためには、それぞれのレベルで、その幅を広げていく以外に方法はない。もちろんそうした努力は、これまでにもなされてきた。例えば、上位のレベル、すなわち心理学からのアプローチとして、PDPグループの研究^[4]がある。

これは、神経細胞のような「ユニット」を下位のレベルに据え、それから構成したモデルによって、認識や記憶といった人間の心的動作を説明しようとするものである。多数の単純なユニットの相互作用によって情報が処理されるという点で、このモデルは従来の心理学モデルよりも実際の脳に近く、心理学としてはかなりの成功を取めた。

しかしながら、PDPのモデルは、生理学に対してあまり有益な知見をもたらしていない。その原因は、設定したユニットのレベルと、神経生理学が扱っているニューロンのレベルとのギャップが大きすぎることにある。そのため、モデルの動作が現実の脳神経回路の動作と結びつかないのである。また、心理学的アプローチは、一般にメカニズムを追求する能力に欠けていることもその原因だと言える。

一方、下位のレベルからのアプローチ、すなわち生理学的知見を積み重ねていくという方法もまた、大きな問題を抱えている。

従来の神経生理学の研究戦略は、主に機能局在論に基づくものであった。つまり、脳のある領域、局所的構造(カラムなど)、更には単一のニューロンの役割を追求していき、できるだけ精密な機能地図を構成することを大きな目標とした。これは、認識系や運動系の研究に対してはかなり有効であったし、記憶系に関しても、脳の破壊実験などを通じて、多くの神経心理学的知見をもたらした。

しかしながら、個々のニューロンを扱う神経生理学のレベルと、その集団ないし脳のある領域の機能を問題とする神経心理学のレベルの隔たりは、予想以上に大きい。なぜならば、相互に結合したニューロンの集団は複雑な力学系を成すため、個々のニューロンの振舞いをいくら追いかけても、それだけでは系全体の動作を理解できないからである。

このギャップを無視して、微小電極で測定した単一ニューロンの反応を、その機能と直接結びつけて解釈することには無理があると言わざるを得ない。しかし、

この点に対する認識は決して十分ではなかったように思われる。実際、このことが、記憶系のニューロンに関する、玉石混交のデータの山と解釈の混乱とをもたらしてきたと言っても過言ではない。

今後も生理学が記憶研究の中心であろうという点を考えると、今求められているのは、機能局在論に代わる理論をもとに、神経生理学の新たな研究戦略を立てることだと言えよう。そして、それは個々のニューロンのレベルとその一つ上の機能単位(神経回路網)のレベル間のギャップを埋めることであり、そのための適切なモデルを示すことである。

改めて言うまでもなく、理論と実験の双方が互いに協調しつつ発展することが、記憶のメカニズムを解明していく上で最も大切である。そして、そのためには、それを可能にするモデルとモデル化の手法とが不可欠なのである。

1.1.3 本研究の目的と方法

本研究の目的は、以上のような考えのもとに記憶の神経回路モデルを構成し、理論的に新たな問題を提起するだけでなく、実験に対しても有用な情報を与えることによって、脳の記憶機構の解明への基礎を築くことである。もう少し具体的に言えば、従来のモデルでは説明できなかった実験事実を説明するモデルを、相当の必然性をもって構築すること、また、それによって実験的に検証可能な予言を導き出し、神経生理学に対して多くの有益な示唆を与えることである。

そのためには、まず従来の理論やモデルの問題点を明らかにしなければならない。本論文では、特に神経回路網の力学について詳しく議論する。次節で述べるように、神経回路網の力学に対する理解不足が多くの問題の原因となっており、これを少しでも解消することがモデルを作る上で非常に重要だからである。

それと共に、実際の脳の記憶系において、情報表現や力学がどうなっているかを考える必要がある。本論文では、側頭葉の短期記憶系を対象としてこの問題を論じ、モデルを構成する。

当然のことながら、これだけで上述の目的が完全に達成できるわけではない。しかし、本研究によって、理論と実験とを結ぶ優れたモデルが一つでも示されれば、それだけで十分大きな意義があると言えよう。

1.2 研究の背景

一般に、脳の情報処理原理を理論的に研究しようとする際、どのような観点からアプローチするかが大切である。記憶の研究の場合、次の三つが考えられる。

まず第一に、情報表現という観点である。外界の情報はどのような形で脳内に表現され、記憶されるのであろうか。短期記憶と長期記憶とでその表現は違っているのか、また違うとしたら、どのように変わっていくのだろうか。

このような問題に答えるためには、どのような表現法があり、それぞれがどのような長所・短所をもつかを明らかにしていかなければならない。しかし、それには、神経回路網の上に表現された情報が、何を目的としてどのように処理されるかを考える必要がある。

これもまた難問であるが、この研究では、とりあえず工学的な連想記憶という枠組みを設定することでこの問題を避けている。つまり、神経回路網によって連想記憶モデルを構成し、その記憶容量や想起能力がどうなるかによって、そこで用いた表現の評価をしようというわけである。

しかし、どのようにして情報を処理するかというのは、情報表現とはまた別の問題である。多くの工学的モデル、特に連想記憶モデルは、神経回路網によって作られる力学系を利用して情報処理を行うが、どのような力学系を用いるべきか、またそれがどういう構造をしているのかといった問題は、情報表現を考えるだけでは答えられない。

そこで、神経回路網の力学という第二の観点が重要となる。上述の問題に加えて、実際の脳神経回路がどのような力学系になっているかを明らかにできれば、前節で述べた単一ニューロンと神経回路網のレベル間のギャップをかなり埋められることになる。

第三は、神経回路網の自己組織化、すなわち、ある機能を果たす神経回路がどのようにしてできあがるのか、またどのような学習を行うことによって外界に適應するのか、という観点である。この方向のアプローチは、認識系の研究において盛んである。

しかしながら、それは情報表現や力学がある程度明らかになってから話である。どのような情報表現が取られていて、どのようなダイナミクスに従って動作しているのか全くわからない系の自己組織化を考えても、あまり意味があるとは思えない。そして、記憶系に関しては、それらの点はまだ明らかではないのである。

以上の議論からわかるように、力学という観点から神経回路網を眺め、それを理解することは、様々な意味で重要である。そこで、次に神経回路網の力学について、理論と実験の両面から問題点を論じることにする。このうち本研究と特に関係の深いものは、次章以降の関連する部分で詳しく述べる。

1.2.1 神経回路網の力学に関する問題点

相互結合のある神経回路網は、非線形・多変数の力学系を構成する。このような系は、一般に複雑で、数学的に最も扱いにくい対象の一つである。そのため、神経回路網の力学に関しては、まだ不明な点が非常に多い。

もちろん、理論的な解析はこれまでも行われてきたし、それによって明らかになった点も少なくない。特に、Hopfield^[5]がある種の神経回路モデルにエネルギー関数が定義できることを示して以来、スピングラス系に対する解析手法を用いて神経回路網を解析することが盛んである^[6]。

しかしながら、こうした手法が使える場合というのは限られている。逆に、従来の理論的研究は、扱いやすいモデルのみを対象としてきた。特に物理学者にその傾向が強く、スピングラス系からの類推を拠り所としているため、それが可能な枠内で、次々と細かな問題を解いているだけのように見える。しかも、そうした比較的単純な力学構造をもつモデルの場合ですら、その性質が十分に理解されたとは言いがたいのである。

一方で、神経回路網が作る力学系は、数学的にはよくわからないまま、工学的に利用されてきた。例えば、Hopfieldの連想記憶モデル^[5]では、記憶すべきパターンを力学系のアトラクタとすることによって連想記憶機能を実現している。

しかし、よく考えてみると、これまで用いられてきた力学系が、連想記憶その他の情報処理を行うのに最も適切なものだという保証はどこにもない、ということに気が付く。そもそもMcCullochとPitts^[7]が細胞のモデルを提案して以来、現在に至るまで、系のダイナミクスは基本的に変わっていないのである。

確かに、並列情報処理の原理や神経回路網の力学を研究する際には、現実の脳にはあまりこだわらず、単純化・理想化したモデルを扱った方がよい。その方が複雑な要因を持ち込まずに済むからである。しかし、それによって、本質的な部分を損なうことがあってはならない。

しかしながら、この点に関する検討は、これまでほとんどなされていなかったように思われる。実際、第3章で示すように、想起のダイナミクスをわざわざ変

えるだけで、従来の連想記憶モデルの能力が飛躍的に向上する。つまり、これまで研究され、用いられてきた力学は、必ずしも工学的に優れたものではないのである。

1.2.2 生理学的知見

では、従来の力学系は、脳のモデルとして適当なのであろうか。

この問題に答えるためには、脳の神経回路網の力学的性質について、なんらかの情報が必要である。しかし、これを調べることを目的とした実験は存在しない。神経生理学はそのための手段をまだもたないし、なにより力学という観点がないからである。

しかしながら、最近になって極めて注目すべき知見が、海馬と並んで最も重要な記憶の座である側頭連合野で得られている(4.1.1参照)。この知見は、ある種の短期記憶情報がどのような形で表現されているかを示すと同時に、間接的に短期記憶を保持する神経回路網の力学的構造を示唆している。

ところが、この知見には、従来の連想記憶モデルではうまく説明できない部分がある。もう少し強く言うならば、側頭葉の短期記憶を保持する神経回路の力学系は、これまで考えられてきたような神経回路モデルでは実現することができない。つまり、従来の力学系は、脳のモデルという観点からも本質的な改善の余地があるのである。

以上の点から考えると、脳の神経回路には、今までのモデルには欠けていた何らかの重要な原理や機構が存在するに違いない。その原理は何であり、それをどうモデル化すればよいのだろうか。これは、本研究の主題の一つである。

1.3 論文の構成

本論文の第2章以下の構成は、次の通りである。

第2章「連想記憶の力学」では、連想記憶モデルの中で最も基本的なものについて、その力学的性質を論じる。ここでの議論は、本研究全体の基礎となる。

この章では、まず従来のモデルについて、これまで理論的に知られている性質を簡単にまとめる。次いで数値実験による検討から、連想記憶の想起過程における問題点を明らかにし、その原因を推察する。また、モデルの改良法について考察する。

第3章「ダイナミクスを改良した神経回路網」では、いったん現実の脳から離れ、連想神経回路網の新しいダイナミクスを考える。これは、モデルの連想記憶能力を向上させるためであると同時に、神経回路網の力学に対する理解を深めるためでもある。

この章では、まず離散型の連想記憶モデルについて、部分反転法というダイナミクスの改良法とその効果を示す。次に、非単調な入出力特性をもつ細胞からなるモデルを提案し、その性質を調べる。更に、後者のモデルに関して、記憶情報をもつ構造と想起能力の関係や学習の問題を論じる。

第4章「側頭葉短期記憶系の力学とそのモデル」では、第3章までの結果をもとに、実際の脳の神経回路網の力学を考え、そのモデルを構成する。

この章では、まず側頭葉TE野に見られる短期記憶ニューロン群の性質について説明し、神経回路網の力学の観点から検討を加える。次いで、従来とは異なる力学系を実現する神経回路モデルを構成し、このモデルによりTE野のニューロンの振舞いがうまく説明されることを示す。最後に、このモデルの興味深い動作に関して、いくつかの考察を行う。

第5章は結論である。本研究の結果をまとめると共に、残された課題や今後の展望について述べる。

連想記憶の力学

自己相関型の連想記憶モデル(以下, 自己相関モデルと呼ぶ)は, 神経回路による連想記憶モデルの中で最も基本的なものである. その研究の歴史は長く, 1969年に提案^[8]されて以来, 数々の解析がなされてきた^[9,10]. 特に最近では, スピングラス系との類似性などから, 記憶容量や平衡点の分布といった性質に関する研究が盛んに行われている^[11~13].

しかしながら, 連想神経回路網の力学的性質を明らかにするという点から見たとき, これらの研究は決して十分なものではない. むしろ, 連想記憶のダイナミクスに対する我々の理解はまだまだ不足している. そして, このことが, 従来モデルがもつ本質的な問題点を長い間解決できなかった原因だと言える.

本章では, この基本的な自己相関モデルを主な対象として, 連想記憶の力学について論じる. また, これを通してより一般の神経回路網のダイナミクスに対する理解を深めることも, 本章の大きな目的の一つである.

2.1 連想記憶モデル

2.1.1 自己相関モデル

n 個の細胞からなる神経回路網を考え、 j 番目の細胞から i 番目の細胞への結合荷重を w_{ij} 、 i 番目の細胞の出力を x_i とする。 x_i が ± 1 の二値をとるものとすれば、ある時刻 t での回路網の状態は、 n 次元のベクトル

$$X_t = (x_1(t), \dots, x_n(t))^T$$

で表される (上添字 T は転置を意味する)。

この回路網に、 m 個のパターン S^1, S^2, \dots, S^m を記憶させるとしよう。但し、 $S^\mu = (s_1^\mu, \dots, s_n^\mu)^T$ は、系のとり得る 2^n 個の状態の中から独立かつランダムに選ばれ、近似的に

$$\frac{1}{n} \sum_{i=1}^n s_i^\mu s_i^\nu = 0 \quad (\mu \neq \nu) \quad (2.1)$$

が成り立っているものとする。

このとき、自己相関モデルでは、結合荷重を

$$w_{ij} = \frac{1}{n} \sum_{\mu=1}^m s_i^\mu s_j^\mu \quad (2.2)$$

とすることによって、これら m 個のパターンを記録する。言い換えれば、荷重行列 $W = [w_{ij}]$ として、自己相関行列

$$W = \frac{1}{n} \sum_{\mu=1}^m S^\mu (S^\mu)^T = \frac{1}{n} \Sigma \Sigma^T \quad (2.3)$$

を用いる。ここで Σ は $n \times m$ の行列 (S^1, \dots, S^m) である。

この荷重行列は、相関学習^[10]を行うことによっても容易に構成することができる (3.4.1 参照) が、これはこのモデルの大きな特長である。

なお、このモデルの場合、 W の対角成分 w_{ii} (細胞の自己結合に相当する) は常に m/n であり、記憶パターンの数以外の情報を全く含んでいない。そこで、便宜上 $w_{ii} = 0$ とすることがしばしばある。後述のように、こうすると系の安定度は低下するが、力学的構造を調べるのに都合が良い。

2.1.2 想起ダイナミクス

想起の際には、回路網に適当な初期状態 X_0 を想起入力として与えた後、系の状態をあるダイナミクスに従って更新していく。この想起ダイナミクスにはいくつかの種類があるが、ここではすべての細胞が

$$x_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^n w_{ij}x_j(t)\right) \quad (2.4)$$

に従って同期的に動作する場合、すなわち、系の状態が離散時刻 $t = 0, 1, 2, \dots$ において

$$X_{t+1} = \operatorname{sgn}(WX_t) \quad (2.5)$$

に従って遷移する場合を扱う。但し、 $\operatorname{sgn}(u)$ は $u > 0$ のとき 1 、 $u \leq 0$ のとき -1 をとる関数(ベクトルに対しては成分ごとくに作用する)である。

このようなダイナミクスを用いると、状態が変化するたびにエネルギー

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij}x_i x_j \quad (2.6)$$

が減少していく^[5]。 E の値は有限であるから、十分に時間が経ったとき、系の状態 X は初期状態に応じて定まるある平衡点 $\bar{X}(X_0)$ に到達することになる[†]。

式(2.3)を用いると、

$$\begin{aligned} E &= -\frac{1}{2} X^T W X \\ &= -\frac{1}{2n} X^T \sum_{\mu=1}^m S^\mu (S^\mu)^T X \\ &= -\frac{1}{2n} \sum_{\mu=1}^m (X^T S^\mu)^2 \end{aligned} \quad (2.7)$$

のように変形されるから、状態 S^μ はその周辺の状態よりもエネルギーが低いと考えられる。従って、 X_0 がある記憶パターン S に十分「近い」(2.2.2 参照)ならば、

$$\bar{X}(X_0) = S \quad (2.8)$$

[†] 厳密には、これは回路が非同期的に動作するときの話で、同期的に動作する場合には成り立たない(周期2で振動する)ことがある。但し、このことは以下の議論にはほとんど関係しない。

となること、すなわち S が正しく想起されることが期待される。

なお、式(2.8)が成り立つような初期状態 X_0 の集合を、 S の引込み領域 (basin of attraction) と呼び、 $B(S)$ で表す。また一般に、ある平衡状態 \bar{X} の引込み領域 $B(\bar{X})$ が \bar{X} 以外の状態を含むとき、 \bar{X} は力学系のアトラクタであると言う。

2.1.3 その他のモデル

以下では、次の2種類の連想記憶モデルにも触れる。

一つはいわゆる一般逆行列を用いた連想記憶モデル^[14] (以下、一般逆行列モデルと呼ぶ)であり、自己相関行列の代わりに

$$W = \Sigma(\Sigma^T \Sigma)^{-1} \Sigma^T \quad (2.9)$$

で与えられる荷重行列を用いる点が自己相関モデルと異なる。このような W を用いると、 S^μ ($\mu = 1, 2, \dots, m$) はすべてエネルギー最小状態となる^[15]。

もう一つは、層状の神経回路を用いたモデル^[16,17] (以下、層状モデルと呼ぶ)である。そのダイナミクスは

$$X_{t+1} = \text{sgn}(W_t X_t) \quad (2.10)$$

$$W_t = \frac{1}{n} \sum_{\mu=1}^m S_{t+1}^\mu (S_t^\mu)^T \quad (2.11)$$

で表される。 S_t^μ がすべて独立であり、時刻ごとに別々の荷重行列 W_t を用いる点が、上述のモデルと異なっている。

2.2 力学的性質

連想記憶モデルの力学的性質に関しては、既にある程度的事実が知られており、理論的な解析もなされている。本節では、その中で特に重要と思われる事柄について述べる。なお、以下の議論において、 S は注目している特定の記憶パターンを表す。 S としてどのパターンを選んでも一般性を失わないので、特に断りのない限り、 $S = S^1$ であるとする。

2.2.1 記憶容量

n 個の細胞からなる連想記憶モデルに m 個のパターンを記憶させたとき、想起がうまくいく最大の m をそのモデルの記憶容量と呼び、 $C(n)$ で表す。但し、 n は十分大きいものとする。

ここで「想起がうまくいく」ことの定義が問題となるが、とりあえず「 S が平衡点であり、 S に十分近い X_0 を選ぶと、系の状態 X は S に収束する (S が力学系のアトラクタである)」ことだとして、このときの自己相関モデルの記憶容量は理論的に導き出すことができ、

$$C(n) = \frac{n}{2 \log n - \log \log n} \quad (2.12)$$

であることが知られている^[12,18]。

式(2.12)からわかるように、細胞数に対する記憶パターンの割合(記憶率)

$$r = \frac{m}{n}$$

を一定にしたまま $n \rightarrow \infty$ にすると、 $C(n)/n \rightarrow 0$ となる。これは、 S が平衡点である確率が0になることを意味する。

それにもかかわらず上述のモデルが連想記憶モデルとして機能し得るのは、 r がある値 r_c 以下であれば、適当な X_0 をとったとき、 $\bar{X}(X_0)$ が S に非常に近いものになるからである^[12]。但し、 r が r_c より大きいと、 X が S の近傍の平衡点に達することはない(2.2.3 参照)。従って、この r_c (最大記憶率と呼ぶ)の値が実質的な記憶容量を表していると言える。

自己相関モデルの場合、 $r_c \simeq 0.15$ であることが実験的に知られていた^[5]が、なぜそうなるのかは不明であった。これに対して、Amitら^[11]は、物理学における

スピングラス系の解析手法を用い、最大記憶率が約0.14になることを理論的に示した。但し、この解析は、レプリカ法と呼ばれる方法が適用できることを前提としているため、必ずしも厳密ではない。

一方、AmariとMaginu^[18]は、神経統計力学の手法を用いて想起過程を解析し、最大記憶率が約0.16になることを示した。しかしこの解析も、いくつかの近似を行っているため、厳密なものではない。

なお、層状モデルの最大記憶率は厳密に求めることができ、 $r_c \approx 0.27$ であることがわかっている^[16]。また一般逆行列モデルでは、すべての記憶パターンが線形独立であればよいから、一般には最大記憶率は1である。但し、式(2.9)の W をそのまま用いると、 $r > 0.5$ のとき連想記憶モデルとして機能しない(図3.14参照)。これは、 w_{ii} の値が大きすぎるために、ほとんどすべての状態が平衡点となってしまうからである(この場合は $w_{ii} = 0$ とした方がよい)。

2.2.2 平衡点の分布

前項でも触れたように、平衡点が S の周辺でどのように分布しているかという問題は、力学系の構造を知る上でかなり重要である。これについては理論的解析が進んでおり、自己相関モデルの場合、次の結果が得られている^[13]。

まず、二つの状態 X^1 と X^2 の間の距離 d を正規化したハミング距離

$$d(X^1, X^2) = \frac{1}{2n} \sum_{i=1}^n |x_i^1 - x_i^2| \quad (2.13)$$

で定義しよう。また、ある状態 X の S からの距離を $\rho = d(X, S)$ で表すことにする。ほとんどすべての状態が、 $\rho \approx 0.5$ のところに集中している点に注意が必要である。

平衡点は $\rho = 0.5$ のところに最も多く分布するが、 ρ が小さくなるにつれてその数が急速に減少し、あるところで0になる。しかし、記憶率 r がある値 $r_c (\approx 0.11)$ 以下のとき、更に ρ が小さいところに多数の平衡点が存在する領域がある。すなわち、平衡点の存在域は、 S と無相関($\rho = 0.5$)の状態を中心とする領域と、 S の近傍($\rho < 0.005$)の領域の二つに分かれている。

この S 近傍の領域は、 r が大きくなるにつれて S から遠ざかり、 $r > r_c$ になるともう一つの領域とつながってしまう。また、一般に r が大きいときほど、 S から最も近くにある平衡点までの距離が大きくなる。

なお、ここでは単に平衡点と呼んだが、この解析は正確にはエネルギーの極小点の分布を扱っている。荷重行列の対角成分 w_{ii} が0のとき両者は一致するが、 $w_{ii} > 0$ のときには、エネルギーの勾配が十分小さい状態はすべて平衡点となる。式(2.3)に従って W を構成した場合、 $w_{ii} = r$ であるから、 r が大きくなると $w_{ii} = 0$ とした場合との違いが現れる。しかし、 r が小さければ、対角成分は平衡点の分布や想起過程にほとんど影響を及ぼさない。

2.2.3 想起過程

平衡点の分布は基本的な力学的性質であるが、 r_c の値が r_c と一致しないことからわかるように、これだけでは想起を理解するのに十分とは言えない。また、 $r > r_c$ の場合にも、 S のある程度近く (例えば $r = 0.2$ のときで $\rho < 0.03$) には平衡点が存在するから、想起がうまくいかないのは S の周りにアトラクタがないためではない。

このように、想起のダイナミクスは、最も近い平衡点に向かって状態が遷移していくといった単純なものではない。従って、それぞれの平衡点がどのような引込み領域をもつかを知る必要があるが、これは解析的にはもちろん、数値実験でも求めるのが困難である。そこで、ある初期状態 X_0 を選んだときに、 X が時間と共にどのように遷移していくかを見ることにする。

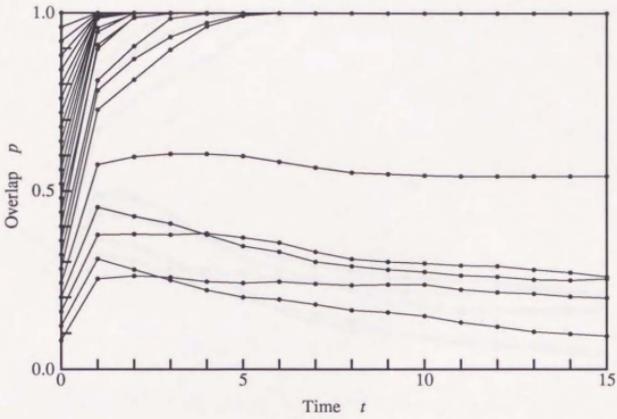
まず、 S からの距離 ρ の代わりに、 X と S との相関 p を考え、これを重なり率 (overlap) と呼ぶ。すなわち、

$$p = \text{Cor}(X, S) = \frac{1}{n} \sum_{i=1}^n x_i s_i \quad (2.14)$$

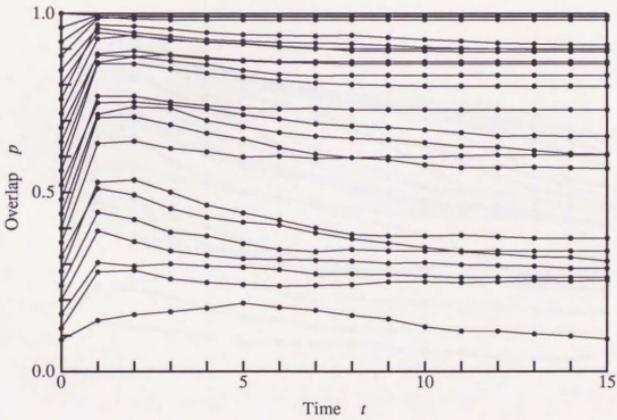
である。また、時刻 t での重なり率を p_t とする。 $p = 1 - 2\rho$ であるから、 p_t が大きいほど X_t が S に近いことを意味する。この重なり率 p の時間変化について、次のような興味深い現象が報告されている^[18]。

図2.1は、初期状態 X_0 と S との距離を変えつつ[†]、 p_t をプロットしていったグラフである。(a)は $r = 0.08$ 、(b)は $r = 0.2$ の場合で、いずれも $n = 1000$ で実験した。また、図2.2は、 W の対角成分 w_{ii} を0にして同じ実験を行った結果である。この実験その他から、以下のことが言える。

[†] 正確には、 S からある距離 ρ だけ離れた状態の中から一つをランダムに選んで X_0 とする。

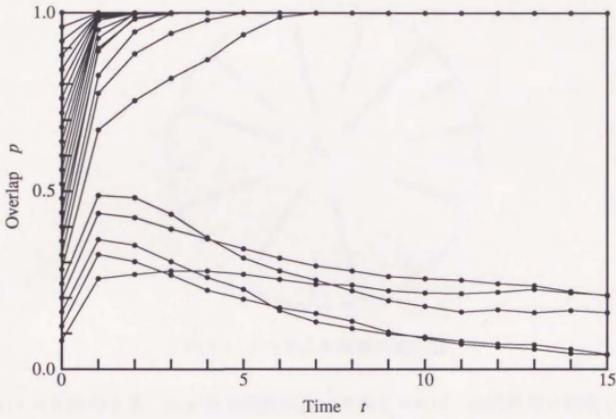


(a) $r = 0.08$

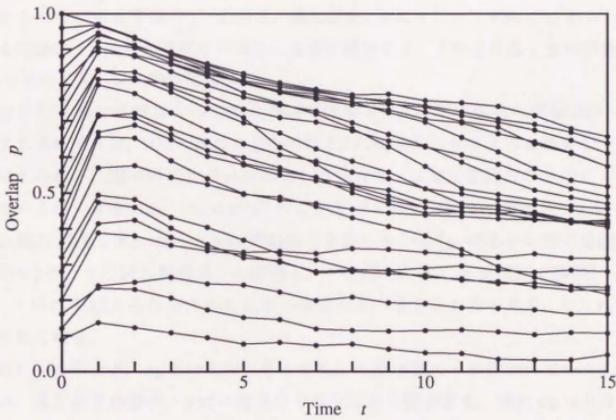


(b) $r = 0.2$

図 2.1 重なり率 p の時間変化 ($w_{ii} = r$)

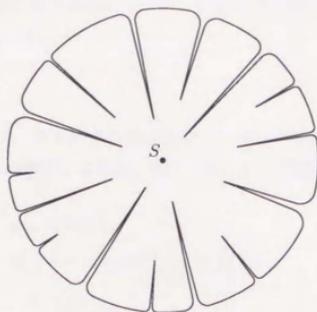


(a) $r = 0.08$



(b) $r = 0.2$

図 2.2 重なり率 p の時間変化 ($w_{ii} = 0$)

図2.3 S の引込み領域の模式図

1) $r = 0.08$ のとき, p_0 がある臨界値 p_c より大きければ, 比較的短い時間で p がほぼ 1 になる. すなわち, X は S の極めて近傍にある平衡点の一つ \hat{S} に達する. しかし, $p_0 < p_c$ のときには, p はある程度までしか大きくなり, X は S から離れたところにある平衡点(いわゆる「偽の記憶」)にトラップされてしまう. また多くの場合, p はある程度まで増加した後減少する. このようなしきい現象は, $w_{ii} = 0$ のとき, より明瞭になる.

2) S の近傍の平衡点 \hat{S} の引込み領域全体をもって, S の引込み領域 $B(S)$ とみなすことにすれば, $B(S)$ はある意味で図2.3に模式的に示すような形をしていると考えられる. 図で内側に切れ込んでいる部分は, n 次元空間のうちのごく一部を占めるにすぎないが, $X_0 \notin B(S)$ のとき X はこの切れ込みの奥へ入っていく.

3) 臨界重なり率 p_c はモデルの想起能力を表している (p_c が小さいほど想起能力が高い)が, その値は記憶率 r に依存し, r と共に大きくなる(図3.5参照). しかし, r が r_c を超えるあたりからしきい現象がはっきりとしなくなり, $r > r_c$ で完全になる.

4) $r > r_c$ のとき, p_0 がいくら大きくても, X が単調に S に近づいていくことはない. ほとんどの場合, p は一度大きくなってから減少する. 特に $w_{ii} = 0$ のときには, たとえ $X_0 = S$ であっても, X は S から遠く離れていってしまう.

なお, 層状モデルでもほとんど同じ現象が見られる^[17]が, 最大記憶率(約0.27)は自己相関モデルの場合に比べてかなり大きい. また, $p_0 < p_c$ ないし $r > r_c$ のと

き、 p が $t > 1$ で指数関数的に減少していき、最終的にはほぼ0になる点も自己相関モデルと多少異なっている。

2.3 考察

2.2.1で触れたように、以上のような性質は Amari と Maginu^[18]によって理論的に解析されている。この解析によれば、重なり率 p の時間変化は

$$p_{t+1} = \Phi(\bar{p}_t) \quad (2.15)$$

$$\sigma_{t+1}^2 = r + 4\{\varphi(\bar{p}_t)\}^2 + 4r\bar{p}_t\varphi(\bar{p}_t)p_{t+1} \quad (2.16)$$

$$\bar{p}_t = p_t/\sigma_t \quad (2.17)$$

で与えられる。但し、 $\sigma_0^2 = r$ であり、

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] \quad (2.18)$$

$$\Phi(s) = \int_{-s}^s \varphi(u) du \quad (2.19)$$

である。

これらの式は、 $t=0$ のときには厳密に成り立つが、 $t \geq 1$ では近似的に成り立つにすぎない。実際、数値実験の結果と比べると、若干の違いが見られる。また、複雑な計算によって導き出されているため、想起過程を理解するのに必ずしも有用ではない。そこで、ここでは想起の過程を少し別の観点から眺め、直観的な理解を試みる。

2.3.1 想起過程の直観的解釈

まず、 $S(=S^1)$ 以外の記憶パターン S^ν ($\nu = 2, 3, \dots, m$)が想起の際にどう影響するかを調べるために、 S^ν と X の相関の絶対値

$$q^\nu = \frac{1}{n} \left| \sum_{i=1}^n x_i s_i^\nu \right| \quad (2.20)$$

を考える。

図2.4は、 q^ν の時間変化を p (太線)と共に示したものである。(a)は正しく想起できた場合で、 q^ν は $t=1$ で大きくなった後は減少する傾向にある。一方、(b)は

正しい想起ができなかった場合であり、 q^ν は $t > 1$ でも減少せず、むしろ全体的に増加する傾向にあることがわかる。

次に、このような現象を理解するため、 q^ν の二乗和

$$\sigma^2 = \sum_{\nu=2}^m (q^\nu)^2 \quad (2.21)$$

を考える[†]。

X_0 をランダムに選んだとき、 q_0^ν ($\nu = 2, 3, \dots, m$) はそれぞれ平均 0、分散 $1/n$ の独立な正規分布に従うから、 $\sigma_0^2 \simeq r$ であることがわかる (q および σ^2 の下添字は時刻を表す)。同様に、 X が S に十分近いときにも $\sigma^2 \simeq r$ が成り立つ。更に、簡単な計算により $q_1^\nu > q_0^\nu$ が示される (式 (2.15)~(2.17) で $t = 0$ のときと同じ) から、 $\sigma_1^2 > \sigma_0^2$ が成立する。従って、正しい想起ができるためには、 σ^2 は $t > 2$ で小さくなる必要がある。

ところが、式 (2.7) より

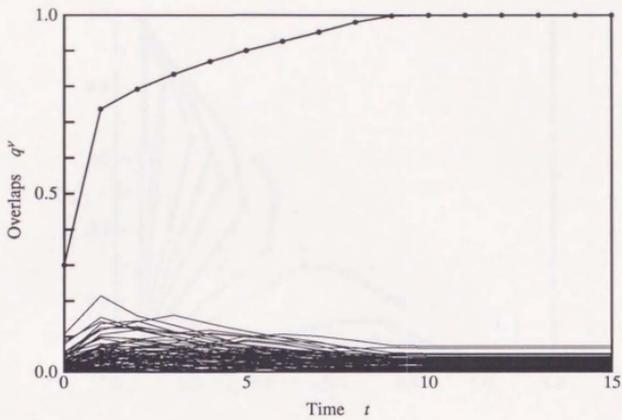
$$\begin{aligned} p^2 + \sigma^2 &= \frac{1}{n^2} \sum_{\mu=1}^m (X^T S^\mu)^2 \\ &= -\frac{2}{n} E \end{aligned} \quad (2.22)$$

であるから、 σ^2 が減少する方向は、一般にはエネルギーの増える方向となる。従って、 p^2 だけが大きく増加する領域——そのような領域は S の近傍に限られる——に X がある場合を除き、 σ^2 は状態遷移のたびに増大する。従って、いったん σ^2 が大きくなりすぎてしまうと、想起は非常に困難になると考えられる。

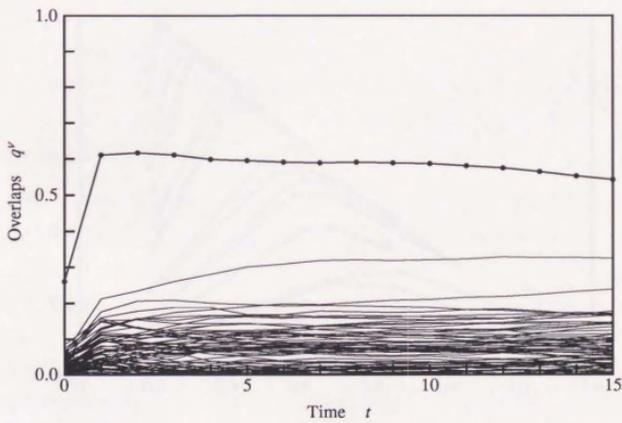
実際に p^2 と σ^2 を状態変数として、 X の遷移の様子を描いたのが図 2.5 である。(a) は $r = 0.08$ の場合で、前述のしきい現象がはっきりと現れている。この図からわかるように、 σ^2 の値がある限度を超えなければ、 $\sigma^2 \rightarrow r$ となって想起がうまくいく。逆に、 σ^2 があまり大きくなってしまうと、言い換えれば X の S への近づき方に比べてエネルギーの減り方が大きすぎると、正しい想起ができない。また (b) は $r = 0.2$ の場合であるが、 σ^2 は増加する一方で減少しないことがわかる。

以上の結果をまとめるならば、回路網の状態 X は、 X_0 から S へ「まっすぐ」に向かうのではなく、他の記憶パターン S^ν に引っ張られて「斜め」に近づいていく

[†] この σ^2 は、式 (2.16) の σ_0^2 に相当する量になっている。

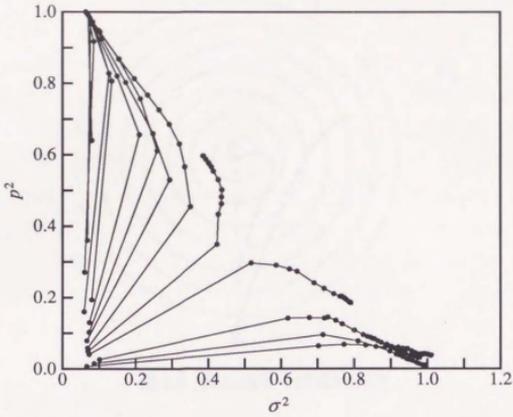


(a) $p_0 = 0.3$

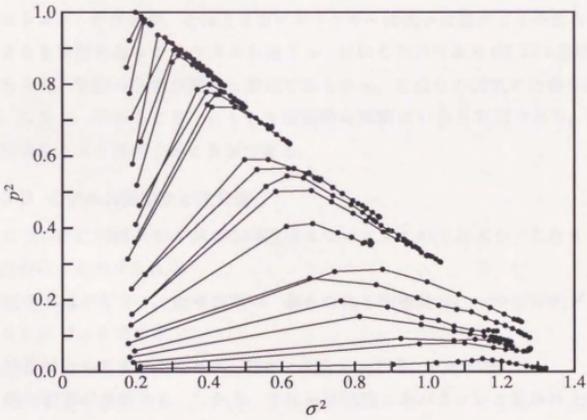


(b) $p_0 = 0.26$

図 2.4 q^ν ($\nu = 2, 3, \dots, m$) の推移 ($r = 0.08$)



(a) $r = 0.08$



(b) $r = 0.2$

図 2.5 状態遷移に伴う σ^2 と p^2 の変化

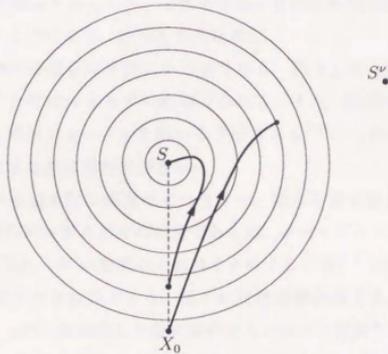


図2.6 想起過程の模式的表現

と行うことができよう(図2.6). また, $r > r_c$ のとき, 状態 S はいわばエネルギー関数の鞍点になっていると考えられる. つまり, S は周囲のほとんどすべての状態よりエネルギーが低いが, それより更にエネルギーの低い状態がごく少数存在し, そのような状態を通して X が S から遠ざかっていくわけである(図3.13参照).

もちろん, 実際の状態空間は n 次元であるから, これらの図式には多少無理がある. しかし, 今のところ, こうした直観的な解釈はかなり有用であり, 種々の実験結果をうまく説明することができる.

2.3.2 モデルの問題点と改良法

ここで, 自己相関モデルがもつ問題点についてまとめてみよう. 主なものとして, 次の四つが挙げられる.

- 1) 記憶容量が小さく, 効率が悪い. 蓄えられる情報は, 一つのシナプスにつき最大0.15ビットである.
- 2) 想起能力があまり高くない. 特に r が大きいとき, それが言える.
- 3) 偽の記憶が多数ある. しかも, それらは記憶したパターンと見かけ上区別がつかない.
- 4) 記憶パターンが相互にほぼ直交している(式(2.1)が成り立つ)必要がある. これは, 情報表現上の大きな制約である.

これらの問題を解決するために、これまで様々なモデルの改良法が提案されてきた。これを大きく分けると、次のようになる。

第一に、回路網の構造自体を変える方法である。例として、前述の層状モデルのほか、BAM^[19]と呼ばれるモデルが挙げられる。また、現実の神経回路とはかけ離れているが、高次ニューロンを用いる方法もある^[20]。しかし、これらの方法は、構造を複雑にする割に効果が少ない。

第二に、モデルの基本的な構造はそのまま、結合荷重を修正するやり方がある。前述の一般逆行列モデルがその典型であるが、パーセプトロン流の誤り訂正学習^[10]を用いる方法や、 W の要素の大部分を0にする方法^[21]も考えられている。

このように荷重行列を改良すると、確かに記憶効率が上がり、想起能力も改善される。しかし、 p が一度増加した後に減少するといった現象はやはり生じ、偽の記憶を想起してしまうことも同様である。また、4)の問題は回避できるものの、積極的に記憶パターンがもつ構造を利用することはできなくなる(3.3.2参照)。

結局、以上のどの方法を用いても、上述の問題点をすべて解決することはできない。特に3)の問題は、エネルギーが時間と共に減少するような力学系を利用する限り避けられない。

そこで、第三の方法として、想起のダイナミクスを変えることが考えられる。この方法は、これまでほとんど取られたことがないが、かなり有望である。というのは、前項で述べたような想起過程の特性は、従来のダイナミクスが荷重行列 W に含まれる情報を十分に利用していないことを示唆しているからである。

例えば、自己相関モデルの記憶容量が層状モデルの半分程度しかないことは、 r が0.15を超えても、 W から記憶情報が失われるわけではないということを示している。また、数値実験によれば、記憶したパターンと偽の記憶とでは、想起時の各細胞への入力値の分布が異なるが、このことから、 W は両者を区別するだけの情報を含んでいることが示唆される。つまり、上述の問題点のほとんどが、 W の選び方ではなく、想起のダイナミクスに問題があるために生じると考えられるのである。

そのほか、モデルの改良法には、記憶パターンの表現ないしコーディングを変える方法もある。例えば、記憶するパターンをスパースな(1をとる要素の割合が非常に小さい)ものに限ると、記憶容量が増加する^[22]。但し、ただスパースなパ

ターンを記憶しただけでは駄目で、想起の際にその情報を利用しなければ想起能力は改善されない。つまり、コーディングを変えることは、想起のダイナミクスを変えることと表裏一体なのである。

以上の議論から、従来のモデルの問題点を本質的に解決するには、単に荷重行列や学習法を改善するだけでなく、想起のダイナミクスを改良する必要があると考えられる。では、どのように改良すればよいのだろうか。この点に関しては、次章で具体的に論じることとする。

第3章

ダイナミクスを改良した神経回路網

前章で述べたように、想起のダイナミクスを改良することによって、連想記憶モデルの能力を高めることが可能だと考えられる。また、ダイナミクスを改良することは、神経回路網が構成する力学系の性質や構造を知るための手段としても有力である。そこで、本章では、主に自己相関モデルについて、ダイナミクスの改良を試み、その効果を検討する^[23]。

ここで提案するのは、離散型のモデルである特定の細胞の出力を強制的に反転させるという方法(部分反転法)と、連続型のモデルで細胞の入出力特性を非単調なものにする方法(非単調ダイナミクス法)である。これらの方法を用いると、従来のモデルの記憶効率や想起能力が大幅に向上する。また、特に非単調ダイナミクスを適用したモデルは、数々の興味深い性質を示す。

ここで扱うモデルは、神経回路網の力学だけでなく、記憶の情報表現や学習の問題を考えるためにも重要である。また、現実の脳を直接モデル化したものではないが、次章で論じるように、脳の神経回路とも深くかかっている。

3.1 部分反転法

本節では、2.1で述べた離散時間一離散情報型の自己相関連想記憶モデルについて、想起のダイナミクスを改良する方法を考える。

但し、ダイナミクスを改良するといっても無制限に変えてよいわけではなく、この種の連想記憶の長所は残しておくべきだという点に注意する必要がある。例えば、各細胞(素子)は独立かつ並列的に動作すべきであるし、一様性も保たれた方がよい。また、想起入力 X_0 以外に、 p_0 の値など特別な情報を与えるべきではない。

つまり、ここで言うダイナミクスの改良とは、従来のダイナミクスでは利用されていなかった情報を利用することである。そして、そのためには、各細胞への入力の符号だけでなく、その大きさに関する情報を系の状態変化に反映させることが必要だと考えられる。

このような条件を満たすダイナミクスの改良法の一つが、以下に述べる「部分反転法」である。

3.1.1 基本的考え方

回路網の状態がエネルギー極小値にトラップされるのを防ぐ方法としてよく用いられるのが、ノイズ(熱的ゆらぎ)を加えるという方法である。もし S の引き込み領域が本当に図2.3のような形をしているのならば、適当なノイズを加えることによって、 X を引き込み領域の内部に飛び込ませることが可能なように思われる。しかし、ただランダムなノイズを加えてもうまくいかないことは、図2.5や図2.6から明らかであろう。

しかしながら、同時にこれらの図は、ランダムではなく σ^2 を減らす方向に X を動かしてやれば、想起の過程が改良されるということを示唆している。つまり、状態が遷移するたび、 S からの距離はあまり変わらないようにしつつ、その他のパターン S^p との距離を遠ざける(相関を減らす)方向に戻す、ということを繰り返せばよい(図3.1)。

これが部分反転法の基本的な考え方である。 X をある方向に動かすとは、ある細胞の出力 x_i の符号を変える(反転させる)ことにほかならないから、どのような x_i を反転させるかが大切であるが、 $x_i = s_i$ が成り立っている可能性が低く、反転

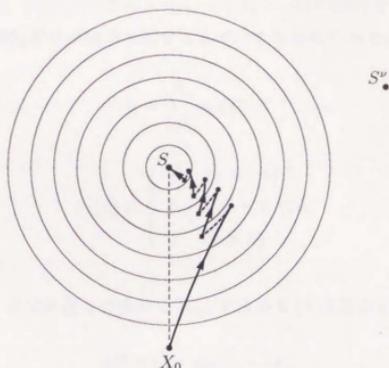


図3.1 部分反転法の考え方

によって全体のエネルギーがなるべく大きくなる ($p^2 + \sigma^2 = -2E/n$ に注意) ような細胞を選ぶべきだと考えられる。

その具体的アルゴリズムは次に示す通りである。これが唯一の方法ではないし、最適であるという保証もないが、単純な割に大きな効果が得られる。

3.1.2 アルゴリズム

まず、想起の1ステップをフェーズIとフェーズIIとに分ける。前者は従来と同様の想起を行うフェーズで、後者が出力の一部を反転するフェーズである。

具体的には、フェーズIで、入力荷重和

$$u_i = \sum_{j=1}^n w_{ij} x_j(t) \quad (3.1)$$

を求める。式の上では必要ないが、ここでいったん

$$x_i = \text{sgn}(u_i) \quad (3.2)$$

によって X を更新することにしよう。そうすると、ここまでのダイナミクスは従来と同じである。

フェーズIIでは、式(3.1)で求めた u_j ($j = 1, 2, \dots, n$) の絶対値が、ある定数 h より大きいような細胞だけが出力を出すとして、それらについての荷重和 v_i を求める。式で表すと

$$v_i = \sum_{j=1}^n w_{ij} \phi_h(u_j) \quad (3.3)$$

となる。但し、

$$\phi_h(u) = \begin{cases} -1 & (u < -h) \\ 0 & (-h \leq u \leq h) \\ 1 & (u > h) \end{cases} \quad (3.4)$$

である。

そうした上で、次式に従って改めて X_{t+1} を求める (λ は正の定数)。

$$x_i(t+1) = \text{sgn}(u_i - \lambda v_i) \quad (3.5)$$

これにより、フェーズIで求めた x_i のうち、

$$0 < \frac{u_i}{v_i} < \lambda \quad (3.6)$$

が成り立つものが反転することになる。式(3.1)~(3.5)をまとめると、簡潔に

$$X_{t+1} = \text{sgn}(W(X_t - \lambda \phi_h(WX_t))) \quad (3.7)$$

と表すこともできる (関数 ϕ_h はベクトルの成分ごとに作用する)。

以下、フェーズIとIIの操作を交互に行って X を順次更新していき、平衡状態に達するまで繰り返す。但し、必ず平衡状態に落ち着くという保証はなく、リミットサイクルに達する場合もある。

このアルゴリズムの意味を直観的に言うならば、「非常に大きな入力(の絶対値)を受けた細胞の出力を信用してはいけない」ということになる。これは、従来のダイナミクスを用いたとき、式(2.6)のエネルギー関数 E が

$$E = -\frac{1}{2} \sum_{i=1}^n |u_i| \quad (3.8)$$

と書けることからわかるように、 $|u|$ の大きな細胞ほど、 σ^2 の増加への寄与が大きいからである。

逆に、一部の細胞の $|u|$ が大きくなりすぎることを防げば、 σ^2 が増大しないようにできると考えられる。但し、 $|u_i|$ が大きいほど $\text{sgn}(u_i) = s_i$ が成り立っている確率が高いから、直接こうした細胞の出力を反転するわけにはいかない。代わりに、そのような細胞への寄与が大きく、かつ $|u|$ が比較的小さい細胞の出力を反転するのである。

このダイナミクスにおいて、パラメータ h および λ の選び方が重要である。 h が大きすぎたり λ が小さすぎたりすると、出力の反転する細胞がほとんどなく、従来のダイナミクスと同じになるし、その逆だとほとんどの x_i が反転してしまい、元の情報が失われてしまうからである。理論的な最適値は求められていないが、数値実験によれば、 λ は 2.7 前後、 h は 1.8 ~ 2.0 程度にするのがよい。また、記憶率 r がわかっているのであれば

$$h = 1 + r + k\sqrt{r} \quad (3.9)$$

に従って h を変える[†]方がよい ($k \simeq 2$)。

$X = S$ のとき、 $s_i u_i$ の値は平均 $1 + r$ 、分散 r の正規分布に従うから、パラメータを適切に選ぶと、 X が S に十分近づいたとき、 $|u_i| > h$ となる細胞の割合は非常に小さくなる。従って、正しく想起できたときには、フェーズ II での x_i の反転がほとんど起こらない。また、その逆が常に成り立つという保証はまだ得られていないが、平衡状態で反転が起こらなければ、正しい想起ができたと判断してよいと思われる。

ところで、式 (3.7) において、 X_t の代わりに $\text{sgn}(WX_t)$ を用いると、

$$X_{t+1} = \text{sgn}(W\psi(WX_t)) \quad (3.10)$$

すなわち、出力関数 (u_i から x_i への関数) として $\psi(u)$ と $\text{sgn}(u)$ を交互に用いるという、より単純なアルゴリズムが得られる。ここで、

$$\psi(u) = \text{sgn}(u) - \phi_h(u) = \begin{cases} \lambda - 1 & (u < -h) \\ -1 & (-h \leq u \leq 0) \\ 1 & (0 < u \leq h) \\ 1 - \lambda & (u > h) \end{cases} \quad (3.11)$$

[†] $u_{ii} = 0$ の場合には、 $h = 1 + k\sqrt{r}$ となる。

である(ベクトルに対しては要素ごとに作用する)。この方法は、従来のダイナミクスよりは優れているが、式(3.7)で表されるダイナミクスに比べると、記憶容量や想起能力を増やす効果が小さい。

3.1.3 数値実験

部分反転法を用いることによって想起特性が実際にどう改善されるかを調べるため、数値実験を行った。

まず、重なり率 p の推移を図3.2に示す。実験は、ダイナミクス以外すべて図2.1と同じ条件で行った。パラメータは、 $\lambda = 2.7$ 、 h は式(3.9)に従って $r = 0.08$ のとき1.65、 $r = 0.2$ のとき2.09とした。

この図から、次のことがわかる。

1) 臨界重なり率 p_c が、従来より小さい。これは、 S の引き込み領域 $B(S)$ が拡大し、想起能力が高まったことを意味する。

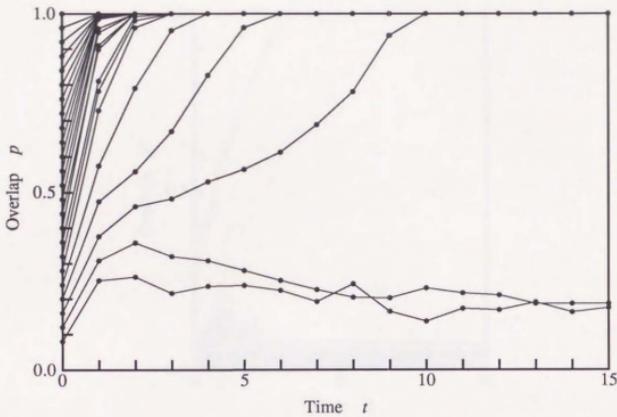
2) 従来のダイナミクスで想起可能な場合においても、 X_t が平衡点 \hat{S} に至るまでに要する時間(ステップ数)が短くなる。また、グラフからは読みとりにくいだが、より S に近いところにある \hat{S} に達する(想起誤差が小さくなる)場合が多い。

3) 記憶容量が増加し、 $r > r_c$ でも想起が可能となる。 r の上限はパラメータ等にも依存するためはっきりしないが、層状モデルの最大記憶率とほぼ等しい。これが偶然の一致であるか、なんらかの理由があつてのことかは、今のところ不明である。

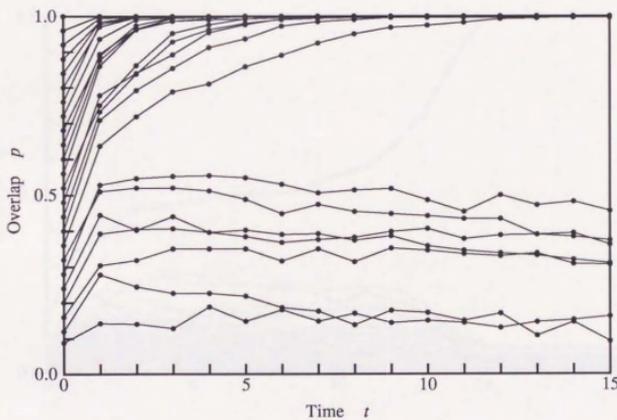
なお、 n および r が大きいとき、想起がほぼうまくいったにもかかわらず、平衡状態に達しないことがある。この場合、 X は S の近傍にある複数の状態間を周期的に移り続ける。

次いで、 q^ν ($\nu = 2, 3, \dots, m$)の推移を調べた結果を図3.3に示す。(a)は図2.4(b)と同じ初期状態を与えた場合であるが、 $t > 1$ で q^ν はすばやく減少し、4ステップで想起が完了している。また、(b)は p_0 がほぼ臨界値 p_c に等しく、想起に成功するまで長い時間を要するような場合の例である。 q^ν の増加が抑えられ、 p との差が徐々に大きくなっていく様子を読み取れるが、これを一種の競合現象とみなすこともできよう。

図3.4は、 σ^2 と p^2 を座標軸にとり、各フェーズの直後の状態をプロットしたグラフである。実線はフェーズIでの状態遷移、点線はフェーズIIでの遷移を表す。

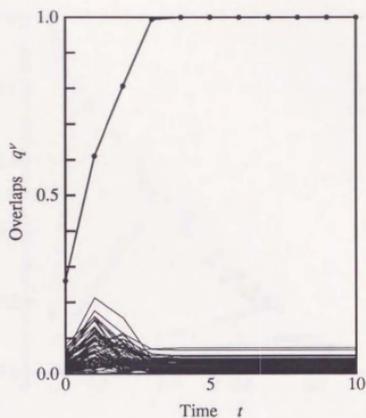


(a) $r = 0.08$

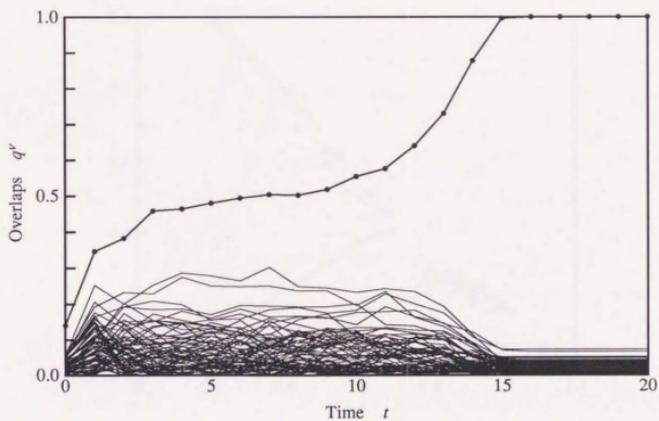


(b) $r = 0.2$

図3.2 部分反転法を用いたときの p の時間変化

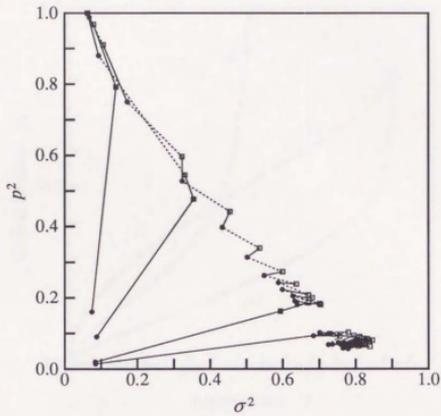


(a) $p_0 = 0.26$

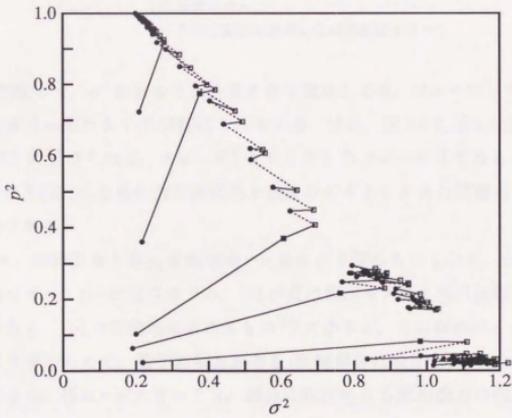


(b) $p_0 = 0.14$

図 3.3 q^ν ($\nu = 2, 3, \dots, m$) の推移 ($r = 0.08$)



(a) $r = 0.08$



(b) $r = 0.2$

図3.4 部分反転法の想起過程

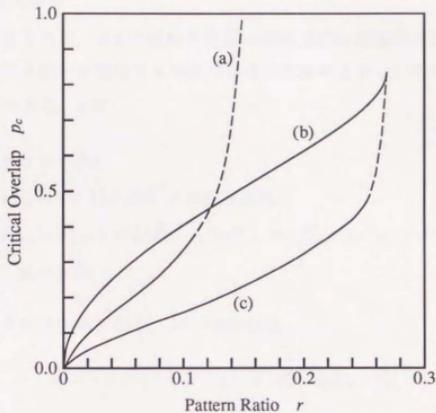


図 3.5 臨界重なり率 p_c と r との関係

- (a) 従来の自己相関モデル
- (b) 層状モデル
- (c) 部分反転法を適用した自己相関モデル

図 2.5 と比較して、 σ^2 がかなりの大ききまで増加しても、フェーズ II で σ^2 が減少することによってうまく S が想起されている。また、図 3.1 に示した図式とは違い、 X が S に近づくのは、フェーズ I よりもむしろフェーズ II であることがわかる。このことは、この系の力学的構造や従来のダイナミクスの問題点を考える上で示唆的である。

図 3.5 に、臨界重なり率 p_c が記憶率 r と共にどう変わるかを示す。(a) は従来の自己相関モデル、(b) は層状モデル、(c) が自己相関モデルに部分反転法を適用した場合である。(b) は理論的に求めたもの^[16]であるが、それ以外は $n = 1000$ の実験により求めたので、若干の誤差がある(破線部分は推定値)。このグラフからわかるように、特に r が大きいとき、部分反転法による想起能力の向上が顕著である。

3.1.4 理論的解析

甘利(私信)によれば, 2.3で触れた従来の想起過程の理論的解析と同様な解析が, 部分反転法を用いた場合にも可能である. これによると, ある初期重なり率 p_0 が与えられたとき, p は

$$p_{t+1} = \Phi(\bar{p}_t) \quad (3.12)$$

$$\sigma_{t+1}^2 = r + 4\{\varphi(\bar{p}_t)\}^2 + 4r\bar{p}_t\varphi(\bar{p}_t)p_{t+1} \quad (3.13)$$

$$\bar{\sigma}_{t+1}^2 = \sigma_{t+1}^2 + r\lambda\{\lambda\tilde{F}_{t+1} - 2p_t F_{t+1} + 2F'_{t+1}(\lambda p_t F_{t+1} - 1)\} \quad (3.14)$$

$$\bar{p}_t = p_t/\bar{\sigma}_t \quad (3.15)$$

に従って変化する ($t \geq 0$). 但し, $\bar{\sigma}^2$ の初期値は

$$\bar{\sigma}_0^2 = r + r\lambda\{\lambda\tilde{F}_0 - 2p_0 F_0 + 2F'_0(\lambda p_0 F_0 - 1)\} \quad (3.16)$$

で与えられる. また, $\varphi(u)$ および $\Phi(u)$ は前述の解析と同じ式(2.18), (2.19)であり, F, \tilde{F}, F' はそれぞれ

$$F_t = \int_{(h-p_t)/\sigma_t}^{(h+p_t)/\sigma_t} \varphi(u) du \quad (3.17)$$

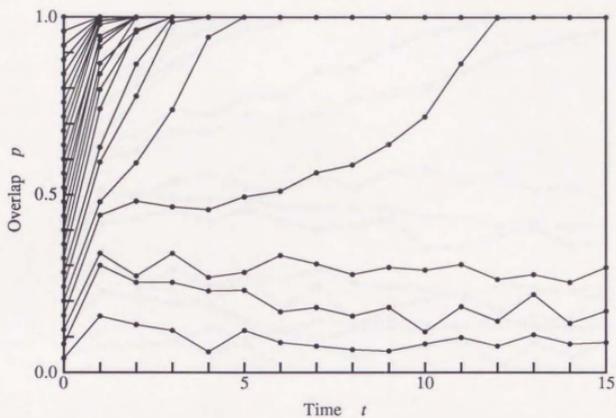
$$\tilde{F}_t = \int_{(h-p_t)/\sigma_t}^{\infty} \varphi(u) du + \int_{(h+p_t)/\sigma_t}^{\infty} \varphi(u) du \quad (3.18)$$

$$F'_t = \frac{1}{\sqrt{2\pi}\sigma_t} \left\{ \exp\left[-\frac{(h-p_t)^2}{2\sigma_t^2}\right] + \exp\left[-\frac{(h+p_t)^2}{2\sigma_t^2}\right] \right\} \quad (3.19)$$

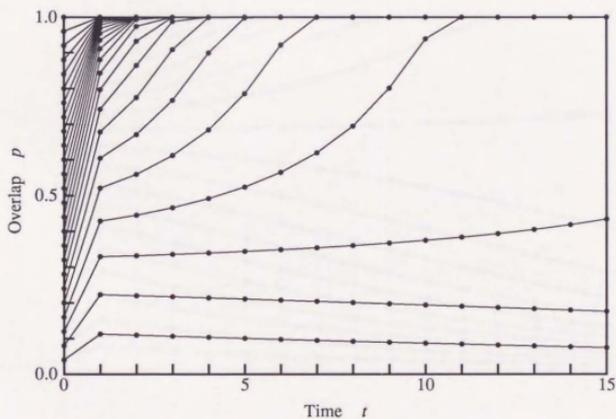
である.

実際に, これらの式を用いて p の推移を計算してみた. いくつかの場合について計算した結果を, $n = 1000$ で行った数値実験の結果と併せて, 図3.6~図3.8に示す. なお, この解析では W の対角成分を0とおいているため, 数値実験の方でも $w_{ii} = 0$ としてある.

これらの図からわかるように, この解析は実験結果とかなりよく一致する. しかし, 理論式に従って順次 p_t を求めていくと, 条件によっては(例えば, $r = 0.2$, $\lambda = 2.7$ で $h < 1.73$ のとき) $\bar{\sigma}^2$ が負になってしまうことがある. その原因は不明であるが, この解析結果にまだ多少の誤りが含まれている可能性もある.

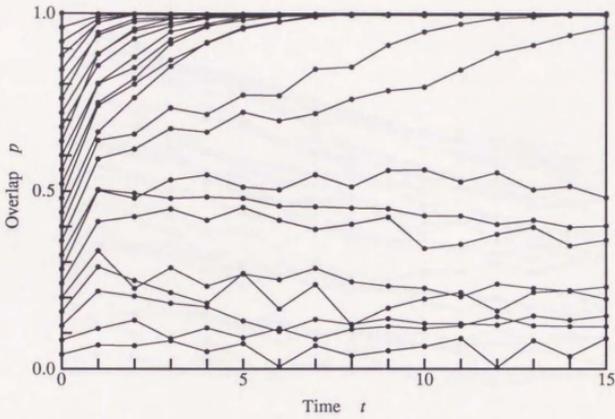


(a) 数値実験

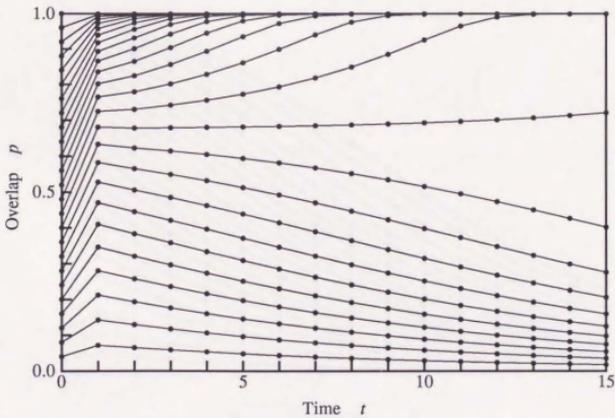


(b) 理論

図 3.6 p の時間変化 ($r = 0.08$, $h = 1.57$, $\lambda = 2.7$)

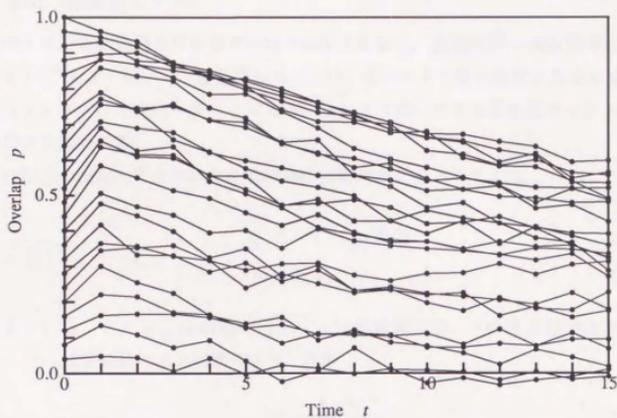


(a) 数値実験

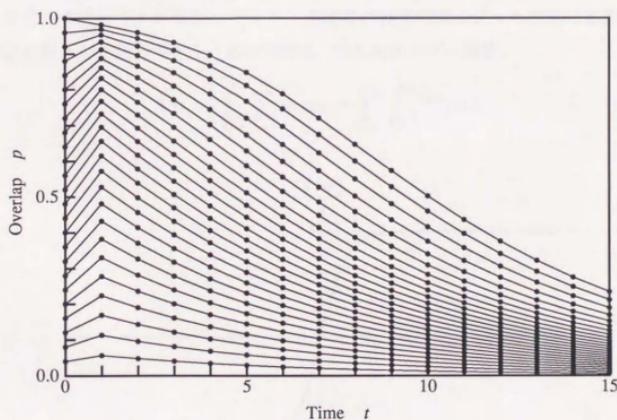


(b) 理論

図 3.7 p の時間変化 ($r = 0.2$, $h = 1.89$, $\lambda = 2.7$)



(a) 数値実験



(b) 理論

図 3.8 p の時間変化 ($r = 0.32$, $h = 2.13$, $\lambda = 2.7$)

3.2 連続型モデル

例えば神経回路網をアナログ回路で実現する場合、連続時間—連続情報の細胞モデル^[10]をよく用いる。連続型のモデルは、その出力を発火頻度とみなせば、現実のニューロンの動作モデルとしても、これまで扱ってきた離散型のモデルより適当だと思われる。

連続型の細胞モデルからなる連想神経回路網のダイナミクスは、

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij} y_j \quad (3.20)$$

$$y_i = f(u_i) \quad (3.21)$$

で表される。ここで y_i は細胞の出力、 u_i は平均膜電位、 τ は時定数である。また、 $f(u)$ は非線形の出力関数であり、通常

$$f(u) = \frac{1 - e^{-cu}}{1 + e^{-cu}} \quad (3.22)$$

で表されるシグモイド形の単調増加関数(図3.9)を用いる(c は正の定数)。

従来のダイナミクスを用いる限り、連続型の連想記憶モデルの動作は離散型の場合とほとんど同じであり、系の状態は、常にエネルギー関数

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} y_i y_j + \sum_{i=1}^n \int_0^{y_i} f^{-1}(z) dz \quad (3.23)$$

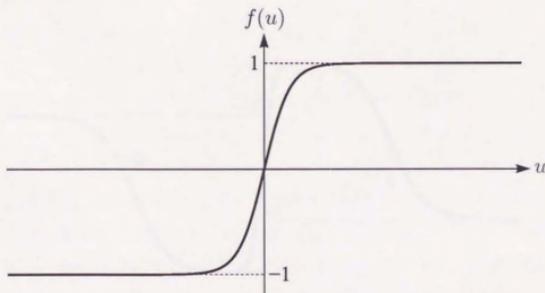


図3.9 シグモイド形出力関数

が最も減少する方向に変化する^[24]。しかし、以下に述べるダイナミクスを用いると、こうしたエネルギー関数は定義できず、離散型のモデルとかなり違った性質が現れる。

本節では、このような連続型の神経回路モデルを扱う。

3.2.1 非単調ダイナミクス

前節で述べた部分反転法のアルゴリズムは、想起の1ステップを二つのフェーズに分けることを必要とする。そのため、これを連続型モデルに対して、そのままの形で適用することはできない。

しかしながら、前述のアルゴリズムの本質は、非常に大きな入力を受ける細胞がないようにすることにあつたから、細胞の出力関数 $f(u)$ を図 3.10 に示すような形の非単調関数にすれば、部分反転法を用いるのと同じような効果が得られると期待される。というのは、このような関数を用いると、 u_i が大きくなりすぎたときには出力 y_i が小さくなり、その結果入力が減少して ($w_{ij} = w_{ji}$ であることに注意) u_i の増加が抑えられるはずだからである。図 3.10 の関数は、式 (3.10) のアルゴリズムで用いる二つの関数 $\psi(u)$ と $\text{sgn}(u)$ を平均し、なめらかにしたものだという見方もできる。

この非単調関数を式で表すと、

$$f(u) = \frac{1 - e^{-cu}}{1 + e^{-cu}} \cdot \frac{1 + \kappa e^{\epsilon(|u|-h)}}{1 + e^{\epsilon(|u|-h)}} \quad (3.24)$$

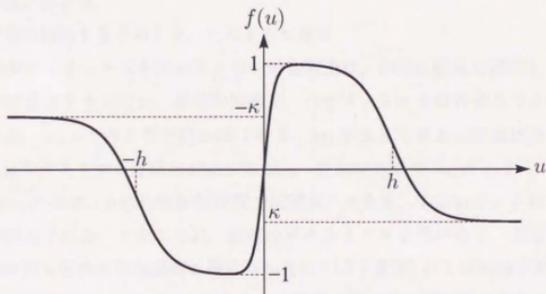


図 3.10 非単調出力関数

となる。ここで d と h は正の定数、 κ は通常負の定数である。但し、 $f(u)$ の形やパラメータの値は、モデルの力学的性質にあまり敏感には効いてこない。

ところで、 $\kappa = 1$ のとき、式(3.24)の $f(u)$ は式(3.22)のシグモイド関数に一致する。以下では、 $f(u)$ は常に式(3.24)で与えられるものとし、特に $\kappa = 1$ とした場合を従来のダイナミクス、 $\kappa < 0$ の場合を非単調ダイナミクスと呼ぶ。

なお、この回路網の状態は、ベクトル (u_1, \dots, u_n) を指定しないと決まらないが、以下のように連想記憶モデルとして動作させる際には、 u_i の符号が主要な意味をもつ。そこで、離散型のモデルのときと同様に $x_i = \text{sgn}(u_i)$ を考え、 $X = (x_1, \dots, x_n)^T$ を回路網の状態と呼ぶことにする。従来のダイナミクスで x_i が y_i とほぼ等しいのに対して、非単調ダイナミクスでは、 $|u_i|$ が大きいとき両者の符号が異なる点に注意が必要である。

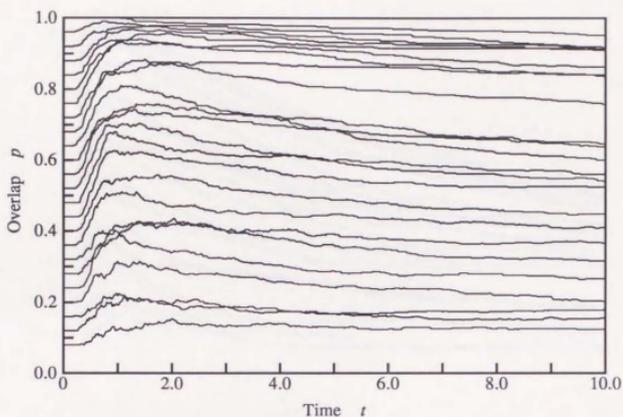
3.2.2 数値実験と考察

連続型の自己相関連想記憶モデルに関して、離散型モデルと同様な数値実験を行った。実験は細胞数 $n = 1000$ で行い、パラメータは $c = 50$, $d' = 15$, $h = 0.5$, $\kappa = -1$ (非単調ダイナミクスの場合) とした。なお、初期状態 X_0 は、 $u_i = 0.5x_i$ とすることによって系に与える。

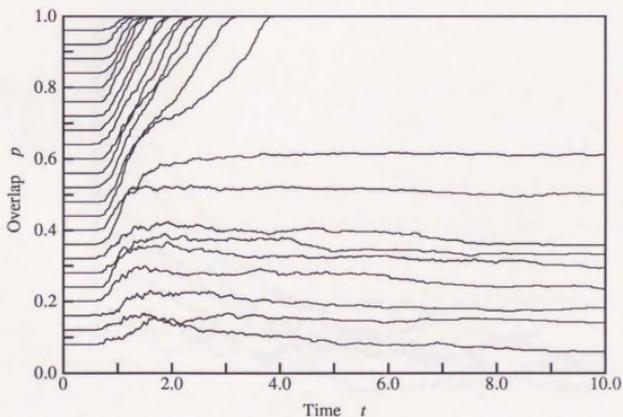
図3.11は、 X と S との重なり率 p (式(2.14)) の変化を示したものである(グラフの横軸は、時定数 τ を単位とする時間を表す)。また、図3.12は、 W の対角成分を0にした場合である。いずれも記憶率 r は0.32であるが、 r が大きいため両者でかなりの違いがある。

この実験の結果をまとめると、次のようになる。

- 1) 非単調ダイナミクスを用いたときの記憶容量は、部分反転法を適用した離散モデルの容量よりも大きい。最大記憶率は、パラメータにも依存するためはつきりしないが、 $w_{ii} = 0$ のときで約0.32である。 w_{ii} を大きくすると容量は更に増えるが、あまり大きくすると偽の記憶が増加し、想起能力も低下してしまう。
- 2) $p_0 > p_c$ のとき、 p は比較的短時間で厳密に1となる。すなわち、 S が想起誤差なしで想起される。このことは、従来のダイナミクスを用いると、想起がうまくいく場合でも多少の想起誤差が避けられない(2.2.1参照)のと対照的である。
- 3) $w_{ii} = 0$ にすると、 $w_{ii} = r$ のときよりも想起に要する時間が短くなる。また、想起がうまくできないとき、 p は急速に小さくなり、その後半永久的に変化し続

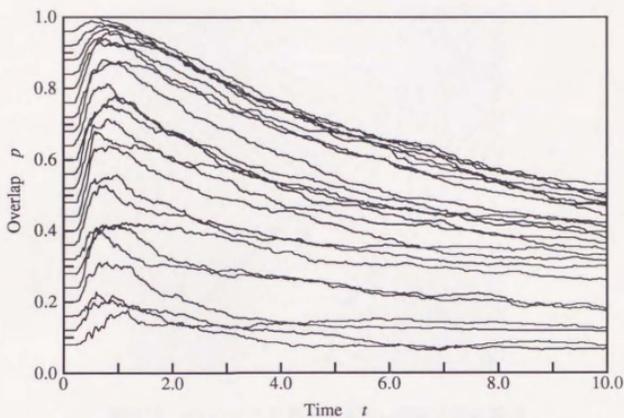


(a) 従来のダイナミクス

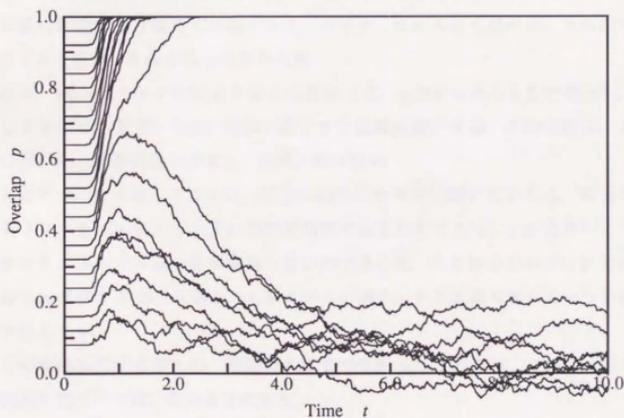


(b) 非単調ダイナミクス

図3.11 重なり率 p の時間変化 ($r = 0.32$, $w_{ii} = r$)

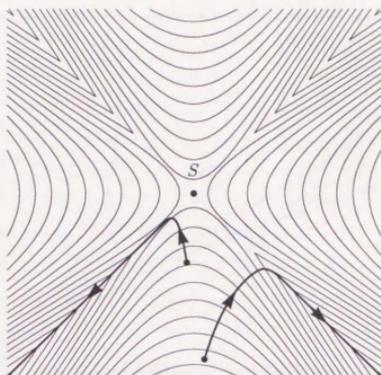


(a) 従来のダイナミクス



(b) 非単調ダイナミクス

図 3.12 重なり率 p の時間変化 ($r = 0.32, w_{ii} = 0$)

図 3.13 $r > r_c$ のときのエネルギー関数と想起過程

ける[†]。このことから直ちに偽の記憶が存在しないと言わわけにはいかないが、安定な偽の記憶状態は極めて少数であり、パラメータをうまく選べば、その数を限りなく0に近づけられるように思われる。

なお、 $w_{ii} = 0$ で r が0.32より少し大きいとき、 p はいったん1まで増加し、しばらくそこにとどまった後、突然0近くまで急激に減少する。この現象は、 S 付近の奇妙な力学的構造を示唆しており、興味深い。

ところで、よく考えてみると、以上の結果はやや不思議に思われる。従来のダイナミクスを用いたとき、 S が不安定状態であるだけでなく、 r が大きいと S の近傍にもエネルギー極小点が存在しないのであった。にもかかわらず、出力関数を変えただけで X が S に達し、しかも多くの場合、そこに落ち着くというのはなぜであろうか。

この疑問に対する答えは、理論的にはまだ与えられていない。しかし、極めて直観的に言うならば、次のようになる。

[†] このときの p の変化はカオス的であり、系の状態は周期アトラクタに達するのではないと思われる。また、 n または m が小さいと、 S 以外の記憶パターン S^p ないしその反転パターン $-S^p$ を想起することがある。

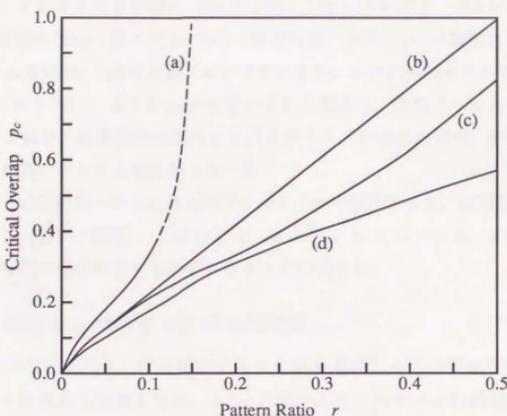


図 3.14 各モデルの臨界重なり率 p_c の変化

- (a) 従来の自己相関モデル
- (b) 一般逆行列モデル
- (c) 一般逆行列モデルで $w_{ii} = 0$ とした場合
- (d) 非単調ダイナミクスを適用した自己相関モデル

2.3.1 で述べたように、従来のモデルにおいて、 S は $r > r_c$ のとき力学系の鞍点になっていると推測される。つまり、エネルギー関数 E は、 S の周辺で図 3.13 に模式的に示すような形をしていると考えてよい (図の細線はエネルギーの等高線を表す)。この図からわかるように、 S の周辺の大部分の領域で、 E は S の側に向かって下っているが、 S まで達するエネルギーの谷があり、その底では S から遠ざかるにつれ E が小さくなる。但し、谷底での E の勾配は緩やかなので、 w_{ii} が大きいと X はそこで止まる。また、谷に近い部分ほど $|u_i|$ の分散が大きく、一部の細胞が非常に強い入力を受けることによってエネルギーが低くなっている。

結局、 S の引き込み領域は、いわば「切れ目」の入った「たらい (basin)」であり、 $|u_i|$ の分散を大きくしない力を加えることによって切れ目がふさがり、 $r > r_c$ でもたらいとして機能する、と解釈することができよう。なお、3.3.3 で行う議論は、この問題を考える上でかなり参考となる。

最後に、 r を変えたときの p_c の変化の様子を図3.14に示す。図の曲線は、それぞれ(a)従来の自己相関モデル、(b)一般逆行列モデル、(c)一般逆行列モデルで $w_{ii}=0$ とした場合、(d)非単調ダイナミクスを用いた自己相関モデルを表す。

(d)の方が(b)や(c)よりも p_c が小さいという事実は、少なくとも自己相関モデルに関する限り、荷重行列の改良よりダイナミクスの改良の方が、想起能力を高めるのに有効であることを物語っている。

なお、一般逆行列モデルに非単調ダイナミクスを適用すると、想起能力は(d)をやや下回る(図3.17参照)。このことは、式(2.9)の W を用いても、 S の周辺力学構造が改善されるにすぎないということを示唆する。

3.3 構造をもつパターンの連想記憶

前節で示したように、非単調ダイナミクスを適用した連続型の自己相関モデルは、数々の優れた特長をもち、2.3.2で述べた従来のモデルの問題点のうち、1)~3)を解決または改善することができる。

では、残された問題、すなわち4)の情報表現に関する問題はどうかであろうか。ここでは、いくつかの数値実験と直観的な考察を通して、この点について考える。なお、本節および次節では、連続型のモデルで荷重行列の対角成分が0の場合のみを扱う。

3.3.1 自己相関モデルと情報表現

従来の自己相関モデルでは、記憶するパターンが相互にほぼ直交している(距離が十分離れている)ことが不可欠であった。そうでなければ、

$$WS = S \quad (3.25)$$

が近似的にすなり成り立たず、 S の近くに平衡点が存在しないからである。

しかし、このような条件は、記憶の情報表現を非常に強く制約する。例えば、ある二つの事物がどれだけ近い関係にあっても、別の記憶事項である限り、全く相関のないパターンとして表現されなければならない。つまり、記憶情報ももつ構造——多くの場合、それは複雑な階層構造を成す——を無視しなければならないのである。このことは、神経回路網による分散型の情報処理の大きな利点を捨てるようなものであり、かなり深刻な問題である。

一般逆行列モデルのように、荷重行列を改良するのは、この問題に対する一つの解決法である。しかし、後述のように、それだけでは記憶パターンに含まれている構造を利用することはできない。情報表現の制約を緩め、記憶の構造をうまく生かすためには、やはりダイナミクスの改良が必要なのである。

では、前節で扱った非単調ダイナミクスを用いたモデルに、相互に相関のあるパターンを記憶させるとどうなるであろうか。ここでは、次のような場合を例に取り、記憶パターンのもつ構造が、モデルの動作にどう影響するかを調べる^[25]。

3.3.2 クラスタを成すパターンの想起

まず、 k 個の二値(± 1)パターンを独立かつランダムに選び、 C^1, C^2, \dots, C^k とする。次に、 $C^\mu = (c_1^\mu, \dots, c_n^\mu)$ とある相関 $a (\geq 0)$ をもつパターンの中から l 個をランダムに選び、 $S^{\mu 1}, S^{\mu 2}, \dots, S^{\mu l}$ とする。すなわち、すべての μ, ν に対して

$$\text{Cor}(C^\mu, S^{\mu\nu}) = \frac{1}{n} \sum_{i=1}^n c_i^\mu s_i^{\mu\nu} = a \quad (3.26)$$

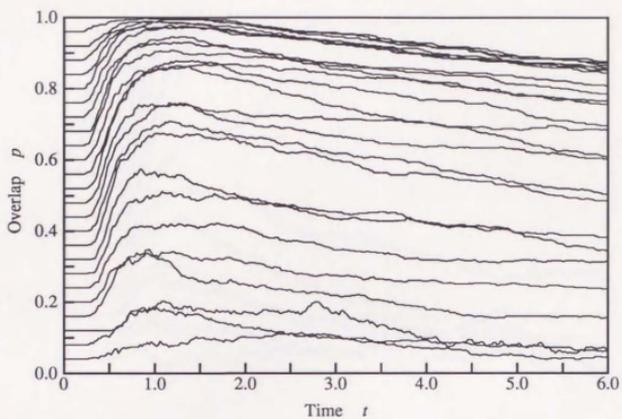
が成り立つようにする。実際には、 C^μ の要素のうち $(1-a)/2$ 個をランダムに選び、その符号を反転することによって $S^{\mu\nu}$ を生産すればよい。

a が十分小さいとき、 $S^{\mu\nu}$ は状態空間中に一様に分布する。しかし、 a が大きくなると、これらは C^μ を中心に l 個ずつクラスタを成して存在するようになる。また、一つのクラスタに属するすべてのパターンは、相互に約 a^2 の相関をもつ。

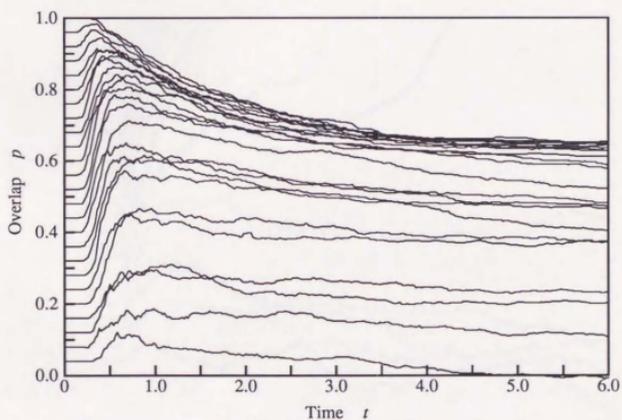
こうして選んだ $k \times l$ 個のパターンを、式(2.3)で与えられる自己相関行列を用いてモデルに記憶させる(但し、 $S^1 = S^{11}, S^2 = S^{12}, \dots, S^m = S^{kl}$ とおく)。

さて、 $n = 1000, k = 50, l = 4$ ($m = 200$)の場合について、数値実験を行った。まず、重なり率 p の推移を、図3.15(従来のダイナミクス)と図3.16(非単調ダイナミクス)に示す。いずれも、(a)は $a = 0$ 、すなわちパターン間の相関がない場合、(b)は $a = 0.6$ の場合である。初期状態 X_0 が $S (= S^{11})$ と p_0 の重なり率(相関)をもつとき、(b)では X_0 と $S^{1\nu}$ ($\nu = 2, 3, \dots, l$)の間にも約 $0.36p_0$ の相関があることを注意しておく。

従来のダイナミクスを用いると、 a の値に関係なく想起がうまくできないが、図3.15の(a)と(b)を比べると、後者では p_0 が1に近いとき、 p が約0.6まで急速に減少することがわかる。このことは、 a がある程度大きくなると、 S の周辺に

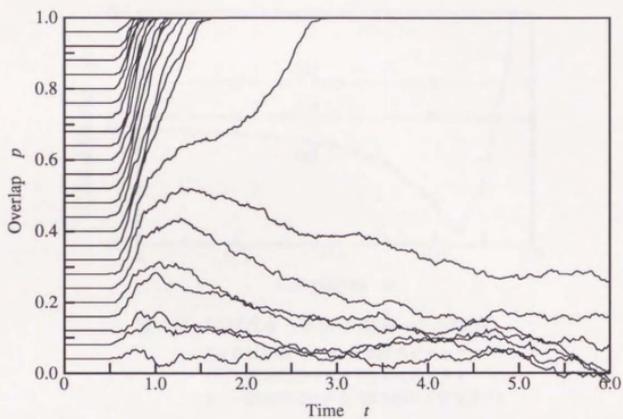


(a) $a = 0$

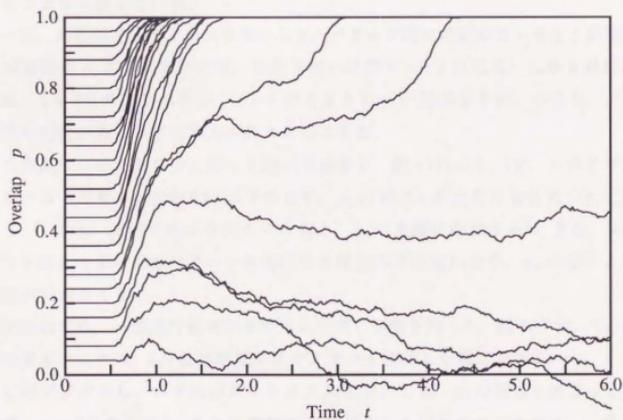


(b) $a = 0.6$

図 3.15 従来のダイナミクスを用いた場合の p の推移



(a) $a = 0$



(b) $a = 0.6$

図 3.16 非単調ダイナミクスを用いた場合の p の推移

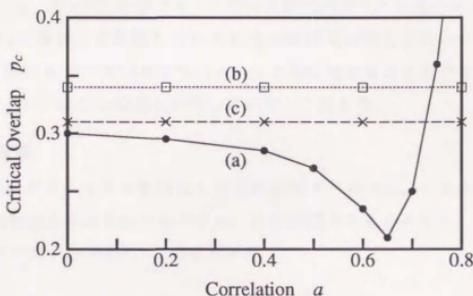


図 3.17 臨界重なり率 p_c と相関 a との関係

- (a) 自己相関モデル (非単調ダイナミクス)
 (b) 一般逆行列モデル (従来のダイナミクス)
 (c) 一般逆行列モデル (非単調ダイナミクス)

は平衡状態がほとんど存在しなくなり、クラスタの中心 C^1 の近くに偽の記憶が集まることを示唆している。

一方、非単調ダイナミクスを用いると、パターン間に相関がないときと同様に、 p_0 が臨界値 p_c より大きければ、かなり短い時間で $p=1$ になる。しかも意外なことに、 $a=0.6$ のときの方が、 $a=0$ のときより p_c の値が小さい。つまり、パターン間に相関があると想起能力が高まるのである。

この点についてももう少し詳しく調べた結果を、図 3.17(a) に示す。このグラフからわかるように、 a が約 0.65 以下のとき、 p_c の値は a が大きくなるにつれて減少する。しかし、 a がそれより大きくなると、 p_c は急激に増加する[†]。また、 a が約 0.75 を超えると、記憶パターンに対応する状態が不安定になり、 p_0 の値によらず想起ができなくなる。

比較のため、一般逆行列モデルについて同じ実験を行った。図 3.17 で、(b) は従来のダイナミクス、(c) は非単調ダイナミクスを適用した場合である。

このグラフから、いずれのダイナミクスを用いても、 p_c の値は a によって変化せず、 $a < 0.7$ のとき (a) よりも想起能力が劣ることがわかる。これは、一般逆行

[†] このとき、 $p_0 < p_c$ で想起が成功したり、 $p_0 > p_c$ でもうまく想起できない場合が非常に多く、 p_c の値ははっきりしない。

列モデルでは、すべての記憶パターンを他の記憶パターンと直交するパターン(共役ベクトル)に変換して記憶していることが原因だと考えられる。つまり、このモデルは、非常に強い相関をもつパターンでも記憶することができるが、その反面、記憶パターンがもつ構造を利用していないと言える。

3.3.3 考察

では、なぜダイナミクスを改良した自己相関モデルでは、パターン間の適度な相関が想起能力を高めるのであろうか。この問題を考えるために、まず回路網の状態 X とクラスタの中心 C^1 との重なり率

$$p' = \text{Cor}(X, C^1) = \frac{1}{n} \sum_{i=1}^n x_i c_i^1 \quad (3.27)$$

を定義する。

p' の時間変化の一例を、 p の推移と共に図 3.18 に示す。これは、 p_0 がほぼ p_c に等しく、想起に長い時間がかかるような場合である。このグラフから読み取れるように、 p' は最初 p よりも大幅に増加し ($p'_0 \simeq a p_0$ に注意)、一時は p を超えるが、その後やや減少して $0.6 (= a)$ に落ち着く。

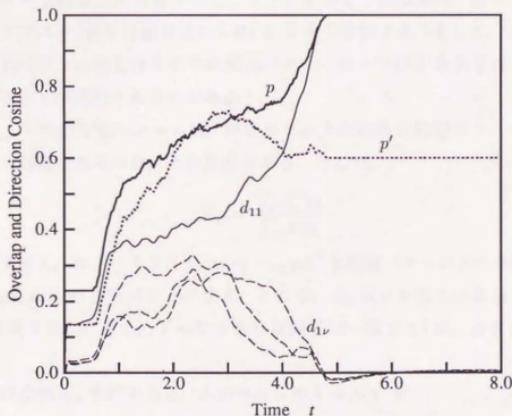


図 3.18 非単調ダイナミクスによる想起の過程 ($p_0 = 0.23$)

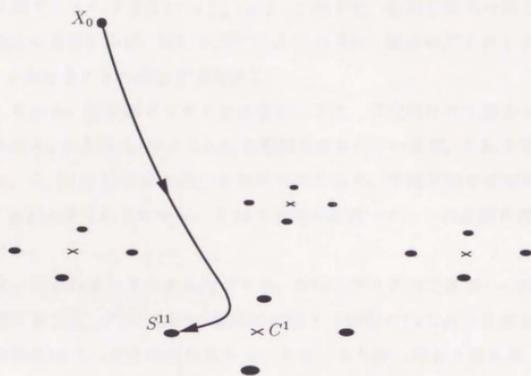


図 3.19 想起過程の模式図

このことから、図 3.19 に模式的に示すような形で想起が行われているという解釈が可能である。すなわち、 X がどの記憶パターンからも十分離れているときには、クラスタ全体があたかも一つのアトラクタのように振舞い、最も引きつける力の強いクラスタ (通常は最も近いもの) が X を引き寄せ、そして、 X がそのクラスタに近づくと、今度はその中の記憶パターンの一つが X を引き寄せ、最終的にそのパターンが想起されるのである。

問題は、一つの記憶パターンが、同じクラスタ内の他の記憶パターンからどのようにして分離されるのか、という点である。そこで、

$$d_{\mu\nu} = \frac{\sum_i s_i^{\mu\nu} y_i}{\sum_i x_i y_i} \quad (3.28)$$

という量を考え、出力ベクトル $Y = (y_1, \dots, y_n)^T$ と記憶パターン $S^{\mu\nu}$ の間の方向余弦 (direction cosine) と呼ぶことにする。ここで、 d_{11} は p と似ている ($p=1$ のとき $d_{11} = 1$ が成り立つ。また、 $Y=X$ であれば両者は一致する) が、必ずしも等価ではない。

この方向余弦 $d_{\mu\nu}$ を用いると、入力ベクトル $V = WY$ を

$$V = \frac{\gamma}{n} (d_{11} S^{11} + d_{12} S^{12} + \dots + d_{kl} S^{kl}) - rY \quad (3.29)$$

のように表すことができる ($\gamma = \sum_i x_i y_i$). この式で、右辺括弧内の第2項以降はノイズ成分に相当するが、特に $d_{1\nu} S^{1\nu}$ ($\nu \neq 1$) の項は、従来のダイナミクスを用いる限り、 p が大きくなれば必ず増大する。

しかしながら、非単調ダイナミクスを用いると、状況はかなり変わってくる。図3.18には d_{11} および $d_{1\nu}$ ($\nu = 2, 3, 4$) の推移も示されているが、これを見るとわかるように、 $d_{1\nu}$ ($\nu \neq 1$) は p の減少と共に小さくなり、平衡状態ではほぼ0となる。つまり、 p が大きくなるにつれ、 X は S 以外の記憶パターンの影響をほとんど受けなくなる。

これは、 X があるクラスタに近づくと、そのクラスタ内の数多くの記憶パターンに共通する ($|\sum_{\mu} s_{\mu}^i|$ が大きい) 要素に対応する細胞の出力が小さくなり、それ以外の細胞の出力が支配的になるからである。もう少し詳しく言えば、 $x_i = s_i^{1\nu}$ が多くの ν に対して成り立つような細胞ほど、他の細胞から強い入力を受けて $|u_i|$ が大きくなり、 $x_i y_i$ の値が減少する。そのため、 Y と $S^{1\nu}$ ($\nu \neq 1$) の内積が(従って $d_{1\nu}$ も) ほぼ0になり、結果として式(3.29)のノイズ項がほとんど消えてなくなるわけである。

ところで、図3.17(a)において、 $a > 0.65$ で p_c が急増し、モデルがうまく動作しなくなるのは、 S 以外の記憶パターンが近すぎるため、 X はどれだけ S の近くにいっても C^1 の方へ引かれていってしまうからだと考えられる。このような状況では、もはや自己相関行列を用いて S を記憶するのは無理であり、荷重行列のある程度改良することが必要になってくる。

3.4 非単調ダイナミクスを利用した学習

前節の最後に述べたように、非常に強い相関をもつパターンを安定に記憶するためには、荷重行列を改良する必要がある。しかし、一般逆行列モデルと同じ W を用いたのでは、記憶パターンがもつ構造をうまく生かすことができない。何か、自己相関モデルと一般逆行列モデルの長所を併せもつ(前節の例で言うならば、 a が小さいとき前者に近く、 a が大きいと後者に近くなる)ような、うまいモデルは作れないだろうか。

単純に考えれば、例えば W を式(2.3)と(2.9)の荷重平均とし、パターン間の相関が大きいきほど後者の重みを増やせば良いように思われる。しかしながら、このようなやり方は、クラスタごとに相関の大きさが違うだけでも適用が難しくなる。しかも、一般には、記憶パターンはもっと複雑な構造をしている。従って、一般の場合について、適切な結合荷重を明示的な式で与えることは、非常に困難だと言える。

結局、学習によって W を適応的に構成することが望ましい。では、どのような学習則を用いるべきであろうか。また、非単調ダイナミクスを用いた神経回路網では、なんらかの意味で従来より優れた学習が可能なのであろうか。

本節では、これらの問題について議論する。

3.4.1 相関学習と直交学習

2.1.1で触れたように、自己相関モデルの荷重行列 W は、相関学習によって構成することができる。具体的には、

$$\tau \frac{dw_{ij}}{dt} = -w_{ij} + \alpha s_i^\mu s_j^\mu \quad (3.30)$$

という学習則(τ および α は正の定数)に従って w_{ij} を更新することを、すべての μ について十分な回数繰り返すと、 $W = [w_{ij}]$ は式(2.3)の W (の定数倍)にほぼ収束する。

一方、一般逆行列モデルの荷重行列は、直交学習^[10]

$$\tau \frac{dw_{ij}}{dt} = -w_{ij} + \varepsilon \left\{ \sum_{j=1}^n w_{ij} s_j^\mu - s_i^\mu \right\} \quad (3.31)$$

によって近似的に実現される (ε は正の定数)[†]。

従って、望ましい学習則は、ある意味で相関学習と直交学習の中間に位置すると考えられる。しかし、単純に式(3.30)と(3.31)を折衷するのでは、先に述べた式(2.3)と(2.9)の平均を取る方法と同じことである。入力しているパターン S の近傍に、他の記憶パターンがあれば直交学習に近く、そうでなければ相関学習に近いような学習則でなければならない。

そのような学習を実現するには、回路網のダイナミクスをうまく利用するのが一番の早道だと思われる。そして、そのためには、やはり改良したダイナミクスを用いるべきであろう。

実は、学習の際に非単調ダイナミクスを用いると、パターン間の相関が大きいとき直交学習の効果をもつよう、相関学習を自然な形で改良することができるのである。以下に、その具体的な方法を示す。

3.4.2 学習則

まず、回路網に外部からの入力を加えられるよう、式(3.20)を少し変え、

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij} y_j + z_i \quad (3.32)$$

とする。ここで、 z_i は i 番目の細胞に対する外部入力である。

さて、パターン S^1, S^2, \dots, S^m を記憶する場合、 $z_i = \beta s_i^\mu$ (β は正の定数) という形で外部から S^μ を順次入力し、式(3.21), (3.24), (3.32) で記述されるダイナミクスに従って系を動作させる。それと並行して、 w_{ij} の学習を

$$\tau' \frac{dw_{ij}}{dt} = -w_{ij} + \alpha y_i y_j \quad (3.33)$$

に従って行う。但し、時定数 τ' は τ より十分大きいものとする。

従来のダイナミクス ($\kappa = 1$) を用いた場合、 β が十分大きければ $y_i = s_i^\mu$ となるから、これは相関学習を行うのと同じである。しかし、非単調ダイナミクスの場合には、必ずしも同じにならない。このことを定性的に説明すると、次のようになる。

[†] ε が小さいほど式(2.9)に近い値に収束するが、収束に時間がかかる。また、 w_{ii} を 0 に固定して学習を行うと、それ以外の成分も式(2.9)とは若干違う値になる。

今、 $S = (1, 1, \dots, 1)^T$ を入力しているとしよう(こう仮定しても一般性を失わない)。また、 W が自己相関行列に近い状態を考える。

S 以外の記憶パターンがすべて S から十分に離れている(相関がほとんどない)とき、 u_i の大きさはどの細胞についてもだいたい同じになる。そうすると、近似的に $Y = kS$ が成り立つ(k は正の定数)から、式(3.33)は式(3.30)とほぼ等価である。すなわち、この学習則は、普通の相関学習と変わらない。

これに対して、もし S の近くに他の記憶パターンがあると、3.3.3で述べたように、他の細胞から受ける入力大きさにばらつきができ、 u_i の分布は分散が大きなものになる。そして、 $f(u)$ の非単調性により、 u が大きい細胞ほど出力 y が小さくなり、逆に u が小さい細胞は大きな出力を出す。そうすると、一般に u の大きな細胞への結合荷重は学習によって小さくなり、その細胞が受ける入力も減少する。 u の小さな細胞はその逆である。

以上のようなメカニズムが働く結果、学習が進むにつれて u_i の分散が小さくなると思われる。このことは、直交学習を行うのと同様な効果をもつ(直交学習を行うと、 u_i の分散は0に収束することに注意)。但し、ある程度 u_i の分散が小さくなると、 y_i のばらつきも減少し、相関学習に近づくことになる。

なお、パラメータの大きさはそれほど重要ではないが、学習がうまくいくにはいくつかの条件があり、パターン間に強い相関があるとき(前節の例では、 a が大きいとき)ほどその条件が厳しくなる。特に、 y_i が x_i と異符号で大きな絶対値をとると、 $|u|$ が大きい細胞同士の結合が増大してしまい、 $|u_i|$ の分散が小さくならないことがあるため、 κ をあまり小さくしてはならない[†]。

ところで、この学習法を用いると、ある程度学習したパターンに関しては、自らが想起したパターンを用いて学習することができる。このとき、外部から与える入力は、完全なパターンでなくてもよいし、その大きさ β も小さいもので十分である。そのため、新たなパターンを記憶する際などに必要な、過去に記憶したパターンの再学習が容易となる。これは、想起のダイナミクスを利用しない普通の教師あり学習にはない特長である。

[†] これを避けるには、式(3.33)を多少変更し、例えば $y_i y_j$ のところを $(y_i - \kappa x_i)(y_j - \kappa x_j)$ などにする必要はあるが、煩雑である。

3.4.3 数値実験

さて、以上のような学習則を、3.3.2で扱ったクラスタを成すパターンの学習に適用してみた。その手順は以下の通りである。

まず、200個のパターン $S^{11}, S^{12}, \dots, S^{kl}$ を適当な順番で入力し、一つのパターンにつき 2τ の間学習を行う。400 τ で一巡するが、これを5回繰り返す。パラメータは $\alpha=4, \beta=0.3, \tau'=4000\tau, \kappa=-0.1$ 、その他は前節と同じである。

このようにして学習を終えた後の回路網に関して、前節と同様な実験を行い、臨界重なり率 p_c と相関 a の関係を調べた。結果を図3.20に示す。なお、前節の実験と合わせるため、 p_c を求める際には $\kappa=-1$ とした($\kappa=-0.1$ のままだと、 p_c が全般に大きくなる)。

グラフから読み取れるように、 a が大きくなるにつれて想起能力が高まるが、 a が0.65を超えても p_c が下がり続ける点で、自己相関モデル(破線)と異なっている。また、 a が0.6前後のとき自己相関モデルより若干劣るものの、それ以外では他のモデルよりも優れた想起能力を示すことがわかる。

なお、学習の回数を増やしても、 p_c の値がこれ以上小さくなることはなく、むしろわずかに増加する。これは、学習を行わずと、必要以上に一般逆行列モデルに近づいてしまうためだと思われる。逆にこのことから、学習回数はかなり少なくてよいと言いうこともできよう。

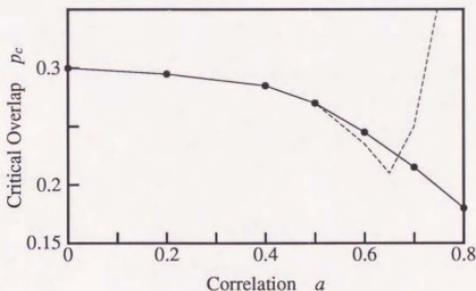


図3.20 学習後の臨界重なり率 p_c

同様に、クラスタが階層を成しており、記憶パターンがより複雑な構造をもつ場合についていくつかの実験を行ったところ、やはり上述の学習をしたモデルが最も高い想起能力を示した。従って、非単調ダイナミクスを用いた学習法は、かなり一般的な場合に適用可能だと考えられる。

側頭葉短期記憶系の力学とそのモデル

第2章および第3章では、連想記憶機能をもつ神経回路モデルの力学を論じてきたが、これらのモデルは、必ずしも現実の脳を反映しているわけではない。では、脳の記憶の神経回路網は、実際にどのような力学的性質をもっているのだろうか。

この問題を考えるための貴重な手がかりを与えてくれるのが、サルの前頭連合野に見られる一群のニューロンである。これらのニューロンは、ある種の力学系を構成することによって短期記憶を担っていると考えられるが、その振舞いは従来の神経回路モデルによって実現される力学系ではうまく説明できない。

本章では、側頭葉の短期記憶系の力学について考察し、これと同様な力学的性質を実現する神経回路モデルを構成する^[26]。このモデルは、その構成単位が局所的な抑制回路とそれによって実現される非単調な入出力特性をもつ点において、従来のモデルと本質的に異なるものであり、脳の記憶回路の動作とそのメカニズムを理解する上で重要な示唆を与えてくれる。

なお、本章で用いる記号の意味は、一部前章までと異なる。但し、特に断りのないものに関しては同じである。

4.1 背景

4.1.1 TE野の短期記憶ニューロン

一瞬(0.2秒)提示されたフラクタル図形と、16秒の遅延期間の後再提示された図形とを比較するという課題(実際の実験手順はもっと緻密である)をサルに課したところ、遅延期間中興奮し続ける一群のニューロンが、下部側頭葉、主にTE野に見つかったと言う^[27,28]。このニューロンには多くの特筆すべき性質があるが、以下の議論と直接関係するものを簡単にまとめると次のようになる(但し、これには甘利ら^[29]および著者の解釈が含まれている)。

1) 過去に何度も学習し、見慣れている図形を提示した場合、一つのニューロンは、ごく少数の図形(100枚のうち2,3枚)に対してだけ強く反応する。それよりやや弱い反応を示す図形も数枚あるが、その他大多数の図形に対してはほとんど反応を示さない。また、強い反応を引き起こす数枚の図形の間に、特に共通する特徴はない。

2) 見慣れた図形に対する反応の再現性は高い。また、回転や拡大・縮小などの変換を施した図形を提示しても、ほとんど同じ反応を示す。

3) 以前に見たことのない新奇な図形を提示したとき、強い反応を示すことはほとんどない。しかし、比較的多数の図形に対して弱く反応する。見慣れた図形の場合に比べ、反応の再現性が低く、時間的な変動も大きい。

以上は個々のニューロンについての実験結果であるが、これからある図形に対するニューロン群全体の反応を、次のように推測することができる。

4) 見慣れた図形を提示したときにTE野に現れるのは、全体のごく一部のニューロンだけが強く興奮している、いわゆるスパースなパターンである。但し、それよりやや弱い反応を示すニューロンや、ごく弱い興奮を持続するニューロンもある程度含んでいる。このパターンは、提示される図形ごとに異なるが、多少の変形に対しては不変である。

5) 新奇な図形を提示すると、見慣れた図形の場合よりやや多数のニューロンが、弱く興奮しているようなパターンが現れる。このパターンもまた図形によって異なるが、同じ図形を見せても同じパターンが現れるとは限らない。

4.1.2 系の力学的性質

さて、これらのニューロンは、互いに結合し合って一つの力学系を構成しており、その相互作用によって短期記憶(どのような図形を提示されたかについて16秒間覚えている情報)を保持しているものと考えられる。短期記憶を担う別のニューロン群があって、そこから強い持続的人力を受けている可能性も残されているが、そのようなニューロン群は発見されていない。また仮にあったとしても、それを含めた全体の系に対して同じ議論が成り立つ。そこで、TE野は、それ自身が短期記憶を保持する力学系になっているとして、次にその性質について考察する。

まず、見慣れた図形はスパースな興奮パターンとして保持されるが、安定で再現性が高いこと、図形の多少の変形に影響されないことから、このパターンは力学系の強い(引き込み領域が広い)アトラクタになっていると考えられる。また、各図形はその図形の特徴に関係なくコードされているから、こうした強いアトラクタはかなり一様に分布していることになる。

一方、新奇な図形もスパースなパターンとして表現されるが、強い興奮を持続するニューロンがほとんどないこと、安定性および再現性があまり高くないことから、このパターンはあまり安定でない平衡状態ないし不完全なアトラクタ(後述)になっていると考えられる。また、このような状態は非常にたくさんあり、初めて見る図形はそのうちのどれか一つにコードされると考えるのが自然であろう^[29~31]。

従って、TE野の短期記憶回路は、少数の強いアトラクタとそうでない多数のアトラクタとが共存し、両方でニューロンの反応強度の分布が異なるような力学系であると言うことができる。

4.1.3 従来の力学系との比較

ところが、以上のような性質をもつ力学系を神経回路網で実現しようとするとき、大きな問題があることに気が付く。その一つは、中くらの強さの興奮を持続するニューロンの割合が大きすぎるという点である。

これについて説明するために、まず、各細胞が0から1の連続値をとる連続時間のHopfieldモデル^[24]を考えよう。そのダイナミクスは次式で与えられる。

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij} x_j - h_i \quad (4.1)$$

$$x_i = f(u_i) \quad (4.2)$$

ここで、 x_i と h_i は、それぞれ i 番目の細胞の出力としきい値である。また、結合は対称 ($w_{ij} = w_{ji}$) で自己結合がない ($w_{ii} = 0$) ものとし、出力関数 $f(u)$ は、 $u \rightarrow -\infty$ で 0, $u \rightarrow \infty$ で 1 をとるシグモイド形関数

$$f(u) = \frac{1}{1 + e^{-cu}} \quad (4.3)$$

とする。

前述の連想記憶モデルと同様、この回路網はエネルギー関数

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j + \sum_{i=1}^n h_i x_i + \sum_{i=1}^n g(x_i) \quad (4.4)$$

が時間と共に減少するように動作し、 E の極小点が系のアトラクタに対応する。ここで、

$$\begin{aligned} g(x) &= \int_{\frac{1}{2}}^x f^{-1}(y) dy \\ &= \frac{1}{c} (x \log x + (1-x) \log(1-x) + \log 2) \end{aligned} \quad (4.5)$$

であり、 $g'(1/2) = 0$ である。

今、式(4.3)の定数 c の値が十分大きいものとしよう。これは、どの細胞も他の細胞と十分強く結合していることと等価である。ここで、 x_i の値が 0.5 に近ければ、 $g'(x_i)$ は十分小さいから、 $\partial E / \partial x_i$ は x_j ($j \neq i$) の 1 次関数とみなすことができる。従って、このとき $\partial E / \partial x_i = 0$ が成り立つことはほとんどない。逆に、 E が極小となるのは、ほとんどすべての x_i が 0 または 1 に非常に近い値をとる場合に限られる。つまり、いくつかの細胞が 0 と 1 の中間的な値を出力するような状態は、強いアトラクタとなり得ない。

一方、 c の値が小さい(相互結合の強さが弱い)と、外部からの入力なくなるととき、初期状態によらず、系の状態はいつも同じになってしまう。また、一部の細胞が他と弱く結合している場合、常にそれらの細胞だけが中間的な出力を出すことになる。

このように、少なくとも Hopfield モデルのような対称結合の回路網では、いくつかの細胞が中間的な出力値をとるアトラクタの存在を説明できない。そして、このような性質は、恐らく式(4.1)~(4.3)で表される力学系の本質的な構造とかかわっており、結合の対称性を多少崩す程度では変化しない。また、ほとんどすべ

での結合を切る ($w_{ij} = 0$ にする) ことによって対称性を完全に無くしても、基本的な力学的性質は同じである^[32,33]。

一方、非対称な神経回路網の力学に関する理論的研究によれば、一様な構造 (w_{ij} の分布が i, j に極端には依存しない) のまま非対称性を強めていくと、平衡状態の数が指数的に減少してしまう^[34]。また、多くの細胞が中間的な値をとる状態が平衡状態となるのは、 W にある特殊な構造が含まれるときに限られることが示唆されている^[35]。これらの点から考えて、たとえ対称性の条件を完全に取っ払ったとしても、一様構造のモデルでは、前述のようなニューロン群の振舞いを説明できないと思われる。

TE野の力学系のモデルを構成しようとする際、もう一つ問題となるのは、見慣れた図形と新規な図形とでニューロンの反応が異なるという点である。従来の連想記憶モデルでは、記憶したパターンとそうでないパターン(偽の記憶)を細胞の出力分布から区別することはできないから、この点においても、従来のモデルはTE野のモデルとして適当でないことがわかる。

4.2 モデルの構成

前節の結果をまとめると、側頭葉短期記憶回路の力学系のモデルを構成するためには、ダイナミクスが従来と全く異なる神経回路網を考えるか、回路網に何らかの構造を導入しなければならないということが言える。仮に後者の方法を取るとして、いったいどのような構造を考えればよいのだろうか。

確かに、大脳の神経回路には比較的整然とした構造があり、何種類ものニューロンからなる局所的な回路がある。しかし、この回路は相当に複雑であり、そのままモデル化することは困難である。また、仮にそれができたとしても、そのようなモデルにあまり意味があるとは思えない。情報処理の原理を明らかにするには、もっと別のアプローチを取る必要があろう。

そこで、ここでは前章までの結果を踏まえ、神経回路網の力学的性質を改善する上で何が本質的であり、どのような構造が必要であるかを探ることにする。

4.2.1 ダイナミクスの改良

既に述べたように、連想神経回路モデルは、出力関数 $f(u)$ を変えるだけで、その力学的性質が大幅に変化する。そして、その際に最も本質的なのは、細胞の入出力関係の非単調性である。というのは、これによって回路網が u の符号以外の情報を利用できるようになり、また式 (3.23) のエネルギー関数が定義されない (E が必ずしも時間と共に減少しない) 力学系になるからである。

非単調な入出力特性には、ダイナミクスを改良する上で、もう一つ重要な意味がある。それは、TE野で保持されるパターンがスパースであることと関係している。

2.3.2 で述べたように、連想記憶モデルでスパースコーディングを行う (記憶パターンをスパースなものに限る) と、1パターン当りの情報量が減るため、より多くのパターンを記憶させることができる。但し、十分な想起能力を得るためには、想起の際に系の活動度 (細胞の出力値の総和) をほぼ一定にする必要がある。つまり、記憶パターンの活動度 (1をとる要素の数) に関する情報を利用して、系のとり得る状態を制限することが、スパースコーディングによるダイナミクスの改良において本質的な役割を果たす。

活動度を低いレベルに保つという方法は、TE野を含めて現実の脳で広く行われていると考えられる。ところが、連想記憶モデルの場合、活動度を一定に保つ機構を自然な形で実現することは、それほど簡単ではない。例えば、単純に活動度に応じた抑制をフィードバック回路によって加える方法は、振動が起りやすいなどの問題があり、なかなかうまくいかない。しかし、もし入力がある程度大きくなると出力が減少に転じるような特性が個々の細胞に備わっていれば、活動度のコントロールはかなり容易になると考えられる。このことは、後述のシミュレーション実験でも確かめられる。

4.2.2 モデルの構造

以上のように、細胞の非単調特性は、連想記憶のダイナミクスの改良と密接にかかわっている。しかしながら、現実のニューロンが単調な特性をもつことは事実であるから、なるべく自然なモデルを構成するために、複数の細胞の組み合わせを考えなければならない。そして、非単調な特性を得るには、細胞の出力で

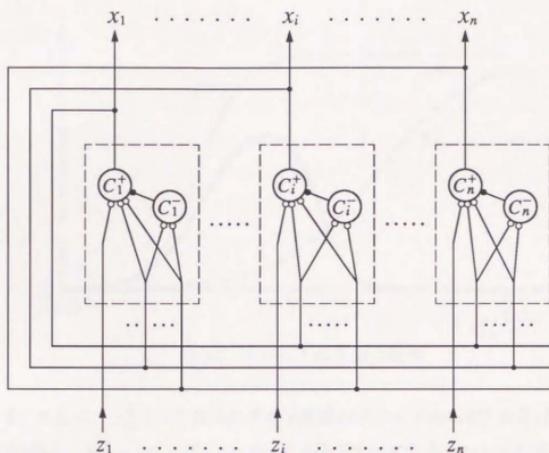


図 4.1 モデルの構成図

なく入力に応じて抑制を加えること，すなわちフィードフォワード型の抑制が不可欠である。

このような考察に基づいて構成したのが，以下に示すモデルである．まず，全体の構造を図 4.1 に示す．

図の破線で囲まれた部分が，ユニットと呼ぶモデルの構成単位であり，従来の一つの細胞に対応する． i 番目のユニットは出力細胞 C_i^+ と抑制細胞 C_i^- の二つからなるが，前者はこのユニットの出力 x_i を出す細胞，後者は y_i を出力することによって前者に強い抑制を加える細胞である．他のユニットからの入力は C_i^+ と C_i^- の両方に入るが，系の外部からの刺激 z_i は C_i^+ だけに入力される．式で表すと，

$$y_i = f\left(\lambda \sum_{j=1}^n w_{ij}^- x_j - \theta\right) \quad (4.6)$$

$$\tau \frac{du_i}{dt} = -u_i + \sum_{j=1}^n w_{ij}^+ x_j - w_j y_i + z_i \quad (4.7)$$

$$x_i = f(u_i) \quad (4.8)$$

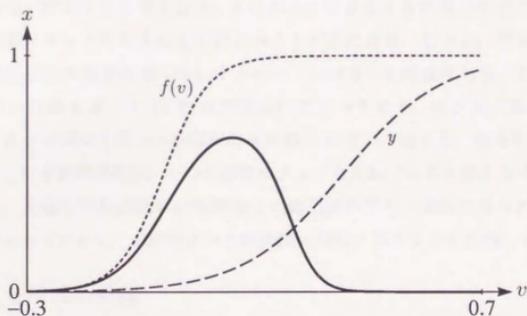


図4.2 ユニットの入出力特性

となる。ここで、 w_{ij}^+ と w_{ij}^- はそれぞれj番目のユニットから C_i^+ および C_i^- への結合荷重(但し、 $w_{ii}^+ = w_{ii}^- = 0$)、 w_I は C^- から C^+ (iによらない)への結合荷重、 λ 、 θ および τ は正の定数である。また、出力関数 $f(u)$ は、0から1の値をとる単調増加関数であり、式(4.3)のシグモイド関数を用いる。

いま、すべてのjについて $w_{ij}^+ = w_{ij}^-$ が成り立っているとしよう。このとき、 C_i^- が受ける入力荷重和は、 C_i^+ への入力荷重和

$$v_i = \sum_{j=1}^n w_{ij}^+ x_j \quad (4.9)$$

と常に等しいから、このユニットの出力 x_i は v_i だけの関数となる。

ここでパラメータを適当に選べば、 $y = f(\lambda v - \theta)$ の非線形性により、 x_i は v_i の増加に対して単調に増加するのではなく、図4.2のようなベル型の特性を示す。つまり、 v_i がある程度大きくなると、 y_i が急激に増加して C_i^+ を強く抑制するため、 v_i の増加に対して x_i が減少するようになる。

w_{ij}^+ と w_{ij}^- とが等しくない場合にも、両者の相関が高ければ、やはり入力がある程度大きくなったところで出力が減少に転じる。但しこの場合には、 v_i の大きさは同じでも、入力のパターン(どのユニットからの入力が大きいか)によって x_i の値が変わることになる。

ところで、図 3.10 から考えると、入力小さすぎるときに出力が増加するような特性を各ユニットにもたせる方がよいように思われる。しかし、そうするとモデルが必要以上に複雑になってしまうので、このような構成にした。また、ここに示したのはあくまでも TE 野の力学系のモデルであり、モデルの細胞と現実のニューロンの間に 1 対 1 の対応があるわけではない。むしろ、ある程度の数のニューロンの平均的動作が、一つの細胞によって表されていると考えるべきであろう。但し、大脳の神経回路には局所的な抑制回路がごく一般的に見られ、その大部分がフィードフォワード型だという興味深い知見があることを指摘しておく。

4.3 モデルの動作

4.3.1 スパースなパターンの連想記憶

さて、上記の回路網に、 m 個のスパースなパターン S^1, S^2, \dots, S^m を記憶させるとしよう。但し、 $S^\mu = (s_1^\mu, \dots, s_n^\mu)$ は、 n 個の要素のうち $l (\ll n)$ 個だけが 1 で残りは 0 であるようなパターンの中から、ランダムに選ばれるものとする。

ここでは、

$$w_{ij} = \frac{1}{l} \sum_{\mu=1}^m \left(s_i^\mu - \frac{1}{n} \right) \left(s_j^\mu - \frac{1}{n} \right) \quad (4.10)$$

で与えられる行列 $W = [w_{ij}]$ を利用する。 $l = n/2$ のとき、この W は自己相関連想記憶モデルで用いられる荷重行列と等価である。但し、本モデルの場合、活動度に応じて全体に抑制を加えることが必要なので、ユニット間の結合荷重を

$$w_{ij}^+ = w_{ij} - \frac{\alpha}{l} \quad (4.11)$$

$$w_{ij}^- = w_{ij} \quad (4.12)$$

とする。ここで、 α は正の定数であって、一様な相互抑制の大きさを表している。

記憶したパターンを想起する際は、適当な想起入力 $P = (p_1, \dots, p_n)$ を

$$z_i = kp_i + z_0 \quad (4.13)$$

という形で入力する (k は想起入力の大きさ、 z_0 は全体の刺激レベルを表す)。想起入力は、十分な時間 (時定数 τ の数倍) 与え続ける必要がある。このとき、 P が記憶パターンの一つ (S^1 とする) に十分近ければ、系の状態は S^1 をコードするア

トラクタ(後述)に引き込まれ、 k が0になった後もその状態が保持されるものと考えられる。

実際に、このモデルについてシミュレーション実験を行い、その動作を調べた。実験は、 $n=1000$, $m=400$, $l=100$ で行った。このとき、一つのユニットは、400個のパターンのうち平均40個をコードすることになる。その他のパラメータは次のように選んだ。

$$\begin{aligned}c &= 50, w_I = 1.0, \lambda = 0.2, \theta = 0.1, \\ \alpha &= 0.2, k = 0.1, z_0 = -0.1\end{aligned}$$

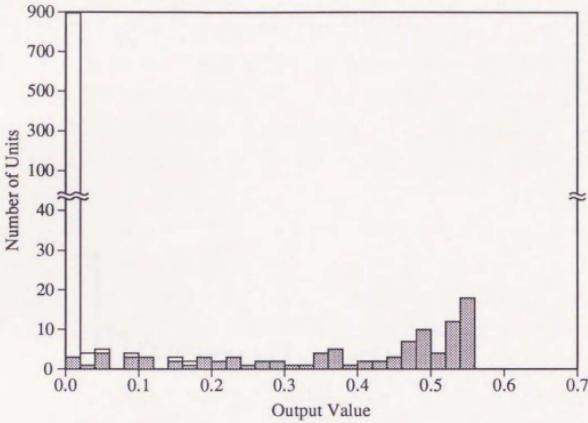
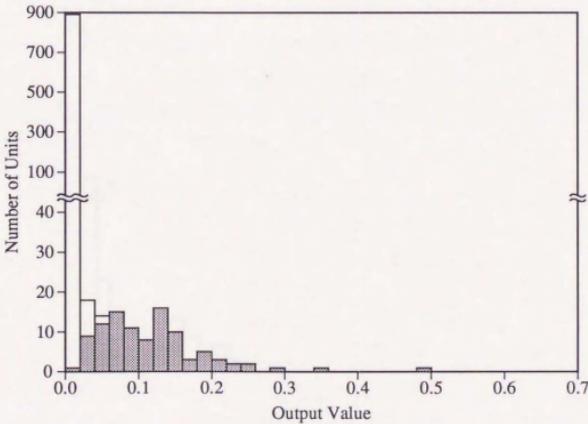
以下、実験結果について定性的に述べることにする。

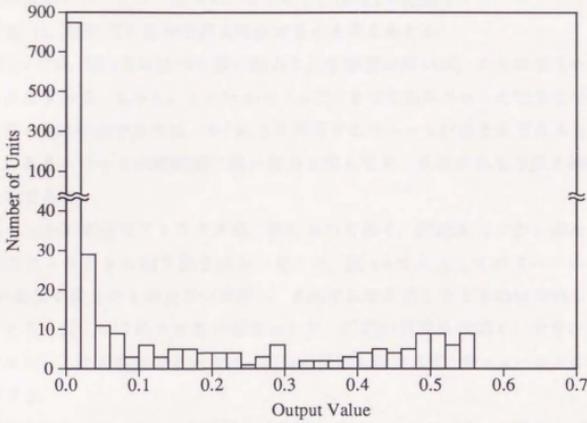
図4.3(a)は、 $P=S^1$ を想起入力として与えた後、十分に時間が経過した時点における各ユニットの出力値 x_i の分布を表すヒストグラムである。灰色の部分 S^1 をコードする($s_i^1=1$ が成り立つ)100個のユニットを表すが、これらが比較的大きな出力を出しているのに対して、その他のユニットはほとんど出力を出さないことがわかる。また、0.1~0.5の中間的な値を出力するユニットも、かなりの数存在している。このような分布は、見慣れた図形を提示したときのTE野ニューロンの興奮強度の分布とよく一致する。

このとき系に多少の外乱を与えてもすぐに元に戻るから、図の状態は、安定な平衡状態である。また、 P が S^1 と相当に異なっても同じ平衡状態に達するので、強いアトラクタになっている。そして、出力パターン $X=(x_1, \dots, x_n)$ は S^1 に最も近く、それ以外の記憶パターンとほとんど相関がないから、このアトラクタは、 S^1 をコードするものだと言ってよいであろう。

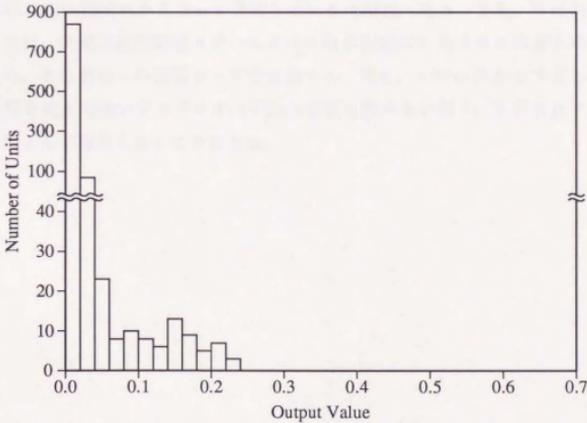
参考のため、このときの抑制細胞の出力 y_i の分布を図4.3(b)に示しておく。ここで注意すべき点は、 C_i^- は C_i^+ をかなり強力に抑制するため、 y_i が非常に大きい数個のユニットでは、 x_i がほとんど0になっているということである。このようなユニットは、全体の活動度が下がると、逆に大きな出力を出すようになる。

これに対して、どの記憶パターンとも全く異なるランダムなパターンを入力すると、十分時間が経過した($t=25\tau$)時点での出力値の分布は、図4.4のようになる。但し、これは完全な平衡状態ではなく、その後も系の状態はゆっくりと変化し続ける。しかし、 10τ 程度の比較的短い時間でこのような状態に達した後は、相当長い時間にわたってほぼ同じパターンが保たれ、多少の外乱を受けてもその

(a) ユニットの出力 x_i の分布(b) 抑制細胞の出力 y_i の分布図4.3 S^1 をコードする平衡状態



(a) ユニットの出力 x_i の分布



(b) 抑制細胞の出力 y_i の分布

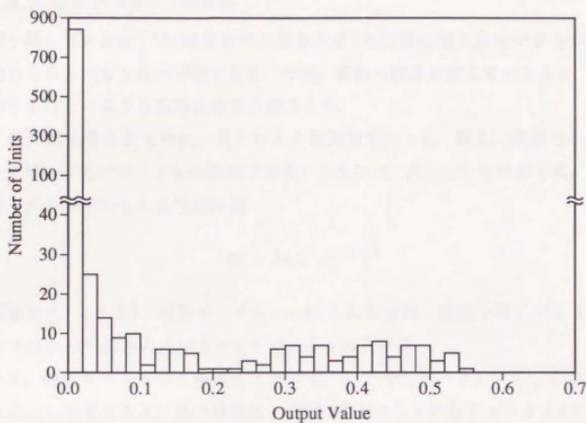
図4.4 ランダムなパターンを入力したときの反応 ($t = 25\tau$)

付近の状態にとどまる。従って、このような状態(不完全なアトラクタと呼ぶことにする)は、短時間の情報保持に利用できると思われる。

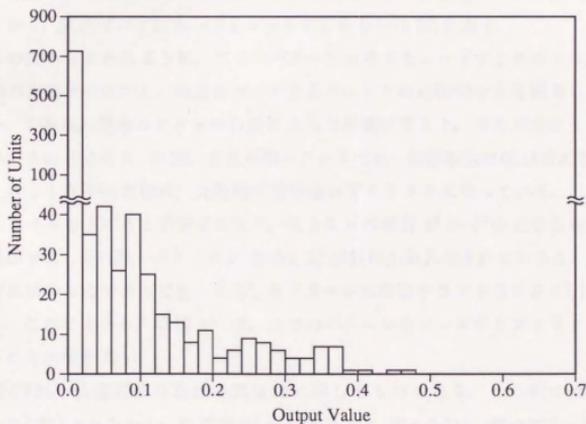
図4.4では、図4.3に比べて弱い出力を出す細胞が多いが、大きな出力を出すものもかなりある。しかし、 $t=5\tau$ から $t=25\tau$ までの各ユニットの出力を平均すると、図4.5(a)に示すように、0.5以上を出力するユニットはほとんどなくなる。これは、あるユニットが瞬間的に強い出力を出しても、それがあまり長く持続しないためである。

こうした不完全なアトラクタは、数が極めて多く、想起入力が少し異なるだけで別のアトラクタに引き込まれる。そこで、図4.4で入力したパターンの1%(10個)の要素を変えたものを5つ用意し、それぞれを入力したときの出力値を平均したところ、図4.5(b)のような分布を示した。反応の再現性が低く、大きな出力を出すユニットほど数が少ない点は、新奇な図形に対するTE野ニューロンの反応と符合する。

なお、シミュレーションの結果を見る限りでは、一つのパターンをコードするアトラクタは一つしかない。但し、 n および m が非常に大きい場合には、複数の平衡状態が狭い領域にかたまって存在している可能性もある。また、不完全なアトラクタは、従来の連想記憶モデルにおける偽の記憶に相当するものだと考えられるから、その数は n の指数オーダーで増加する。逆に、 n や m が小さすぎると、荷重行列を変えて強いアトラクタの引込み領域を狭めない限り、不完全なアトラクタはほとんど存在しないことになる。



(a) 1 回の反応



(b) 5 回の平均

図4.5 ランダムな入力パターンに対する出力 x_i の時間平均の分布

4.3.2 複数のパターンの保持

我々は一度に複数(7個程度までと言われる)の短期記憶を保持できるが, サルの場合もある程度これが可能である。では, 複数の図形を覚えているとき, TE野の力学系はどのような状態にあるのだろうか。

この問題を考えるために, 以下のような実験を行った。但し, 前節の条件下で一度に一つのパターンしか保持できなかったので, 式(4.10)の代わりに, 一般逆行列モデルで用いられる荷重行列

$$W = \Sigma(\Sigma^T \Sigma)^{-1} \Sigma^T \quad (4.14)$$

の要素を w_{ij} とした[†]。パラメータは $c=40$ とした以外, 前節と同じであるが, 刺激レベル z_0 は外部からコントロールできるものとする。

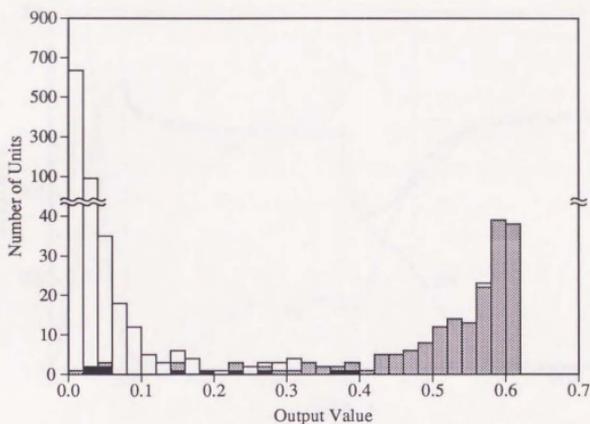
さて, 刺激レベルを少し高くしておいて, 二つの記憶パターン S^1, S^2 の和を想起入力として与えると, 系の状態は, S^1 や S^2 をコードするアトラクタ(それぞれ A^1, A^2 とする)とは別の平衡状態 $A^{1,2}$ に達する。このときの出力値 x_i および y_i の分布を図4.6に示す。黒い部分は S^1 と S^2 の両方をコードするユニット(10個)を表しており, 灰色はいずれか一方をコードするもの(190個)を表す。

この図からわかるように, 二つのパターンの片方をコードするユニットが大きな出力を出すのに対し, 両方をコードするユニットは比較的小きな出力しか出さない。これは, 他のユニットから受ける入力が強すぎると, 出力が小さくなる特性があるからである。実際, これらのユニットでは, 抑制細胞の出力が大きい。

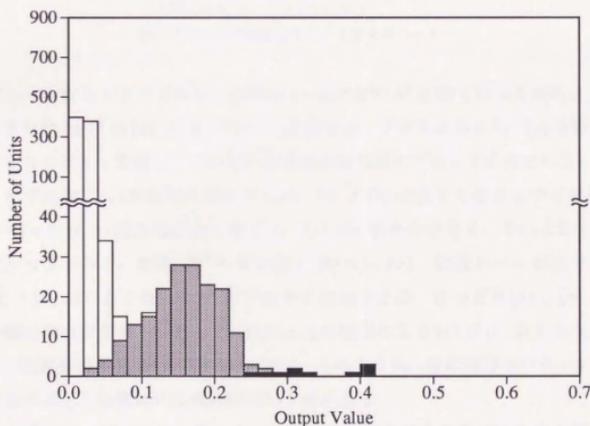
このような平衡状態は, 比較的安定で強いアトラクタになっている。しかし, 刺激レベルを下げると不安定になり, ほとんどの場合 A^1 か A^2 のどちらかに状態が遷移する。その際, パターン S^1 ないし S^2 が外部から入力されていると, それが非常に弱いものであっても, 入力したパターンに対応するアトラクタに引き込まれる。このことから, 状態 $A^{1,2}$ は, 二つのパターンをコードするアトラクタだと言うことができよう。

図4.7は, 状態遷移の過程を具体的に示したものである。この例では, まず $z_0 = 0.1$ で $t=0$ から $t=2\tau$ まで $S^1 + S^2$ を入力し ($k=0.05$), 次いで $t=8\tau$ で

[†] 3.3.2で示したように, w_{ij} をこのように選ぶのが必ずしも最良というわけではない。また, m/n ないし l/n が小さければ, 式(4.10)のままでも複数のパターンを保持できる。



(a) ユニットの出力 x_i の分布



(b) 抑制細胞の出力 y_i の分布

図 4.6 二つの記憶パターンを入力 ($P = S^1 + S^2$) したときの反応

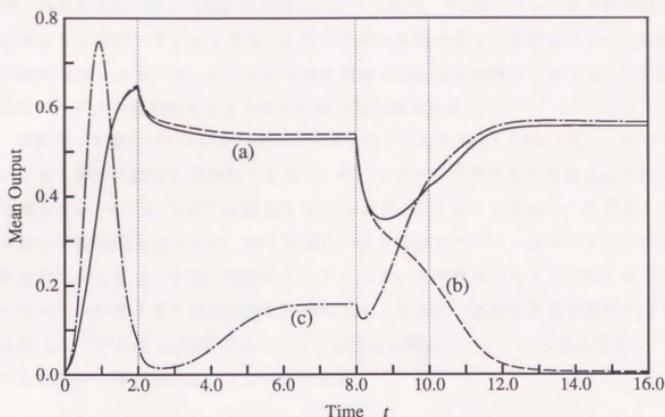


図4.7 各パターンをコードするユニットの平均出力の変化

- (a) S^1 のみをコードするユニット
- (b) S^2 のみをコードするユニット
- (c) S^1, S^2 の共通部分をコードするユニット

刺激レベルを $z_0 = 0$ に下げると同時に $t = 10\tau$ まで S^1 を弱く ($k = 0.005$) 入力した。グラフの (a) と (b) は S^1 と S^2 の一方だけをコードするユニット、(c) は両方をコードするユニットを表し、それぞれの平均の出力値がプロットされている。

まず最初、(c) が急速に増大するが、(a) と (b) が大きくなるとすぐに減少する ($0 < t < 2\tau$)。外部入力がなくなると、(a)、(b) がやや小さく、それに応じて (c) が多少大きくなり、状態 $A^{1,2}$ に落ち着く ($2\tau < t < 8\tau$)。刺激レベルが低くなると、(a) と (b) が小さくなって (c) がすばやく増加するが、このとき (a) と (b) との間に一種の競合が生じる。そして、外部入力の効果により (a) が (b) よりも大きくなると、最終的に状態 A^1 に達する ($t > 8\tau$)。このように、特に両方のパターンをコードするユニットの振舞いには興味深いものがある。

ところで、以上のことをサルでの実験に対応させるならば、二つの異なる図形のどちらにも強く反応するニューロンは、両方の図形を同時に提示したとき、弱い反応しか示さないことになる。これは、このモデルが予言する最も重要な現象

の一つであるが、宮下の最近の実験(私信)によれば、TE野のニューロンは実際にこのような振舞いを示すと言う。従来のモデルを用いると、複数のパターンの保持は実現できるものの、こうした奇妙な現象は決して生じない。従って、この知見は本モデルの妥当性を示すかなり決定的な証拠である。

同様に、3個のパターンを同時に入力すると、そのうちの一つのみをコードするユニットが相対的に大きな出力を出し、すべてをコードするものはほとんど出力を出さない。但し、上述の実験よりも m/n か l/n を小さくしないと、そのような状態は平衡状態にならない。多くの場合、より少数のパターンをコードする状態に変化してしまう。また、同時に入力するパターンの数が増えると、この系だけですべてを保持することは非常に困難になる。しかし、適当な間隔で想起入力を順次与えてやれば(心理学で言うところの短期記憶のリハーサルに相当する)、それらの情報を保ち続けることも可能である。

4.4 考察

このモデルを通して、数々の有益な示唆が得られるが、以下にそのいくつかを挙げる。

1) 細胞数 n を一定にしたとき、記憶パターンがスパースになる(l/n が小さい)ほど、記憶容量が大きくなり、同時に保持できるパターンの数も増える。しかし、想起能力(アトラクタの引込み領域)は小さくなる。両者のバランスを考えたとき、TE野で用いられているような適度にスパースなコーディングは、非常に優れた方法だと言える。

2) 大脳皮質の神経回路の構造はどこでもほぼ同じであり、TE野だけが特殊なわけではない。従って、脳のその他の領域も、このモデルによって実現される力学系の一つになっている可能性が高い。領域によってニューロンの振舞いが異なるのは、情報表現の違いや結合のパラメータの違いが原因のように思われる。

実際、視覚系のニューロンなどで見られる高いパターン選択性——どんなに強い刺激を与えても、それがあつた特定のパターンでなければほとんど反応を示さないが、特定のパターンであれば、非常に弱い刺激に対しても反応する——は、従来の認識系のモデルでは説明しにくい。4.3.1のモデルでパラメータを少し変える(主に刺激レベルを下げる)と、そのような反応の選択性が得られる。

3) ある神経回路網を一つの力学系として見るとき、系全体の状態を考えずに個々の細胞を取り上げて論じることにはあまり意味がない。本モデルのように、構成要素が非単調特性をもつ場合は特にそうである。そして、記憶系や高次の認識系の神経回路に対しては、そうした力学系としての見方が不可欠だと思われる。このことは、単一ニューロンの反応の測定をもとにして研究を進める際、十分に注意しなければならない。

このように、このモデルは脳のモデルとしてかなり一般的な意味をもつものであり、より複雑なシステムをモデル化する際のモジュール^[29]としても、大きな役割を果たすと考えられる。

ところで、本モデルは、1種類の細胞からなる神経回路網の荷重行列に、ある非常に特殊な構造を導入したものとみなすこともできる。しかし、2種類の細胞からなるユニットを考えることには、系の動作を理解しやすくすること以外に、もっと重要な意味がある。それは、学習の問題とかかわっている。

前述のように、ユニット間の結合荷重 w_{ij}^+ と w_{ij}^- との間には、ある程度の相関がなければならない。従って、これらを学習によって構成する場合、 w_{ij}^+ の学習と w_{ij}^- の学習を全く独立に行うことはできない。一方の学習には、他方の荷重の大きさか、少なくともその増減に関する情報が必要である。但し、物理的に離れているシナプスの強度が直接伝わるというのは不自然であるから、このことは、一方の学習に用いられる情報が他方でも利用可能でなければならないことを意味する。例えば、 w_{ij}^+ が Hebb 則に従って変化するならば、 w_{ij}^- の学習には、 C_i^+ の出力 x_i が教師信号として必要である。ただ、具体的な学習則に関しては、様々な可能性があるため、これ以上のことは言えない。

しかし、もし 3.4 で示したような学習がこの回路網で行われるとするならば、「 w_{ij}^+ は相関学習 (Hebb 則) に従って強化され、 w_{ij}^- は x_i がある一定値を超えたとき、 x_j の大きさに応じて強化される」という学習則が、有力なものとして導き出される。このような学習則を用いると、ユニットの入出力特性が非単調なものになり、また入力されているパターンが過去に記憶したパターンに近いときだけ直交学習的な効果が現れるからである。なお、このとき、 C_i^- から C_i^+ への抑制性の結合荷重 w_i を学習によって変化させることにはあまり意味がない。 C_i^- への入力シナプスの荷重 w_{ij}^- を変えることが重要だということを強調しておく。

以上の議論は、大脳皮質のフィードフォワード型抑制回路の可塑性にもそのまま当てはまり、恐らく上述の学習則に近いものが実際に行われているのではないかと思われる。少なくとも、抑制細胞への入力シナプスの学習に、抑制されている細胞の出力がなんらかの形で関与しているはずである。また、もしそうだとすれば、抑制されている細胞がある程度強く興奮したとき、その情報を抑制細胞まで伝達する経路がなくてはならない。伝達手段としては、電気的な信号のほかに化学物質も考えられるが、いずれにせよ抑制回路の局所性(物理的な近さ)が重要な意味をもつ。

これらの予想に関しては生理学的な検証を待ちたいが、モデルの立場からも更に検討が必要である。

結 論

5.1 本研究のまとめ

本研究では、脳の記憶機構の解明を目指して、主に神経回路網の力学という観点から神経回路モデルを構成した。また、その動作を調べることによって、従来のモデルの問題点や脳の記憶回路の力学などに関する考察を行った。以下にその結果をまとめる。

まず、第2章では、自己相関連想記憶モデルの力学について検討し、その想起過程が、想起しようとするパターン以外の記憶パターンの影響を考えることによって直観的に理解できることを示した。また、従来のダイナミクスは、神経回路網がもつ情報を十分に利用しておらず、これを改良することによって、従来のモデルの問題点を解決できる可能性があることを論じた。

次いで、第3章の前半において、具体的なダイナミクスの改良法として部分反転法と非単調ダイナミクス法を提案し、数値実験によってその効果を調べた。その結果、これらの方法は、学習法や回路の基本的構造を変えることなく、モデルの記憶容量を増やし、想起能力を高めることがわかった。このことは、神経回路網の力学を理解するのに有用なだけでなく、従来のダイナミクスが情報処理を行うのに最適ではないことを示している点で、工学的にも重要である。

更に、第3章の後半では、ダイナミクスを改良した連想記憶モデルについて、情報表現および学習の観点から検討を行った。その結果、非単調ダイナミクスを適用した自己相関モデルは、相互にかなり相関があるパターンを記憶できるだけでなく、記憶パターンの構造をうまく利用することによって、高い想起能力を示すことがわかった。また、出力関数の非単調性が、学習の際にも重要な役割を果たすことが示された。これらことは、連想神経回路網において、分散型の情報表現の利点を十分に生かすには、細胞の非単調特性が不可欠であることを示唆している。

以上の結果を踏まえて、第4章では、側頭葉の短期記憶系の力学に関する考察を行い、それが従来の神経回路モデルでは説明できないことを示した。また、非単調な入出力特性をもつユニットを考える必要性を論じ、フィードフォワード型の抑制回路を含む神経回路モデルを構成した。更に、それが側頭葉短期記憶系と同様な力学的性質をもつことを、シミュレーション実験により示した。

ここで主張しておきたい点として、第4章で示したモデルにはかなりの必然性があり、これと本質的に異なる神経回路モデルでは恐らくTE野に見られるような力学系は実現できない、ということがある。これに関する理論的な裏付けはまだ不十分だが、このような必然性は、モデルから単なる可能性や示唆以上のものを引き出すことを可能にする。実際、このモデルは、二つの異なる図形をコードするニューロンの反応が、両者を同時に提示したとき小さくなることを予言し、また局所的な抑制回路において、抑制細胞への入力シナプスの伝達効率が抑制される細胞の出力に応じて変化することを予想する。しかも、前者については、実験によって既にほぼ確認されているのである。

以上のように、本研究は、理論と実験とを結ぶことによって、その双方に新たな視点を提供するものであり、記憶のメカニズムの解明へ向けて、一つの大きな足掛りを築くものだと言える。

5.2 今後の課題と展望

この研究は、ある意味でまだ始まったばかりであり、数多くの課題が残されている。ここですべてを論じることはできないが、そのいくつかを列挙してみよう。

まず、神経回路網の力学に関する理論的解析を進めることが挙げられる。特にダイナミクスを改良したモデルは、偽の記憶の有無など数学的に興味深い問題を多数提供する。本論文で直観に頼っていた部分をきちんとした理論によって裏付けることは、脳のモデル化の研究を進めていく上でも大切である。

また、側頭葉の短期記憶系に関連する問題のうち、第4章で扱わなかったものがいくつかある。

その一つは、力学系に保持されている短期記憶情報をどのように利用するのか(サルの場合で言えば、16秒後に再提示された図形が最初の図形と一致している否かをどうやって判断しているのか)、という問題である。これに答えるために

は、第4章で示したモデルやTE野の力学的性質を更に詳しく調べると共に、脳のその他の部位との関係を考慮する必要がある。

もう一つは、側頭葉の短期記憶力学系がどのようにして自己組織化するのか、ということである。これは、どのようにして長期記憶が形成されるかという非常に大きな問題の一部だとも言える。

TE野の場合、特に興味をもたれるのは、「新奇な図形」がいかにして「見慣れた図形」へと変わっていくのかという点である。これに関連して、提示した順番が近い図形ほど似たパターンにコードされる傾向にある^[28]という興味深い知見が得られているが、このことは、学習の過程で各図形をコードするアトラクタの性質が変わるだけでなく、アトラクタ間の関係、すなわち情報表現も変化することを示唆する。

また、数々の知見から、この過程にはTE野以外の領域、特に海馬が関係すると考えられる^[27]。従って、学習の過程をモデル化するには、単に前述のモデルの学習則を考えるだけでなく、海馬や側頭連合野を中心とする記憶システム全体をモデル化していかなければならない。しかし、海馬一つを取ってみても、その機能すら明らかとは言えず^[3,36]、モデル^[30,31]も十分なものではない。

そのほかにも、本論文では扱っていない重要な問題がある。その一つが、時間的なパターンの記憶である。

有名なPenfieldの実験^[37]によれば、ヒトの側頭連合野を電気刺激すると、古いエピソード記憶が次々と想起されると言う。いったい、そのメカニズムはどうなっているのだろうか。また、メロディーや運動のパターンを覚える際、時間的な情報はどこにどうやって記憶されるのだろうか。このような疑問に対して、満足な回答を与えてくれる理論やモデルは、今のところ存在しない。

更に、記憶は認識や行動とも密接に関係している。また、認識系や行動系の神経回路網が記憶系に近い力学的構造をもっていると信じるだけの理由がある(4.4参照)。従って、本論文で構成したモデルを認識系や行動系に応用することは、より優れた認識や行動のモデルを作るのに必要であると同時に、記憶の研究にとっても重要だと思われる。

以上のような課題は、いずれも簡単に解決できるものではない。しかし、神経回路網の力学をよりよく理解し、うまく利用することが必要だという点は、すべてに共通している。その意味でも、本研究の結果は今後の研究の基礎となるもの

だと言えよう。もちろん、記憶のメカニズムを解明するためには、モデルを発展させるだけでなく、数学的・理論的な研究や生理学的・実験的な研究の発展も不可欠である。

もとより脳の解明は、一つ分野、一つのアプローチによって達成できるものではない。本研究が多くの研究者に対して刺激を与え、記憶の謎を解く糸口となることを期待しつつ、論文の結びとする。

謝 辞

まず、指導教官として、5年間にわたって著者の面倒を見て下さった中野馨助教授に、心より感謝する。研究をはじめとする様々な面において、中野先生から受けた影響は非常に大きい。また、中野研究室の自由でなごやかな雰囲気の中でなければ、著者にとって満足のいく研究をするのは不可能だったであろう。

吉澤修治助教授からは、本研究に関する議論のほか、数々の助言と激励を賜った。日頃お世話になっていることを含めて、ここに深く感謝する。

甘利俊一教授には、神経回路網の理論についてたびたび御教示して頂いた。本研究の着想の多くは、甘利先生の下された示唆がもとになっている。また、鈴木良次教授と石川正俊助教授から、貴重な御意見を頂いた。

以上の先生方をはじめ、御教示や御議論を賜ったすべての方々に、深い感謝の意を表する。

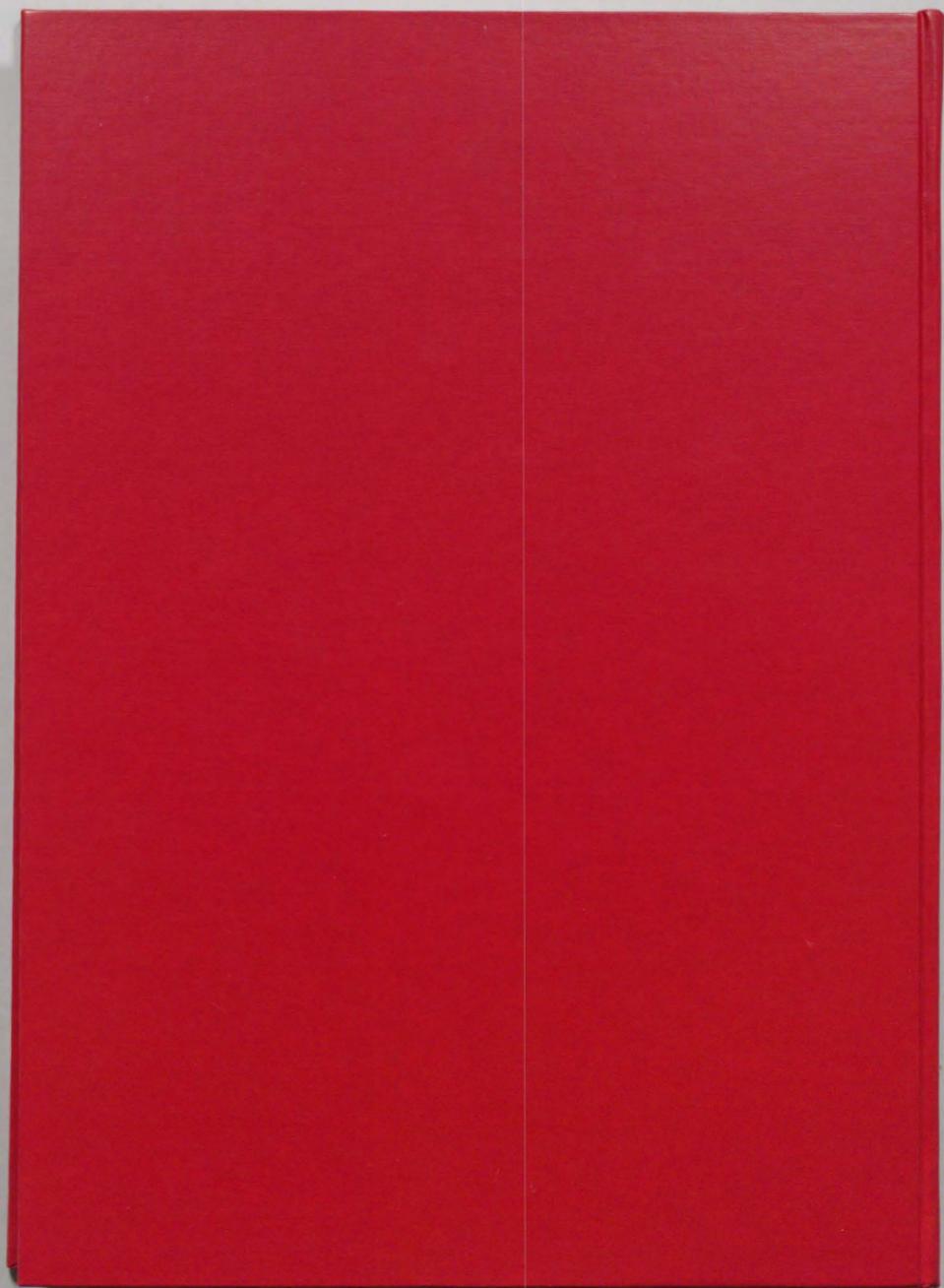
最後に、研究室の関係者全員に対して、御協力を感謝する。それ以外にも、多くの方から暖かい励ましの言葉を頂いた。著者の家族の理解と共に、これらがなければ、ここまで研究を進められなかったに違いない。ここに、心より感謝の意を表したい。

参考文献

- [1] 酒田英夫：記憶は脳のどこにあるか，岩波書店(1987).
- [2] 塚原伸見：脳の可塑性と記憶，紀伊國屋書店(1987).
- [3] 森田昌彦：“海馬の神経回路モデル”，東京大学工学部修士論文(1988).
- [4] Rumelhart, D.E., McClelland, J.L. and PDP Research Group: *Parallel Distributed Processing*, MIT Press, Cambridge (1986).
- [5] Hopfield, J.J.: “Neural networks and physical systems with emergent collective computational abilities”, *Proc. Natl. Acad. Sci. USA*, **79**, 2554-2558 (1982).
- [6] Amit, D.J., Gutfreund, H. and Sompolinsky, H.: “Spin-glass models of neural networks”, *Phys. Rev.*, **A32**, 1007-1018 (1985).
- [7] McCulloch, W.S. and Pitts, W.: “A logical calculus of the ideas immanent in nervous activity”, *Bull. Math. Biophys.*, **5**, 115-133 (1943).
- [8] 中野 馨：“アソシアトロンとその応用——連想記憶装置に関する研究”，電子通信学会インホメーション理論研究会資料，IT 69-27 (1969).
- [9] 上坂吉則，尾関和彦：“連想記憶の二，三の性質”，電子通信学会論文誌(D)，**55-D**, 323-330 (1972).
- [10] 甘利俊一：神経回路網の数理，産業図書(1978).
- [11] Amit, D.J., Gutfreund, H. and Sompolinsky, H.: “Storing infinite numbers of patterns in a spin-glass model of neural networks”, *Phys. Rev. Lett.*, **55**, 1530-1533 (1985).
- [12] McEliece, R.J., Posner, E.C., Rodemich, E.R. and Venkatesh, S.S.: “The capacity of the Hopfield associative memory”, *IEEE Trans. Inform. Theory*, **IT-33**, 461-482 (1987).
- [13] Gardner, E.: “Structure of metastable states in the Hopfield model”, *J. Phys.*, **A19**, L1047-L1052 (1986).
- [14] Kohonen, T.: *Self-Organization and Associative Memory* (2nd ed.), Springer, New York (1988).

- [15] Cottrell, M.: "Stability and attractivity in associative memory networks", *Biol. Cybern.*, **58**, 129-139 (1988).
- [16] Meir, R. and Domany, E.: "Exact solution of a layered neural network model", *Phys. Rev. Lett.*, **59**, 359-362 (1987).
- [17] Meir, R. and Domany, E.: "Layered feed-forward neural network with exactly soluble dynamics", *Phys. Rev.*, **A37**, 608-618 (1987).
- [18] Amari, S. and Maginu, K.: "Statistical neurodynamics of associative memory", *Neural Networks*, **1**, 63-73 (1988).
- [19] Kosko, B.: "Bidirectional associative memories", *IEEE Trans. Sys., Man, Cybern.*, **SMC-18**, 49-60 (1988).
- [20] Bak, C.S. and Little, M.J.: "Memory capacity of artificial neural networks with high order node connections", *Proc. ICNN'88*, **I**, 207-216 (1988).
- [21] Derrida, B., Gardner, E. and Zippelius, A.: "An exactly solvable asymmetric neural network model", *Europhys. Lett.*, **4**, 167-173 (1987).
- [22] Amari, S.: "Characteristics of sparsely encoded associative memory", *Neural Networks*, **2**, 451-457 (1989).
- [23] 森田昌彦, 吉澤修治, 中野 馨: "自己相関連想記憶の想起過程とその改良", 電子情報通信学会論文誌 (D-II), **J73-D-II**, 232-242 (1990).
- [24] Hopfield, J.J.: "Neurons with graded response have collective computational properties like those of two-state neurons", *Proc. Natl. Acad. Sci. USA*, **81**, 3088-3092 (1984).
- [25] Morita, M., Yoshizawa, S. and Nakano, K.: "Memory of correlated patterns by associative neural networks with improved dynamics", *Proc. INNC-90-Paris*, 868-871 (1990).
- [26] 森田昌彦: "側頭葉短期記憶力学系の神経回路モデル", 電子情報通信学会論文誌 (D-II), **J74-D-II**, 54-63 (1991).
- [27] 宮下保司: "視覚再認記憶のニューロン機構", *神経進歩*, **32**, 553-565 (1988).
- [28] Miyashita, Y.: "Neuronal correlate of visual associative long-term memory in the primate temporal cortex", *Nature*, **335**, 817-820 (1988).
- [29] 甘利俊一, 倉田耕治, 赤穂昭太郎: "短期および長期記憶の神経回路モデル", 電子情報通信学会技術報告, **MBE88-143** (1989).
- [30] 森田昌彦: "連想記憶の海馬モデル", 電子情報通信学会論文誌 (D-II), **J72-D-II**, 279-288 (1989).

- [31] 中野 馨 編: ニューロコンピュータの基礎, コロナ社(1990).
- [32] Kree, R. and Zippelius, A.: "Continuous-time dynamics of asymmetrically diluted neural networks", *Phys. Rev.*, **A36**, 4421-4427 (1987).
- [33] Treves, A. and Amit, D.J.: "Metastable states in asymmetrically diluted Hopfield networks", *J. Phys.*, **A21**, 3155-3169 (1988).
- [34] Crisanti, A. and Sompolinsky, H.: "Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics", *Phys. Rev.*, **A37**, 4865-4878 (1988).
- [35] 浦浜 喜一: "ニューラルネットの局所安定性", 電子情報通信学会論文誌 (D-II), **J72-D-II**, 1599-1600 (1989).
- [36] 宮下保司: "海馬の機能", 科学, **55**, 423-432 (1985).
- [37] Penfield, W.: "Memory mechanisms", *Arch. Neurol. Psychiat.*, **67**, 178-191 (1952).





Kodak Color Control Patches

Blue Cyan Green Yellow Red Magenta White 3/Color Black

Kodak Gray Scale

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19

C **Y** **M**

© Kodak, 2007 TM Kodak