

神経回路モデルの数理的研究

倉田耕治



①

神経回路モデルの数理的研究

倉田耕治

10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

目次

まえがき	
第1章 ランダム対称結合を持つ回路の平衡状態の数	1
§ 1.0 はじめに	1
§ 1.1 宮下の実験	2
§ 1.2 ランダム対称結合を持つ回路の平衡状態の数 文献	3 8
第2章 ボルツマン・マシンの幾何学	11
§ 2.0 はじめに	11
§ 2.1 ボルツマン・マシン	11
2.1.1 ボルツマン・マシンの動作	10
2.1.2 ボルツマン・マシンの学習	15
§ 2.2 ボルツマン・マシンの学習とKullback divergence	17
§ 2.3 ボルツマン・マシンの情報幾何学的構造	19
§ 2.4 重みが限りなく増大する場合	25
§ 2.5 ボルツマン・マシンの一般化	28
2.5.1 連続値型ボルツマン・マシン	28
2.5.2 $\Pi\Sigma$ 素子のボルツマン・マシンとk次のボルツマン・マシン	30
§ 2.6 H素子があるときの幾何学 文献	31 34
第3章 競合的な隠れユニットをもつ三層神経回路網の学習	35
§ 3.0 はじめに	35
§ 3.1 基本競合系	36
§ 3.2 モデルと学習則	41
§ 3.3 隠れユニットの分布	44
§ 3.4 平均自乗誤差を最小にする隠れユニットの分布密度	47
§ 3.5 誤差について	50
§ 3.6 バック・プロパゲーションを用いた場合	51
§ 3.7 モデルの簡単化	56

文献	57
第4章 トポグラフィック・マッピング形成の数理	59
§ 4.0 はじめに	59
§ 4.1 ボルツマン・マシンを応用したトポグラフィック・マッピング の形成モデル	65
4.1.0 ボルツマン・マシンをモデルに應用することの意味	65
4.1.1 ボルツマン神経場	65
4.1.2 1次元から1次元へのトポグラフィック・マッピング	66
4.1.3 2次元から1次元へのトポグラフィック・マッピング	68
§ 4.2 Kohonenのモデルにおけるハイパーコラム形成とコラム形成	70
4.2.1 Kohonenのモデル	70
4.2.2 Kohonenのモデルの簡単化と連続化	72
4.2.3 Kohonenのモデルの単純連続解の安定性	74
4.2.3.1 帯状2次元領域の信号空間から1次元の神経場への トポグラフィック・マッピング	74
4.2.3.2 拡散を含まないモデル	85
4.2.3.3 出力層にアナログ素子を使った場合	86
4.2.3.4 円柱状3次元領域の信号空間から1次元の神経場へ のトポグラフィック・マッピング	90
4.2.3.5 発火パターンが変化するKohonenのモデル	96
4.2.3.5.1 1次元から1次元への場合	96
4.2.3.5.2 2次元から2次元への場合	103
4.2.4 Amariのモデルとの比較	107
4.2.5 Malsburgのモデルとの比較	110
4.2.6 相互結合のないモデル 文献	113 115
あとがき	117
謝辞	118
付録1	119
付録2	121

まえがき

人間の脳は、ニューロン（神経細胞）という比較的単純な素子が多数集まってできている。ひとつひとつのニューロンの特性はだいたい分かっている（と思われている）が、それらがどのようなネットワークをつくって脳の機能を実現しているのか、また、そのネットワークがどのように形成されるのか（発生、学習の問題）は、一部を除いて殆ど分かっておらず、これを解明することは人類に残された最大の謎のひとつである。この謎を攻略するためには生理学のみならず様々な手法が必要であり、数理工学的な手法もそのひとつである。脳内の神経回路網は複雑かつ膨大な数のニューロンを含んでいるので、個々のニューロンの特性から回路網全体の動作を知ることは容易ではない。ここに脳の数理モデルをたて、それを理論的に、あるいはシミュレーションによって、解析する数理的な研究方法の活躍の場があると期待されている。

脳の研究、理解には幾つかのレベルがある。第一に、脳を形作っているニューロンは、その膜の電氣的興奮によって信号処理をしているのであるから、神経膜やシナプスの働きに着目し、その分子化学的メカニズムを明らかにしようとする研究分野がある。すなわち分子のレベルとニューロンのレベルの間をこなごなとする研究である。ここでは、解明されるべき現象は「ニューロンの興奮」や「シナプスの増強」であり、そのために使われる言葉は「リセプタ」や「イオン・チャンネル」などである。この分野には、遺伝学における分子生物学の大成功に勇気づけられた数多くの分子生物学者、生化学者が参入しており、現在も盛んに研究が続けられている。

第二に、ニューロンやシナプスの働きは事実として受け入れ、脳内を流れる情報の処理様式を明らかにすることによって脳の機能を説明しようとする研究分野がある。これは個々のニューロンと脳の機能をつなぐ神経回路網のレベルの研究である。

第三は、脳の機能に関する「巨視的」な現象論である。これには自分自身や他人の「こころ」の動きに関する日常的な観察や精神医学、心理学などが含まれる。

この論文は言うまでもなく、第二の、神経回路網のレベルに位置づけられるものであるが、このレベルの研究の目標をはっきりさせるために、これを統計力学になぞらえてみよう。よく知られているように、気体の巨視的な性質に関しては現象論的な熱力学があり、そこで用いられる概念は「温度」、「圧力」

など我々の日常的な感覚に馴染みの深いものである。巨視的な実験によりボイルの法則、シャルルの法則などが得られ、これらは自己完結した体系を成すが、これらの法則の成立する理由をその体系の中に求めることはできなかった。ところが、気体は実は膨大な数の分子の集合体であることが理解されると、微視的な分子の運動に基づいて「温度」、「圧力」などの巨視的な概念が再定義され、これらの間に成立する法則が説明された。これが統計力学である。

これを神経回路網の研究に対応させると次のようになる。すなわち現象論的な熱力学に対応するのが、第三のレベルの脳研究である精神医学や心理学であり、そこでは「温度」、「圧力」と同じように「記憶」、「錯覚」、「夢」、「本能」、「欲望」、「自我」など我々の感覚に馴染みの深い概念が用いられる。我々は、これらの概念に関する観察結果の膨大な蓄積を所有しており、幾つかの法則も知られているが、それらが成立する理由に関するいかなる説明も精神医学や心理学のなかに求めることはできない。そればかりか、例えば、精神医学や心理学は「欲望」という概念に関する客観的な定義を持ち合わせていない。強いていえば「欲望」は「人間をある行動に駆り立てる感情」であるが、こう定義すると、「彼にはこのような欲望があったからこういった行動を起こした」という説明は全くのトートロジーになってしまう。このあたりの事情は、統計力学以前の温度の定義は結局「温度計で計られるもの」であり、つきつめればボイルの法則から逆に定義しているのと同じことであつたと酷似している。

もちろんこれは、精神医学や心理学者の怠慢によるものではなく、精神医学や心理学が本来そういうものだからである。すなわち、これらの学問の目的は「巨視的な」概念によって脳の振舞いを観察し、観察結果をそれらの概念の使える範囲内でできるだけ少数の簡潔な法則にまとめ上げることなのである。そしてこれらの学問は、現象論的な熱力学と同様に有用である。

統計力学になぞらえれば、第二のレベルの脳研究の最終目的は、脳がニューロンという比較的単純な機能を持つ多数の素子の集合体であるという事実に基づいて、精神医学や心理学で用いられる概念を再定義し、それらの間に成り立つ法則を説明することにある。しかし全く新しい概念が用いられ、それらによる新しい法則によって脳の機能が説明される可能性もある。

さて回路網レベルの研究は主に生理学者によって進められているのであるが、彼らの研究は脳の周辺部から出発して徐々に脳の中核に向かっている。周辺部とは、情報の入口と出口、すなわち感覚器とそれに連なる感覚野、及び運動野とそれによって制御される反射系である。また大脳の下部プロセッサと考えられる小脳についてもある程度理解が進んでいる。もちろんこれらの部分につい

ても理解が十分とはいえない。しかし最も深い謎に包まれているのは大脳皮質連合野や視床下部、大脳辺縁系などである。これらの部位は記憶、感情など脳の高次機能に関係していると考えられているが、どの部分がどういった機能に関与しているといった漠然とした知識は得られているものの、その機能がどのような回路によって実現されているかについては殆ど分かっていない。そんななかで大脳辺縁系の海馬およびその周辺の回路については最近急速に研究が進んでいる。

生理学者の用いる研究の手法は近年急速に進歩しているが、基本的には微小電極による比較的小数のニューロンの活動の記録とHRPなどを用いた化学的な手法によるニューロン間の結合関係の推定、および実験動物の脳のある領域を破壊してなげがでなくなるかを見、それからその領域の可っている機能を推定しようとする破壊実験などである。微小電極によるニューロンの活動記録は、そのニューロンに関してかなり詳細な情報を与えてくれるが、活動を記録できるニューロンの数が回路網をかたち作るニューロンの数に較べてあまりに少ないので、群盲象を撫でるといった結果に終わることがある。例えばある行動に伴って必ず発火するニューロンが発見されたとしても、そのことだけではそのニューロンがその行動を引き起こしているのか、それとも、その行動の結果そのニューロンの発火が誘発されるのかは決定できない。

このような局所的な情報を寄せ集めて神経回路網の推定をする際に役立つと期待されるのが、化学的な手法によるニューロン間の結合関係の推定や破壊実験なのであるが、これらの手法で分かることは領域のレベルの分解能しか持っていないので、神経回路の推定に関しては十分な情報を与えてくれないことが多い。例えば、HRPによる結合関係の推定では、脳のどの領域が他のどの領域から投射を受けているかということが分かるだけであるが、神経回路網の推定に必要なのは、どの発火パターンを示すニューロンがどの発火パターンを示すニューロンとつながっているかというような情報なのである。そこで、神経回路網の推定は、十分な根拠のない仮説に、ある程度頼らざるを得ないことになる。

また、実は仮説には次のような積極的な重要性がある。仮に脳内の全てのニューロン間の結合をなんらかの方法で調べることができ、これを超大型超高速のコンピュータでシミュレートして人間の脳と同じ機能を実現できたとしても、これによって何が分かるのであろうか。せいぜい脳の機能にはなんの超自然的実在の関与もないことが証明されるだけであり、これは殆どの脳研究者にとっては証明の必要のない大前提であろう。このことから明らかなように、神経回路網の研究者が目指しているのは単なるニューロンの結合関係の記載ではな

く、それを情報処理という観点から解釈することである。神経回路の推定に用いられる仮説には、通常この解釈が含まれている。すなわち、我々は、脳のこの部分はこのような機能がある、それにはこのような情報処理をおこなっているはずだ、そのためにはこのような結合があるはずだ、というふうな仮説を立てるのである。

工学者や数学者が神経回路網の研究に貢献できるのは、この仮説作りの段階においてである。なぜならば、脳を形作っているニューロンの数はあまりに多いので、ある構造の神経回路網を仮定したとき、それがどのような機能を持つかは、それほど明らかではなく、これを予測するには、理論的な解析やコンピュータによるシミュレーションが必要だからである。

一方、工学の立場からみれば、脳は情報処理という明らかな目的を持ったシステムであり、しかも我々にとって未知の原理に基づいて設計されており、連想記憶、高速高精度のパターン認識、例示による学習など、既存の情報処理装置にはない数々の特長を持っている。脳という情報処理装置の設計原理を解き明かすことができれば、われわれの持つコンピュータも飛躍的な発展を遂げることになるはずである。

神経回路網の理論的あるいは工学的な研究は、1943年のW. S. McCulloch と W. Pittsの形式ニューロンの研究に始まるといってよいだろう。彼らはニューロンをしきい値論理素子としてモデル化し、これを組み合わせることによって任意の論理関数が実現できることを示した。

それから15年後の1958年には、F. Rosenblattが最初の学習神経回路網であるパーセプトロンを提案した。パーセプトロンは脳の大きな特長である例示による学習を実現している。パーセプトロンの中で学習を行っている部分だけを取り出したものを単純パーセプトロンと呼ぶが、パーセプトロンが提案されたのと同じ1962年、H. D. Blockは、学習すべき判別関数が、単純パーセプトロンによって実現可能であるならば、単純パーセプトロンは、どんな初期状態から学習を開始しても、有限回の学習で、その判別関数を実現できることを示した。これが単純パーセプトロンの収束定理である。ところが、1969年、D. Marrは、小脳はパーセプトロンであるという説を発表した。少し遅れて1971年、J. S. Albusも同様の説を発表した。この説は、その後、生理学者によって実験的検証が進められ、現在では最も有力視されている。ここに、神経回路網の理論と生理学における実験的研究の理想的な相互作用をみることができよう。

これらは、神経回路網理論の歴史の初期における重要な結果であるが、その後アメリカでは、1969年に人工知能の大家であるM. MinskyとS. Papertによって書かれた"Perceptron: An Introduction to Computational Geometry"という

本の影響で神経回路網の研究はあまり行われなくなってしまった。彼らはこの本の中で、数多くの重要な問題が、パーセプトロンには学習不可能であることを示した。しかしパーセプトロンはたった1層の学習層しか持っておらず、比較的簡単な問題を非常に短い時間で解くに適した回路である。小脳のモデルになっていることから明らかなように、逐次的な論理の積み上げによって解くような問題に適さないのは当然であり、そのような問題にはまた別の回路を考えなければならぬのである。ところが当時は、パーセプトロンの限界が神経回路網そのものの限界のように考えられてしまった。また、この種の神経回路に、理論的に可能なあらゆる入力に正解を出すことを要求するのは適当でない場合がある。問題によっては90%の正答率でも十分役に立つこともあるのだが、彼らはあくまで100%の正答率が得られるかどうかを問題にしたのである。

アメリカで神経回路網の研究が下火になっているあいだ、神経回路網の研究は主に日本とヨーロッパで続けられ、幾つかの重要な発展があった。パーセプトロンに約15年遅れて1972年、中野馨、T. Kohonen等の数人の研究者によって独立に、連想記憶モデルが提案された。これは現在、海馬との関係が議論されている。また、C. von der Malsburgによるトポグラフィック・マッピングの形成モデルの提案も重要である。こちらはその後、甘利俊一によって数学的な解析が行われている。また、甘利は神経集団ダイナミクス、ランダム神経回路網の統計神経力学、Hebb学習による概念形成、基本競合系や神経場における興奮パターンの解析など、神経回路網理論全体にわたって、ほぼ共通のニューロンモデルを用いた統一的な仕事をしている。

1980年代に入って、神経回路網の理論は新しい時代を迎えた。そのきっかけは、D. E. Rumelhartらによって提案されたバック・プロパゲーションと、J. J. HopfieldやG. E. Hintonらによって神経回路網のダイナミクスに導入された「エネルギー」であろう。多層神経回路網の学習法のひとつであるバック・プロパゲーションの考え方は、実は以前から知られていたのであるが、1980年代に入ってから多くの研究者がこれを様々な問題に応用にはじめた。特にT. J. Sejnowskiは、これをソナーの反射音による岩石と鉄の判別に使って、熟練したソナー員に匹敵する成績をあげた。Hopfieldは神経回路のダイナミクスにエネルギーを定義し、これが単調に減少することを利用して、神経回路を損失関数最小化問題に適用し、NP-完全問題のひとつである巡回セールスマン問題を解いてみせた。Hintonの提案したボルツマン・マシンは情報理論の観点からも非常に興味深いものがあり、また統計力学と神経回路網の類似性を明らかにした。ここにきて前にのべた統計力学におけるスピングラスの理論と神経回路網のアナロジーが単なる哲学以上の意味を持ち始めたのである。

現実問題への応用の可能性が見えてきたことによって、多くの工学者が神経回路網の分野に参入し、また、統計力学と神経回路網の類似性が明らかになったことにより、多くの物理学者が神経回路網の分野に参入してきた。神経回路網の研究者の数があまりに急速にふくれあがったために、過去の研究の歴史が正しく評価され色々なことが「再発見」された時期があったが、その後新しい結果が着実に蓄積されはじめる時期に入ったようである。

前述したように、脳の研究は出入口から進められているのだが、中枢部の研究が難しいのは、そこで用いられている情報表現の形式がよくわかっていないからである。例えば、網膜の出力細胞は、網膜状の一点に到達する光の量に応じて発火する。ここでは情報の表現形式は明らかである。また、筋肉は司令繊維の発火の頻度に応じた強さで収縮する。ここでも情報の表現形式は明らかである。しかし脳の中核に近付けば近づくほど情報の表現形式は曖昧になり、個々のニューロンの発火の意味は不明となる。情報表現の形式が分からなければ情報処理のモデルを立てることは不可能である。

古くからある、おばあさん細胞説とパターン認識説の対立もこれと関わってくる。生理学では1個の細胞の活動を微小電極を用いて調べる研究方法が主流であるため、おばあさん細胞説が根強く、サルの顔細胞の発見など、これを裏付ける研究結果も多い。しかし人間が扱う概念には、幾らでも特殊で高度なものがあつた（「おばあさん」という概念も非常に特殊なもの代表として選ばれたのであろうが、今思うと十分でなかったようである）、これらすべてに別々のニューロンが対応しているとはとても考えられない。

しかしパターン認識説に立つとしても、個々の概念に全く相関のないパターンが割り振られているとしたのでは、情報処理の観点からは、おばあさん細胞説と同様、あまり意味のあるモデルは立てられないだろう。問題にすべきは、外界における概念同士の関係が、情報の表現にどの様に反映されているかである。例えば、我々が二等辺三角形、直角三角形、二等辺直角三角形を思い浮かべたとき、各々に対応するパターンが脳内に生ずるとすれば、それらのパターン同士の関係はどうなっているのだろうか。

本論文には、大きく分けて四つの研究が納められているが、それらはすべて脳内における情報表現の問題と深く関わっている。第1章では、短期記憶のモデルと目されるランダム回路の平衡状態を取り扱う。脳内に固定対称結合のランダム回路があり、この回路に存在する平衡状態が短期記憶を担っているとする説があるが、回路の規模にたいしてどの程度の平衡状態があるのか、また平衡状態の数を最大にする発火率はどのくらいなのかを求めた。

第2章では、ボルツマン・マシンの学習を幾何学的な見地から研究し、Hebb

学習によって隠れユニット群の中に作られる結合の幾何学的な意味を明らかにした。

第3章では競合的な隠れユニットをもつ三層神経回路網の学習モデルを理論的に取り扱う。このモデルでは、さらに直接的に情報の内部表現の問題が取り扱われ、学習によって形成された隠れユニットの情報空間における分布が何によって決定されるかが明らかにされた。

第4章では、脳のあちこちに見られるトポグラフィック・マッピングの自己組織化による形成モデルについて述べる。トポグラフィック・マッピングの形成は、信号空間のトポロジーが学習によって神経場のうえにそのままつとられてしまうという、情報表現に関しては、現在最も興味深い現象の一つである。

また、このモデルをつかえば、以下に述べる二つの情報表現形式の間の変換器を自動的に形成することができる。平面上の位置を表すのに次のような二つの情報表現を考えてみよう。ひとつは2次元の神経場を考え、その上の一つの細胞が発火することによって平面上の位置を表す方法。これを場表現と呼ぼう。もう一つは平面上の位置を座標を使って2個の実数におきかえ、これを2個の細胞の発火率によって表す。これを座標値表現と呼ぼう。場表現のほうは随分無駄の多いやり方のようなのだが、利点も多い。例えば、この神経場にたった一個の出力細胞を付け加えれば、平面上の任意の関数が学習可能となる。

いま、暗闇の中の一個の光点を目で追っている人間を考えると、視覚領域では場表現、眼球運動を司令するニューロンの出力は座標値表現をとっていることになる。従って、脳のどこかでこのふたつの表現形式の間の変換がおこっているはずである。場表現から座標値表現への変換器は誤り訂正学習によって作ることができるが、その逆の座標値表現から場表現への変換器はそれほど簡単にはいかない。トポグラフィック・マッピングの形成モデルを使えばこの変換器を学習によって作ることができる。

第4章では、トポグラフィック・マッピングの形成モデルにおけるハイパー・コラム構造とコラム構造の形成に焦点を当て、この種のモデルとしては最も単純なKohonenのモデルを用いてこれらの構造が形成されるための条件を求めら



## 第1章 ランダム対称結合を持つ回路の平衡状態の数

### §1.0 はじめに

パーセプトロン、アソシエトロン[3]、バック・プロパゲーションなど従来の学習、記憶モデルは、中～長期の記憶に関するものであり、すべて、記憶の実体をシナプス荷重の変化に置くものである。現在、長期記憶の少なくとも一部はシナプス荷重の変化によるものだとする考えは、研究者の間で広く認められているようである。これに対して、短期記憶の正体については、いまだによく分かっていない。フィードバック・ループを持った神経回路の中に保持される興奮状態（平衡状態）であるとする考え方が古くからあるが、一方では、10秒程度でコンダクタンスが大きく変化するシナプスが海馬で見つかっており、短期記憶の正体もまたシナプス荷重の変化であるとする説も有力になってきている。また、次に述べる事実もまた、短期記憶がシナプス荷重の変化によって起こるとする説を有力視させるものである。

シナプスでは、シナプス前細胞からシナプス後細胞へ信号が伝わるのであるが、ここではシナプス前膜とシナプス後膜が、シナプス間隙と呼ばれるごく小さな間隙を挟んで接している。シナプス前膜はシナプス前細胞の軸索の終端部分の細胞膜であり、シナプス後膜はシナプス後細胞の樹状突起からさらに乳頭状に突出した、スパインと呼ばれる微小な突起の先端部の細胞膜である。シナプスにおいて、電気的な信号は一旦化学的な信号に変換される。すなわち、シナプスに到着した電気的信号の強さに応じて何個かのシナプス小胞がシナプス前膜から放出される。シナプス小胞は伝達物質と呼ばれる化学物質を含んでおり、これがシナプス後膜を刺激して、そこに電気的興奮または抑制を引き起こす。

シナプス荷重の変化のメカニズムに関しては様々な説があるが、スパインの形の変化による説もその一つである。スパインの形状の変化は実際に観察されているし、またスパインから筋繊維の収縮に関係する蛋白質が発見されているのもこの説の傍証の一つである。ところで、本当にシナプス・コンダクタンスの変化がこの蛋白質によって起こされるのであれば、それは我々が手足を動かすのと同じくらい素早く起こるはずであり、これは短期記憶のメカニズムとしても十分な速さなのである。

しかしこれらの二つの可能なメカニズムの一方だけが脳内で使われていると

考える必要はなく、実際はシナプス荷重の変化と神経回路のフィードバック・ループに保持される興奮状態の二つが複雑に絡み合っているというのが、現在最も妥当な見方であろう。

本章では、まず宮下が短期記憶の正体に迫るためにおこなった実験と、それに関連した森田のモデルを紹介し、つぎに森田のモデルで最も重要な役割を果たしているランダム対称結合回路の平衡状態について理論的な解析をおこなう。

### § 1.1 宮下の実験 [7,8]

宮下の実験は、遅延見本合わせ課題と呼ばれるもので、サルにある図形を短時間(0.2秒間)見せ、その後16秒間何も見せないで置いて、16秒後にふたたび図形を0.2秒間見せる。二度目に見せる図形は、一度目に見せたものと同じ場合もあれば、違う場合もある。サルに課せられた課題は、この二つが同じか違うかを判定することである。

第一の図形を見ると、サルは図形が消えてからも16秒の間、この情報をなんらかの形で符号化して蓄えて、次に見る図形と一致しているかどうかを判定すると考えられる。宮下は、サルが初めて見るような無意味図形を用いて記憶期間中のIT野におけるニューロンの興奮を調べた。

宮下が用いた図形は、コンピュータを使ってフラクタルを作る要領で生成したものである。こうすれば、サルが今までに見たことがない無意味な図形をいくらでも必要なだけ簡単に生成できる。生成した図形を二つに分け、100個の図形からなる一つの図形群については、これをもちいて何回もゆっくりと予備実験をおこなった。これによりサルは、実験課題を習得するとともに、これらの図形を長期記憶に蓄えるはずである。一方、やはり100個の別の図形からなる図形群があって、これらの図形は本実験の際に初めて用いられた。そして、「使用済み」の馴れた図形と、初めて見る新しい図形が、短期記憶の情報表現の上でどう違うかが調べられた。

実験の結果をまとめると、次のように解釈できる。

- 1) サルが図形を見てからの16秒間、IT野には、ある興奮パターンが保持される。16秒が経過し、次の図形が提示された時点でこの興奮は消える。
- 2) この興奮パターンは、見せた図形によって異なるが、図形の移動、拡大、縮小、回転、色の付いた図形から白黒画像への変換などによって影響を受けない。
- 3) この興奮パターンは、図形のいかなる幾何学的特徴を反映しているようにも見えない。つまり、比較的似通った興奮パターンを引き起こす二つの図形

の間には何の共通性も見いだせない。

- 4) 初めてみる図形に対する興奮パターンと、見馴れた図形に対する興奮パターンの間には質的に大きな差がある。初めて見る図形に対しては、比較的低い頻度で、小さい比率のニューロンが発火する。見馴れた図形に対しては、興奮が少し強くなると共に、約2~3%のニューロンは、極めて強い興奮を保持する。
- 5) 学習時に見せる図形の順序を一定にすると、学習した図形に対応する興奮パターンは、時間的に近接したものほど強い相関を持つ。すなわち、一つの細胞に着目すると、その細胞は、それが最も強く反応するパターンとその前後に見せた数個のパターンに反応する。反応するパターンの数と反応の強さは、前にも後ろにもほぼ同程度である。

この実験において、サルが一枚目の図形を見ている0.2秒間にIT野に現れる興奮パターンは、それにつづく16秒間の興奮パターンとは異なるものであるという。これは、図形を見ることによって神経回路に初期状態が与えられ、そこから数秒の平衡状態へと回路の状態が変化したのだと考えられそうである。このような平衡状態を持つ最も簡単な神経回路モデルは、ランダムで対称な結合を持つ回路である。

ただ単に各結合を独立で平均0の確率分布で定めると、リミット・サイクルは数多く存在するが、回路の平衡状態の個数の期待値は、ニューロンの数にかかわらず1となる。ところが、任意の二つのニューロンの間の結合が対称であるという条件を入れると、各素子の動作を非同期に行う場合は、ポテンシャルが存在し、リミット・サイクルはなくなる。素子の動作を同期させた場合でも、リミット・サイクルの周期は高々2であることが証明できる。よってランダムで対称な結合を持つ回路には、多くの平衡状態が存在すると期待される。

### § 1.2 ランダム対称結合を持つ回路の平衡状態の数

この回路の平衡状態と外界からの入力との間を連想記憶で結び付けるのが、森田のモデル [2]であるが、これによれば、宮下の実験結果の最も奇妙な部分、すなわち、3)比較的似通った興奮パターンを引き起こす二つの図形の間には何の共通性も見いだせないということをうまく説明できる。

森田のモデルでは、外界からの刺激は、適当な前処理のち、海馬及びその周辺に存在する(と仮定する)ランダムで対称な結合を持つ回路における平衡状態と、海馬における連想学習によって対応づけられる。ランダム対称回路の状態は次々に各安定状態を巡って変化しており、各刺激に対しては、たまたま

その刺激が到着したときのランダム対称回路の状態が、いわば「内部コード」として割り当てられる。したがって、似通った刺激に対しても全く無関係なコードが割り当てられることになり、うまく3)を説明できるのである。ランダム対称回路の状態が、ある平衡状態からそれに比較的「近い」つまり相関の高い状態へと変化していると考えれば、宮下の実験の5)もうまく説明できる可能性がある。

森田のモデルの鍵を握るのは、いうまでもなくランダムで対称な結合を持つ回路の平衡状態であるが、それならばこの平衡状態はいったい何個ぐらいあるのだろうか。Tanaka & Edward[9]は、次の±1-モデルに関して平衡状態の数を計算した。

[±1-モデル]

$$x_i(t+1) = \text{sgn}[\sum_j w_{ji} x_j(t)], \quad i=1, \dots, n, \quad (1.1)$$

$$\text{但し,} \quad \text{sgn}[x] = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0, \end{cases}$$

$$w_{ij} = w_{ji} \sim N(0,1), \quad \text{i. i. d.,} \quad i, j=1, \dots, n.$$

結果は、 $n$ 個の神経細胞に対して約 $e^{0.199n}$ 個であった。

彼らの方法は統計物理学的な手法によるもので、非常に複雑であるが、我々は、新たに一つの確率変数を導入することによって計算を簡単化できることを使って、01-モデルの平衡状態の数を計算した。

[01-モデル]

$$x_i(t+1) = 1[\sum_j w_{ji} x_j(t) - \theta], \quad i=1, \dots, n, \quad (1.2)$$

$$\text{但し,} \quad 1[x] = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

$$w_{ij} = w_{ji} \sim N(0,1), \quad \text{i. i. d.,} \quad i, j=1, \dots, n.$$

その刺激が到着したときのランダム対称回路の状態が、いわば「内部コード」として割り当てられる。したがって、似通った刺激に対しても全く無関係なコードが割り当てられることになり、うまく3)を説明できるのである。ランダム対称回路の状態が、ある平衡状態からそれに比較的「近い」つまり相関の高い状態へと変化していると考えれば、宮下の実験の5)もうまく説明できる可能性がある。

森田のモデルの鍵を握るのは、いうまでもなくランダムで対称な結合を持つ回路の平衡状態であるが、それならばこの平衡状態はいったい何個ぐらいあるのだろうか。Tanaka & Edward[9]は、次の±1-モデルに関して平衡状態の数を計算した。

[±1-モデル]

$$x_i(t+1) = \text{sgn}[\sum_j w_{ji} x_j(t)], \quad i=1, \dots, n, \quad (1.1)$$

$$\text{但し,} \quad \text{sgn}[x] = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0, \end{cases}$$

$$w_{ij} = w_{ji} \sim N(0,1), \quad \text{i. i. d.,} \quad i, j=1, \dots, n.$$

結果は、 $n$ 個の神経細胞に対して約 $e^{0.199n}$ 個であった。

彼らの方法は統計物理学的な手法によるもので、非常に複雑であるが、我々は、新たに一つの確率変数を導入することによって計算を簡単化できることを使って、01-モデルの平衡状態の数を計算した。

[01-モデル]

$$x_i(t+1) = 1[\sum_j w_{ji} x_j(t) - \theta], \quad i=1, \dots, n, \quad (1.2)$$

$$\text{但し,} \quad 1[x] = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

$$w_{ij} = w_{ji} \sim N(0,1), \quad \text{i. i. d.,} \quad i, j=1, \dots, n.$$

$$u_i = \begin{cases} \alpha x_i + \beta a, & i=1, 2, \dots, m, \\ \alpha x_i, & i=m+1, m+2, \dots, n, \end{cases} \quad (1.6)$$

$x_i, a \sim N(0,1), i.i.d.$

として、(1.5)を満たすように $\alpha, \beta$ の値を決めると、

$$\alpha = \sqrt{1-1/m}, \quad \beta = 1/\sqrt{m} \quad (1.7)$$

となる。 $\alpha$ と $\beta$ が、この値を持つとき、(1.4)と(1.6)で与えられた $u_i$ の分布は等しくなる。そこで以後は、(1.6)で考えることにする。

状態(1.3)が、平衡状態であるためには、

$$\begin{aligned} u_i &> \theta, & i=1, 2, \dots, m, \\ u_i &< \theta, & i=m+1, m+2, \dots, n \end{aligned}$$

が成立することが必要十分である。この確率は、 $\theta$ によって変化するが、いまは確率が最も大きくなる

$$\theta = -F^{-1}(p), \quad (1.8)$$

の場合を考えることにする。ここに $F$ は標準正規分布の分布関数

$$F(x) \triangleq (1/\sqrt{2\pi}) \int_{-\infty}^x \exp(-t^2/2) dt,$$

である。 $q=p-1$ とおくと、求める確率 $P$ は

$$\begin{aligned} P &= \text{prob}\{(\bigwedge_{i=1}^m u_i > \theta) \wedge (\bigwedge_{i=m+1}^n u_i < \theta)\} \\ &= \text{prob}\{\bigwedge_{i=1}^m \alpha x_i + \beta a > \theta\} q^{n-m}, \end{aligned} \quad (1.9)$$

となる。ここで事象 $u_i < \theta, i=m+1, m+2, \dots, n$ は、(1.6)より、(1.9)のなかで $\wedge$ で結ばれている他の事象と独立であることを使っている。さらに、事象

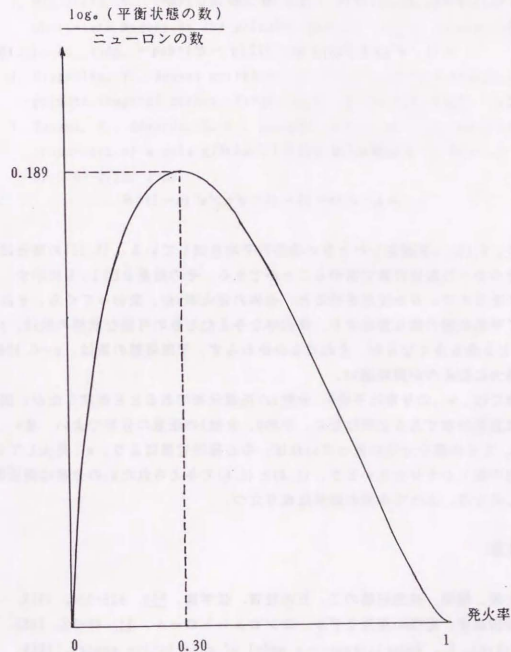


図 1. 1

01-モデルにおける発火率と平衡状態の数の関係

可能な状態の数は、 $p=0.5$ のとき最も多くなるが、それにもかかわらず、平衡状態の数は、 $p=0.30$ 付近で最大になる

$\alpha x_i + \beta a > \theta$ ,  $i = 1, 2, \dots, n$ は,  $a$ を固定すれば互いに独立であるので,

$$P = E_a [\prod_{i=1}^n \text{prob}(x_i + \beta a / \alpha > \theta / \alpha)] q^n \\ = (1/\sqrt{2\pi}) \int \exp\{-a^2/2\} F(\gamma - \theta) da q^n \quad (1.10)$$

ここに,

$$\gamma = \beta a / \alpha + (1 - 1/\alpha)\theta \\ = a/\sqrt{m-1} + \{1 - \sqrt{m}/\sqrt{m-1}\}\theta$$

であり,  $E_a$ は,  $a$ を固定したときの条件付平均を表している。(1.10)の積分は鞍点法をつかった数値計算で求めることができる。その結果を図1.1に示す。

このモデルで,  $\theta$ を変化させると, 全体の発火率 $p$ が, 変わってくる。それにつれて平衡状態の数も変化する。発火率を与えたときの可能な状態の数は,  $p = 0.5$ のとき最も多くなるが, それにもかかわらず, 平衡状態の数は,  $p = 0.30$ 付近で最大になるのが興味深い。

本章では,  $w_{ij}$ の分布は平均0, 分散1の正規分布であると仮定したが, 実は, 分布は正規分布である必要はなく, 平均0, 分散1の任意の分布でよい。各 $w_{ij}$ が, i. i. d. でその様な分布に従っていれば, 中心極限定理により,  $m$ (発火している細胞の数)が十分大きいとき, (1.4)と(1.6)で与えられた $u_i$ の分布は漸近的に等しくなる。よって本章の結果は成り立つ。

#### 文献

1. 上坂, 尾関, 連想記憶の二, 三の性質. 信学論, 55D, 323-330, 1972.
2. 森田昌彦, 記憶の海馬モデル. コンピュータロー, 24, 46-52, 1988.
3. Nakano, K., Associatron—a model of associative memory. IEEE trans. SMC-2, pp.381-388, 1972.
4. Amari, S., Learning patterns and pattern sequences by self organizing nets of analog neuron-like elements. IEEE trans, C-21, pp.1197-1206, 1972.
5. Amari, S., Characteristics of sparsely encoded associative memory. Neural networks, to appear.

6. Amit, D. J. et al., Spin-glass models of neural networks. Phys. Rev. A2, pp.1107-1018, 1985.
7. Miyashita, Y., Chang, H. S., Neural correlate of the pictorial short-term memory in the primate temporal cortex. Nature, 331, pp. 68-70, 1988.
8. Miyashita, Y., Neural correlate of visual long-term memory in the primate temporal cortex. Nature, 335, pp.817-820, 1988.
9. Tanaka, F., Edwards, S. F., Analytic theory of the ground state properties of a spin glass: I. Ising spin glass. J. Phys. F 10, pp.2769-2778, 1980.

## § 2.0 はじめに

Hintonら[3,5]は、統計力学的なアイデアを取り入れた神経回路モデル、ボルツマン・マシンを提案し、さらに、このモデルに対し学習と反学習を組み合わせた新しい学習則を提案した。ボルツマン・マシンは個々の細胞が確率的に発火する自己組織神経回路モデルとして、いくつかの特徴と理論的な見通しの良さがあり、思考能力をもつ知能機械を構成する際の構成要素の候補として看過することのできないものがある。Hintonの提案した学習則によって、ボルツマン・マシンには自分の出力の確率分布を外部から与えられた確率分布に近づけるように学習していくことができる。本章では、ボルツマン・マシンが実現できる確率分布の集合を、あらゆる分布の集合の作る多様体のなかに埋め込まれた部分多様体としてとらえ、これを情報幾何学的手法を用いて解析した。

ボルツマン・マシンのニューロン素子間の結合荷重としきい値には、スピングラスの理論とのアナロジーにおいては、二つのスピンの間の相互作用および外部磁場という意味づけがなされていたが、情報幾何学の立場からの結合荷重としきい値の確率論的な意味をあきらかにする。さらに与えられた分布と、それに対してボルツマン・マシンの学習則が実現する分布の関係を幾何学的に論じ、学習則の幾何学的な意味づけをおこなう。

ここでの考察はアナログ・ボルツマン・マシンや $\Pi\Sigma$ マシンなどの場合にも一般化できる。

## § 2.1 ボルツマン・マシン

## 2.1.1 ボルツマン・マシンの動作

まずHintonらによって提案されたボルツマン・マシンについて簡単に紹介しよう。ボルツマン・マシンは、神経細胞に似た $n$ 個の素子とそれらをつなぐ結合から成り立っている。各々の素子は二つの状態をとることができる。これを $s_i = 0, 1, i=1, \dots, n$ , で表す。 $i$ 番目の素子と $j$ 番目の素子との結合に対して、その強さを表す実数値  $w_{ij} = w_{ji}, i \neq j, i, j=1, \dots, n$ , が定義されていて、その値は学習によって変化する。また、それぞれの素子には、その素子の発火し易さを表す実数  $w_i, i=1, \dots, n$ , が定義されていて、これも学習によって変

化する。場合によっては、このうちのいくつかを0または他の値に固定してもよい。

各素子は、自分への総入力に応じて自分の状態を更新する。i番目の素子の総入力は、

$$u_i = \sum_j w_{ij} s_j + w_i$$

で定義される。つまり神経回路の言葉に直せば、 $w_{ij}$ ,  $i=j$ , はシナプス強度、しきい値は $-w_i$ である。これらのパラメタをまとめて $w$ とかくことにする。従来の神経回路モデルとの重要な違いは、状態の更新が素子間で非同期に行われること、確率的に行われることである。話を簡単にするため、時間は離散的に進み、各時刻に、状態を更新する素子が $n$ 個の素子の中から、全くランダムにそれぞれ $1/n$ づつの確率で選ばれるものとしよう。実はこの仮定は重要ではなく、例えば、ある順序に従って次々と素子を選んで状態を更新することをくり返しても構わない。重要なのは、更新が非同期に行われることと、つぎの確率である。i番目の素子が状態を更新するとき新しい状態が1になる確率は、

$$p(s_i=1) = 1 / (1 + \exp\{-u_i/T\}) \quad (2.1)$$

で与えられる(図2.1)。ここで $T$ は「温度」と呼ばれる正のパラメタである。各素子は、 $T$ が大きいときはほぼランダムに確率 $1/2$ で0, 1の値をとり、 $T$ が0に近いときは、ほぼ決定論的にしきい値論理に従う。

ある時刻におけるボルツマン・マシンの状態は全ての素子の on, off の組み合わせ  $s = (s_i) \in \{0, 1\}^n$  で表される。この $2^n$ 種類の状態に番号  $\alpha = 1, 2, \dots, 2^n$  をつけ、i番目の状態を  $s(\alpha) = (s_i(\alpha))$  と書く。 $\alpha$ 番目の状態に対して「エネルギー」 $En(\alpha)$  が次の様に定義される。

$$En(\alpha) = -\sum_i w_i s_i(\alpha) - \sum_{i,j} w_{ij} s_j(\alpha) s_i(\alpha)$$

定理 2.1 [5]

任意の初期状態から出発して各素子が動作を続けて行くとボルツマン・マシンは $w_{ij}$ によって定まるある確率的な平衡状態に近づく。そのときボルツマン・マシンが状態 $\alpha$ をとる確率 $p(\alpha)$ は次の式で与えられる。

$$p(\alpha) = C \exp\{-En(\alpha)/T\} \quad (2.2)$$

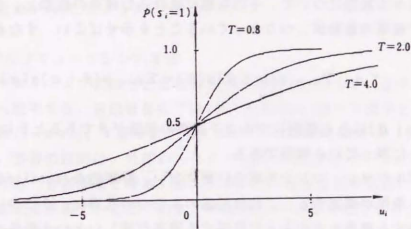


図 2.1 ボルツマン・マシンの素子の発火確率

(2.1)によって与えられる素子の状態変化の確率。入力 $u_i$ が大きいほど新しい状態が1になる確率が大きい。また $T$ が大きいときは、ほぼランダムに確率 $1/2$ で0, 1の値をとり、 $T$ が0に近いときは、ほぼ決定論的にしきい値論理に従う。

証明

式(2.1)で与えられたボルツマン・マシンの状態変化は一つのマルコフ連鎖を定義する。このマルコフ連鎖においては、ハミング距離が1であるような二つの状態間で状態が推移する確率は全て1より小さい正数である。またその他の推移確率は全て0である。よって、このマルコフ連鎖は強連結であり、任意の初期状態から出発して各素子が動作を続けて行くと、初期状態によらないある確率的な平衡状態に近づく。

つぎに、その確率的な平衡状態が(2.2)によって与えられることを示そう。それには、各々の状態について、その状態に流れ込む確率の総和と、その状態から流れ出す確率の総和が、つり合っていることを示せばよい。すなわち、

$$\forall \alpha: \sum_{\beta \neq \alpha} p(\alpha | \beta) p(\beta) = \sum_{\beta \neq \alpha} p(\beta | \alpha) p(\alpha),$$

ここに $p(\alpha | \beta)$ はある時刻にマルコフ連鎖の状態が $\beta$ であるときに、次の時刻に状態が $\alpha$ に移っている確率である。

しかしボルツマン・マシンの場合は更に強い、詳細約合(detailed balance)と呼ばれる条件が成立する。これは任意のふたつの状態 $\alpha, \beta$ について、 $\alpha$ から $\beta$ に推移する確率と $\beta$ から $\alpha$ に推移する確率が等しいという条件である。

$$\forall \alpha: \forall \beta: p(\alpha | \beta) p(\beta) = p(\beta | \alpha) p(\alpha)$$

これが成り立てば、各々の状態について、その状態に流れ込む確率の総和と、その状態から流れ出す確率の総和が、つり合っていることは明らかである。ボルツマン・マシンの場合、 $p(\alpha | \beta)$ は、 $\alpha$ と $\beta$ のハミング距離が2以上のとき0であり、詳細約合の式は成り立っている。 $\alpha$ と $\beta$ のハミング距離が1のときは(2.1)より、

$$p(\alpha | \beta) = 1 / \{n(1 + \exp(-u_i/T))\},$$

$$p(\beta | \alpha) = (1/n) - p(\alpha | \beta),$$

となる。但し、 $\alpha$ と $\beta$ で状態が異なっている素子は $s_i$ で、 $s_i(\alpha) = 1, s_i(\beta) = 0$ であるとする。これと、(2.2)を用い、エネルギーが、 $s_i$ の状態による部分とよらない部分に分けて、

$$En(\alpha) = -u_i \cdot (\alpha) s_i(\alpha) + [s_i(\alpha) \text{によらない部分}]$$

と書けることに注意すれば、詳細約合の式が成立することは簡単に確かめられる。 $\alpha$ と $\beta$ のハミング距離が0であるときは $\alpha = \beta$ であり、このとき詳細約合の式は自明に成立する。  
q. e. d.

式(2.2)をBoltzmann分布と呼ぶ。Cは確率の総和を1にするための正規化定数である。ボルツマン・マシンは全く自由に動作する場合もあるが、いくつかの素子を0または1に固定して動作させることもある。その場合は可能な $\alpha$ に関する確率の総和が1になるようにCをとる。

### 2.1.2 ボルツマン・マシンの学習

ボルツマン・マシンには自己想起型と相互想起型があるが、まず自己想起型の学習から説明する。自己想起型では $n$ 個の全素子は1個のV素子と $m$ 個のH素子に分けられ( $1+m=n$ )、V素子群が入出力ポートの役割を果たす。H素子は無くてもよい。学習の目的は、外部から与えたV素子群上のパターンの出現確率(望ましい分布、分布環境と呼ぶ)を、入力のない自由な動作状態において、できるだけ忠実に再現することである。学習は $w_{ij}$ と $w_i$ を変化させることによって行われ、その学習則は、

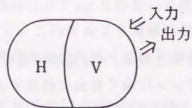
$$\Delta w_{ij} = \varepsilon (p_{ij} - p'_{ij}),$$

$$\Delta w_i = \varepsilon (p_i - p'_i), \quad 0 < \varepsilon < 1$$

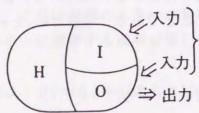
で与えられる。右辺の $p'_{ij}$ と $p'_i$ は、それぞれボルツマン・マシン全体を全く自由に動作させたときの平衡状態における $s_i s_j$ と $s_i$ の期待値である。よって式の上で明記されていないがこれらは $w$ の関数である。 $p_{ij}$ と $p_i$ については以下のとおりである。V素子群の状態に関して外部から与えられた分布環境を $p(r)$ としよう。ここに $r=1, 2, \dots, 2^n$ はV素子群の状態 $s \in \{0, 1\}^n$ に対してつけられた番号である。V素子群を $r$ に固定してH素子群を動作させたときの平衡状態における $s_i s_j$ および $s_i$ の期待値をそれぞれ $p_{ij}(r)$ と $p_i(r)$ とすると、 $p_{ij}$ と $p_i$ は

$$p_{ij} = \sum_r p(r) p_{ij}(r), \quad p_i = \sum_r p(r) p_i(r)$$





自己起型ボルツマン・マシン



相互起型ボルツマン・マシン

図2.2 自己起型と相互起型のボルツマン・マシン

自己起型のボルツマン・マシンは入力ポートの役割を果たすV素子群と、それ以外のH素子群からなる。相互起型のボルツマン・マシンは、I素子群、O素子群、およびH素子群からなる。I素子群とO素子群は、それぞれ入力ポート、出力ポートの役割を果たす。いずれの場合も特殊な場合としてH素子群を欠くものを与えることができる。

で与えられる。よって、i番目とj番目の素子が共にV素子である場合は $p_{ij}$ は $p(\gamma)$ のみによって決定され $w$ にはよらないが、それ以外のばあいは、 $p_{ij}$ は $w$ の関数である。同様に $p_{ji}$ は、i番目の素子がV素子である場合は $w$ にはよらないが、H素子である場合は $w$ の関数である。

相互起型のボルツマン・マシンでは、全素子は $k$ 個のI素子、 $h$ 個のO素子、 $m$ 個のH素子からなる。I素子群とO素子群は、それぞれ入力ポート、出力ポートの役割を果たす。学習則は、やはり(2.3)を用いるが、この場合、 $p_{ij}$ 、 $p'_{ij}$ 、 $p_i$ 、 $p'_i$ は

$$p_{ij} = \sum_{\xi} p(\xi) p(\gamma | \xi) E_{\xi} [s_i s_j; \xi, \gamma], \quad p'_{ij} = \sum_{\xi} p(\xi) E_{\xi} [s_i s_j; \xi],$$

$$p_i = \sum_{\xi} p(\xi) p(\gamma | \xi) E_{\xi} [s_i; \xi, \gamma], \quad p'_i = \sum_{\xi} p(\xi) E_{\xi} [s_i; \xi],$$

で与えられる。ここに $\xi$ と $\lambda$ 、 $\xi = 1, 2, \dots, 2^k$ 、 $\lambda = 1, 2, \dots, 2^h$ はI素子群とO素子群の状態に与えられた番号で、 $p(\xi)$ は外部からの入力でI素子群が $\xi$ に固定される確率、 $p(\gamma | \xi)$ は入力 $\xi$ に対して出力側に状態 $\gamma$ が出現すべき確率、 $E_{\xi}[\cdot; \xi, \gamma]$ 、 $E_{\lambda}[\cdot; \xi]$ は各ポートをそれぞれの状態に固定したとき、 $w$ の与える平衡状態での期待値である。この場合は、 $p(\xi \cap \gamma)$ を分布環境と呼ぶ。シミュレーションでは、V素子群あるいはI素子群とO素子群を分布環境によって固定して動作させて $p_{ij}$ を求め、I素子群だけを固定して動作させて $p'_{ij}$ を求める。 $p_{ij}$ を求めるフェーズを比喩的に覚醒状態、 $p'_{ij}$ を求めるフェーズを睡眠状態と呼んでいる。

言うまでもなく(2.3)の $p_{ij}$ の項はHebb則に対応するが、 $-p'_{ij}$ はその逆で、同時に発火する素子の間の結合を弱めることを意味している。これはCrickやHopfieldの夢のモデル[6]に似ている点でも興味深い。実はこの学習則には次に述べるような理論的根拠がある[5]。

## §2.2 ボルツマン・マシンの学習とKullback divergence

先ず自己起型の場合を考えよう。外から与える分布環境 $p(\gamma)$ と、実際にパラメタ $w$ のボルツマン・マシンによって実現される分布 $p'(\gamma; w)$ との「違い」を表す尺度として

$$G(w) \equiv D(p(\gamma), p'(\gamma; w)) \\ = \sum_{\gamma} p(\gamma) \log(p(\gamma) / p'(\gamma; w)),$$

を考える。これは統計学ではKullback divergenceとよばれ、 $D(p, p') \geq 0$

(等号成立は  $p=p'$  のときのみ) を満たす. Hinton らはこの  $G$  に関して次の定理を得た.

定理 2.2 [5]

$$\begin{aligned} \partial G / \partial w_{ij} &= -(p_{ij} - p'_{ij}) / T \\ \partial G / \partial w_i &= -(p_i - p'_i) / T \end{aligned} \quad (2.4)$$

証明

ここでは,  $w_{ij}$  に関する式にたいする証明を与えておく.  $w_i$  に関する式の証明も同様に証明できる.  $G(w)$  を  $w_{ij}$  で微分すると,

$$\partial G / \partial w_{ij} = -\sum_{\gamma} (p(\gamma) / p'(\gamma)) (\partial p'(\gamma) / \partial w_{ij}) \quad (2.5)$$

を得る. さらに, この式に現れる  $p'$  は (2.2) の  $p$  であるから,

$$\begin{aligned} \partial p'(\gamma) / \partial w_{ij} &= \sum_{\xi} s_i(\gamma, \xi) s_j(\gamma, \xi) A(s(\gamma, \xi)) / B^2 T \\ &\quad - \sum_{\xi} A(s(\gamma, \xi)) \sum_{\alpha} s_i(\gamma, \xi) s_j(\gamma, \xi) A(s(\gamma, \xi)) / B^2 T \end{aligned} \quad (2.6)$$

を得る. ここに

$$\begin{aligned} A(s(\gamma, \xi)) &\equiv \exp\{-E_n(\gamma, \xi) / T\} \\ &= \exp\{-\sum_{i,j} w_{ij} s_i(\gamma, \xi) s_j(\gamma, \xi) + \sum_i w_i s_i(\gamma, \xi)\} / T \\ B &= 1 / C \\ &= \sum_{\alpha} A(s(\gamma, \xi)) \end{aligned}$$

である. (2.6) の右辺の第 1 項と第 2 項はそれぞれ,

$$\begin{aligned} \sum_{\xi} s_i(\gamma, \xi) s_j(\gamma, \xi) A(s(\gamma, \xi)) / B^2 T &= E[s_i s_j | \gamma] p'(\gamma) / T \\ - \sum_{\xi} A(s(\gamma, \xi)) \sum_{\alpha} s_i(\gamma, \xi) s_j(\gamma, \xi) A(s(\gamma, \xi)) / B^2 T &= p'(\gamma) E[s_i s_j] = p'(\gamma) p'_{ij} \end{aligned}$$

となる. これらを, (2.6) に代入し, さらに (2.5) に代入すれば, 証明すべき式 (2.4) を得る. q. e. d

この定理によれば, 学習則 (2.3) は  $G$  を極小にする  $w$  を最急降下法によって求めているわけである. さらに  $H$  素子の無い場合には, 次の定理が成り立つ.

定理 2.3 [5]

$H$  素子が無ければ,  $G$  の極小値は高々 1 個である.

この定理の証明は, 定理 2.5 の証明の過程で得られる. 相互想起型のボルツマン・マシンに対しては  $G$  の定義を

$$G(w) = \sum_{\gamma} p(\xi) D(p(\gamma | \xi), p'(\gamma | \xi; w))$$

とすれば, やはり定理 2.2, 2.3 が成り立つ.

### § 2.3 ボルツマン・マシンの情報幾何学的構造

本節と次の § 2.4 では  $H$  素子なしのボルツマン・マシンについて考える. 以後簡単のため  $T=1$  と仮定する.

ボルツマン・マシンの平衡状態における分布関数は (2.2) で与えられるが, これは

$$p'(X_i) = \exp\{W^i X_i - \psi(W^i)\} \quad (2.7)$$

と書き直すことができる. ここに  $X_i$  と  $W^i$  ( $1 \leq i \leq K$ ,  $K = n(n+1)/2$ ) はそれぞれ,  $X_i(s(\alpha)) = s_i(\alpha) s_j(\alpha)$ ,  $s_i(\alpha)$  および  $W^i = w_{ij}$ ,  $w_i$  ( $1 \leq i < j \leq n$ ) を表し, 二つの添字  $i, j$  をまとめて一つの添字  $l$  として書き直したものである. したがってベクトル  $X = (X_i)$  は  $\{0, 1\}^K$  上の任意の値をとることができるわけではなく,  $s(\alpha)$  の関数として決まるもの, すなわち  $l = (i, j)$  のとき,  $X_l = s_i s_j$  という制約を満たすものだけを考える. 分布 (2.7) も, その様な  $X_l$  についてのみ定義されているとする.  $\psi(W^i)$  は  $\sum p'(X_i) = 1$  とするための正規化定数である. なお式 (2.7) のなかで,  $W^i X_i$  は  $\sum_i W^i X_i$  を意味し, 添字  $l$  について和をとる記号を省略した記法である.

(2.7) のような形の分布を統計学では指数分布族と呼んでいるが, この構造を知るためには情報幾何学的アプローチが極めて有効である [2]. 以下, 本節では

ボルツマン・マシンの情報幾何学的構造について考察を加える。

まず、ボルツマン・マシンのとり得るすべての状態の集合上の確率分布を考え、これを  $p(\alpha)$  と表す。  $\alpha$  は、集合  $\{1, 2, \dots, 2^n\}$  上の確率変数で、状態  $s(\alpha)$  の出現する確率が  $p(\alpha)$  である。  $p(\alpha)$  のうち、全ての  $\alpha$  について  $p(\alpha) > 0$  を満たすものの作る空間を  $S$  とする。このうち、適当な重み  $W$  を持つボルツマン・マシンで実現できる分布  $p(\alpha, w)$  の作る部分空間を  $M$  とする。  $S$  の元  $p(\alpha)$  に対して  $l(\alpha) = \log p(\alpha)$  とすると、  $l(\alpha)$  は  $s_i(\alpha)$  を用いて次の様に展開できる。

$$l(\alpha) = \theta_1^{s_1} s_1 + \theta_2^{s_2} s_2 + \theta_3^{s_3} s_3 + \theta_4^{s_4} s_4 + \dots + \theta_n^{s_n} s_n - \psi(\theta). \quad (2.8)$$

ただし、ここでも総和記号  $\Sigma$  が省略されており、  $\theta$  の冗長性を除くために、右辺を第2項については、  $i < j$ 、第3項  $i < j < k$  を満たすものだけの和を考える。第4項以下も同様である。(2.8)は、  $S$  もまた指数分布族であることを示している。さらに(2.8)の第1項と第2項は(2.7)の  $W^T X_1(\alpha)$  ( $W^T = \theta_2^{s_2}$ ,  $X_1(\alpha) = s_1(\alpha)$ ) に対応しており、  $M$  は3次以上の  $\theta$  を0と置いてできる部分空間であることが分かる。  $w_i, w_j$  ( $i \neq j$ ) は、  $l(\alpha)$  を展開したときの1次と2次の係数に対応する。

$$w_i = \theta_1^{s_1}, \quad w_{ij} = \theta_2^{s_2}, \quad i \neq j.$$

ここで、与えられた確率分布に対する  $\theta$  座標の意味を明らかにしておく必要がある。  $\theta$  のうちで、第  $j$  次の座標  $\theta_j^{s_j}$  は、  $j$  個の確率変数  $s_1, s_2, \dots, s_j$  の同時確率分布に対して、より低次の交互作用に帰着できない  $j$  次の交互作用を表している。とくに、  $w_{ij} = \theta_2^{s_2}$  は、各素子の発火頻度の積では表せない同時興奮の度合、  $\theta_3^{s_3}$  は2次以下の相互作用には分解できない直接の3次の交互作用を表す。これより、ボルツマン・マシンにより、実現出来る確率分布は、各素子間の3次以上の交互作用を含まないものであることがわかる。これは神経回路網が、直接には、2素子間の結合で構成されていることからくる制約といえる。後のべる  $k$  次ボルツマン・マシンでは、ある  $k$  に関して、  $k$  次以下の交互作用が実現できる。

では、  $\theta$  の直接的な意味は何であろうか。とくに、ボルツマン・マシンに關係の深いニューロンのしきい値  $-w_i$  とニューロン間のシナプス荷重  $w_{ij}$  の確率論的な意味は何であろうか。これはつぎの定理で与えられる。いま、第  $i$  番目の

素子に着目し、他のすべての素子が興奮しないという条件のもとで、  $i$  番目の素子が興奮する確率を  $q_i$ 、興奮しない確率を  $q_{0i} \triangleq 1 - q_i$  とする。同じく  $i$  番目と  $j$  番目の二つの素子に着目し、他の素子がすべて興奮しないという条件のもとで、両方が共に興奮する確率を  $q_{11}$ 、  $i$  番目の素子だけが興奮する確率を  $q_{10}$ 、  $j$  番目の素子だけが興奮する確率を  $q_{01}$ 、  $i$  番目も  $j$  番目も共に興奮しない確率を  $q_{00} \triangleq 1 - q_{11} - q_{10} - q_{01}$  とする。

定理 2.4

しきい値およびシナプス荷重は、確率分布と次の関係で結ばれている。

$$w_i = \log(q_i / q_{0i}), \quad w_{ij} = \log(q_{11} q_{00} / q_{10} q_{01}) \quad (2.9)$$

証明

(2.2)より、

$$q_{0i} = 1 / (1 + \exp\{w_i\}), \quad q_i = \exp\{w_i\} / (1 + \exp\{w_i\}).$$

これより、

$$w_i = \log(q_i / q_{0i}),$$

を得る。また、

$$q_{11} = \exp\{w_{ij} + w_i + w_j\} / C', \quad q_{00} = 1 / C',$$

$$q_{10} = \exp\{w_i\} / C', \quad q_{01} = \exp\{w_j\} / C',$$

ここに、

$$C' \triangleq \exp\{w_{ij} + w_i + w_j\} + \exp\{w_i\} + \exp\{w_j\} + 1.$$

これより、

$$w_{ij} = \log(q_{11} q_{00} / q_{10} q_{01}).$$

を得る.

q. e. d.

なお、高次の項  $\theta_1, \dots, \theta_n$  も同様の解釈ができるが、これは省略する.

S内の分布  $p$  を表すパラメタとして  $\theta$  のかわりに、その双対座標  $\eta$  をとることができる.

$$\eta^{1,1,2,\dots,1,j}(p) = \text{Ep}[s_{1,1}s_{1,2}\dots s_{1,j}]$$

ここに  $\text{Ep}[\ ]$  は分布  $p$  の下での期待値である. 特に  $\eta^{1,i} = p^{1,i}$ ,  $\eta^{2,i} = p^{2,i}$  である.

これからは  $2^n - 1$  個の  $\theta$  座標を一列に並べて1個の添字  $i$  を使って  $\theta = (\theta^i)$ ,  $i = 1, 2, \dots, 2^n - 1$ , と書くことにする.  $i = 1, \dots, K$ ,  $K = n(n+1)/2$  までが1次と2次の項  $\theta^{1,i}$ ,  $\theta^{2,i}$  を表している.  $\eta$  座標も同様に  $\eta = (\eta^i)$ ,  $i = 1, \dots, 2^n - 1$ , と表す.  $\eta$  座標と  $\theta$  座標は次のLegendre変換によって結ばれている.

$$\theta^i = \partial \phi(\eta) / \partial \eta^i, \quad \eta^i = \partial \psi(\theta) / \partial \theta^i,$$

$$\psi(\theta) + \phi(\eta) - \theta^i \eta^i = 0.$$

ここに、 $\psi$  は、式(2.8)の正規化定数すなわちキウムラント母関数、 $-\phi$  は、エントロピーである.

さて、情報幾何学に従ってフィッシャー計量  $g_{ij}$  と2種類の接続  $\Gamma^{(1)ijk}$ ,  $\Gamma^{(-1)ijk}$  を導入しよう.

$$g_{ij}(\theta) = E[\partial_i l(\alpha) \partial_j l(\alpha)].$$

$$\Gamma^{(1)ijk}(\theta) = E[\partial_i \partial_j l(\alpha) \partial_k l(\alpha)],$$

$$\Gamma^{(-1)ijk}(\theta) = E[\partial_i \partial_j l(\alpha) + \partial_k l(\alpha) \partial_i \partial_j l(\alpha)].$$

ただし、 $\partial_i \equiv \partial / \partial \theta^i$  である. これらの計量と接続に関して次のことが知られている[2].

### 補題 2.1

$\theta^i(\eta^i)$  のうちの脱つかある値に固定して得られる S の部分空間は  $\Gamma^{(1)ijk}(\Gamma^{(-1)ijk})$  に関して flat である.

### 補題 2.2

$\theta^i$  のうち、 $i \in \Lambda_1 \subset \Lambda = \{1, 2, \dots, 2^n - 1\}$  であるものを固定して得られる部分空間と、 $\eta^i$  のうちで  $i \in \Lambda_2 \subseteq \Lambda$  を固定して得られる部分空間は、 $\Lambda_1 \cap \Lambda_2 = \emptyset$  のとき、計量  $g_{ij}$  の下で直交する.

### 補題 2.3 (ピタゴラスの定理)

S内に3点  $p$ ,  $p'$ ,  $p''$  をとり、 $p$  と  $p'$  -1測地線(-1-flat な曲線),  $p'$  と  $p''$  を1測地線で結んだとき、二つの測地線が  $p'$  で直交するならば、ピタゴラスの定理.

$$D(p, p'') = D(p, p') + D(p', p''), \quad (2.9)$$

が成り立つ(図 2.3).

以上の三つの補題により、外部から与えられた分布にたいして、ボルツマン・マシンの実現する分布を隅に与える次の定理を得る. そのために、Sの分布を次の混存座標系で表そう. Sを表す座標系として  $\eta^{1,i}$ ,  $\eta^{2,i}$  と3次以上の  $\theta$  座標  $\theta_{3^{1,2k}}, \theta_{4^{1,2k}}, \dots, \theta_{n^{1,2^{n-k}}}$  をとると、Sは図 2.4 のように表される.

### 定理 2.5

H素子の無いボルツマン・マシンに確率分布  $p = (\eta^{0,1}, \dots, \eta^{0,K}, \theta_{3^{K+1}}, \dots, \theta_{n^0})$  を分布環境として与えたととき、ボルツマン・マシンの実現する分布は、射影  $p \rightarrow p' = (\eta^{0,1}, \dots, \eta^{0,K}, 0, 0, \dots, 0)$  で与えられる. ただし、 $K = n(n+1)/2$ ,  $N = 2^n - 1$  である.

証明  $p$  と  $p'$  を -1測地線で結ぶと、それは  $\eta^i = \eta^{0,i}$ ,  $i = 1, \dots, K$  で表される部分空間に含まれ、従って、 $\theta^i = 0$ ,  $i = K+1, \dots, N$  で表される M と直交している. M は 1-flat なので、M上に  $p''$  をとると  $p'$  と  $p''$  を結ぶ測地線  $g'$  は Mに含まれ、よって  $g$  と  $g'$  は直交している. 従って補題 2.3 より

$$D(p, p'') \leq D(p, p')$$

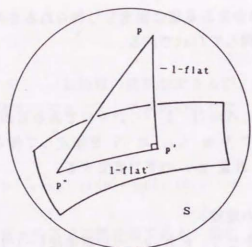


図 2.3 S におけるピタゴラスの定理

S 内に 3 点  $p, p', p''$  をとり,  $p$  と  $p'$  1-測地線,  $p'$  と  $p''$  を 1-測地線で結んだとき, 二つの測地線が  $p'$  で直交するならば, ピタゴラスの定理が成り立つ (図 2.3) .

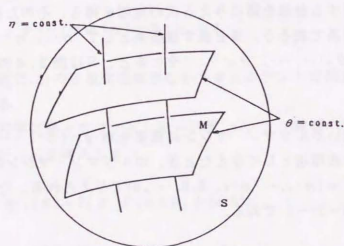


図 2.4 確率分布の空間 S の構造

S は, 3 次以上の  $\theta = \text{const}$  によって定義される 1-flat な多様体の積み重ねに分解され, そのうちで 3 次以上の  $\theta = 0$  に対応するのが M である.

が, すべての M 上の点  $p''$  に対してなりたつ.

q. e. d.

ここで定理 2.3 の証明をしておく. ある  $p$  に対して,  $D(p, p')$  を極小にする  $p'$  が二つ存在したとする. これを  $p'_1, p'_2$  とすれば,  $p$  とこの二つを結ぶ 1-測地線は, M に直交している. よって, (2.9) より

$$D(p, p'_2) = D(p, p'_1) + D(p'_1, p'_2),$$

$$D(p, p'_1) = D(p, p'_2) + D(p'_2, p'_1).$$

この 2 式を辺々加えて,

$$D(p'_1, p'_2) + D(p'_2, p'_1) = 0,$$

を得る. これより,

$$D(p'_1, p'_2) = D(p'_2, p'_1) = 0.$$

よって,  $p'_1 = p'_2$  となり, G の極小値は高々 1 個であることが証明された.

q. e. d.

#### S 2.4 重みが限りなく増大する場合

ボルツマン・マシンを動作させたとき, 平衡状態に達するまでの時間は,  $\|w\|$  が大きいほど長くなる. 従って学習が進むにつれて  $\|w\|$  が無限大に発散する場合には,  $p_{i,j}, p'_{i,j}$  を測定するための動作時間をどんどん長くしていかなければならない. よって, どのとうな環境分布を与えると  $\|w\|$  が発散するのかわかることは有意義である. 前節で述べた M に対して, 座標系として  $(w_1, w_{1,j}) = (\theta_1^i, \theta_2^i, j)$  あるいは  $(p_1, p_{1,j}) = (w_1^i, w_2^i, j)$  をとることができるが,  $(p_1, p_{1,j})$  によって M を  $[0, 1]^k$  に埋めこむことができる. この埋め込みによって, M は  $[0, 1]^k$  中で開集合となる.  $\|w\| \rightarrow \infty$  としたとき, ボルツマン・マシンは M の境界に近づく.

$(p_1, p_{1,j})$  が, ある確率分布 (ボルツマン・マシンで実現できなくてもよい) の 1 次と 2 次の相関となっている領域を  $S'$  とする.

$$S' = \{(p_1, p_{1,j}) \mid$$

$$\exists p(\alpha), p_i = \sum_{\alpha} s_i(\alpha) p(\alpha), p_{i,j} = \sum_{\alpha} s_i(\alpha) s_j(\alpha) p(\alpha)\},$$

$S' \subset [0, 1]^k$  は  $S$  の  $(p_i, p_{ij})$  による像である。

$S$  の内点 (全ての状態  $\alpha$  について,  $p(\alpha) > 0$  であるような分布) は,  $M$  の内点に写る。しかし  $S$  の内点でない点のなかにも  $M$  の内点に写るものがある。

### 例 2.1

$$p(\alpha) = \begin{cases} 0 & \text{状態 } \alpha \text{ において発火している素子の数が偶数であるとき,} \\ 1/2^{n-1} & \text{状態 } \alpha \text{ において発火している素子の数が奇数であるとき,} \end{cases}$$

とすると,  $p(\alpha)$  は,  $S$  の内点ではないが, 任意の  $m$  個 ( $m \leq n-1$ ) に関する  $p(\alpha)$  の周辺分布は一様分布のそれと一致する。よって,  $p(\alpha)$  は  $M$  の内点  $(p_i, p_{ij}) = (0.5, 0.5, \dots, 0.5, 0.25, 0.25, \dots, 0.25)$  に写る。

また,  $M \subset S'$  なので  $S'$  の境界は  $M$  の外側にある。よって  $S'$  の境界上の 2 次相関  $p_{ij}$  を持つ分布環境は  $M$  の内点に射影され得ず, 学習時に  $\|w\|$  が無限大に発散してしまう。これより次の定理を得る。

### 定理 2.6

次の三つの条件の内の少なくとも一つが成り立てば, その分布環境は  $\|w\|$  を無限大におしやる。

- a)  $\exists i: p_i = 1$  or  $0$ .
- b)  $\exists i: \exists j: p_i = p_{ij}$ .
- c) ある素子の集合  $Q$  に対して,  $1 - \sum_{i \in Q} p_i + \sum_{i, j \in Q} p_{ij} = 0$ .

ここで,  $\sum_{i \in Q}$  は,  $Q$  に属している素子  $i, j$  についてのみ和をとることを意味している。

証明  $p_{ij}$  が 2 次の相関であるためには

$$0 \leq p_i \leq 1, \quad p_{ij} \leq p_i$$

が必要である。よって等号が成立すれば,  $p_{ij}$  は  $S'$  の境界上にある。c) も同様であるが c) の式の左辺は「 $Q$  に含まれない素子は何れも発火しない」という事象の確率の上限を与えている。しかし, これらの条件をすべて合わせても  $\|w\|$  が, 無限大に発散するための必要条件にはならない。それを示すのがつぎの例である。

### 例 2.2

$n$  個の  $V$  素子だけからなるボルツマン・マシンを考え, その内の  $k$  個 ( $0 \leq k \leq n$ ) が発火するパターン ( $nC_k$  個ある) を等確率で出現させるような分布は  $\|w\| \rightarrow \infty$  で, ちょうど実現できる。

証明 対称性より  $w_i = u, w_{ij} = v, (i \neq j)$  においてよい。すると  $n$  個中,  $m$  個が発火するパターンのエネルギーは,

$$E_m = -mu - m(m-1)v/2,$$

である。よって,

$$a_m = (m, m(m-1)/2)^T, \quad w = (u, v)^T,$$

とすると, そのパターンの出現確率は,

$$p_m(w) = C(w) \exp\{(a_m, w)\},$$

で与えられる。ここで  $^T$  は転置,  $(\cdot, \cdot)$  は内積を表す。  $p_m(w)/p_k(w) \rightarrow 0, m \neq k$  とすればよいのだが, そのためには,

$$\log(p_m(w)/p_k(w)) = (a_m - a_k, w),$$

であるから, この式の右辺を全ての  $m \neq k$  について負にする  $w_a$  をみつけて,  $w = rw_a, r \rightarrow \infty$  とすればよい。実際,  $w_a = (k-0.5, -1)^T$  とすれば,

$$(a_m - a_k, w) = ((m-k, \frac{1}{2}(m(m-1) - k(k-1)))/2)^T, (k-0.5, -1)^T)^T \\ = -(m-k)^2/2 < 0, \quad m \neq k,$$

を得る。この例で  $n=5$ ,  $k=3$  と置くと、定理 2.6 の条件 c) を満たさないことが分かる。

## § 2.5 ボルツマン・マシンの一般化

### 2.5.1 連続値型ボルツマン・マシン

通常のボルツマン・マシンの素子は  $+1$  と  $-1$  の 2 値しかとらないが、これを、3 個以上の離散値、あるいは連続値をとるように一般化することができる [4]。ここでは、つぎの連続値型ボルツマン・マシンにたいして、§ 2.3 の幾何学的理論が成り立つことを示す。

このボルツマン・マシンでは、素子の状態は、 $[-1, 1]$  の連続値をとり、その状態変化則は、

$$p(s_i) = D \exp\{w_i s_i + \sum_j w_{ij} s_j\}, \quad (2.10)$$

で与えられる。この式は  $s_i$  が状態変化を起こすとき、新しい状態  $s_i$  の確率密度が、他の素子の状態  $s_j$ ,  $j \neq i$ , によって決まる様子を表している。よって、 $D$  は、右辺を  $s_i$  で  $-1$  から  $+1$  まで積分したとき、積分値が 1 となるようにきめられた正規化定数あり、 $s_i$  以外のすべての  $s_j$  の関数である。 $s_i$  が  $-1$  と  $+1$  以外の値をもとるので、 $s_i$  と  $s_i s_i$  は一般に異なる。そのため、 $w_i$  と  $w_{ii}$  は別物である。

この系の状態を  $\alpha \in [-1, 1]^n$  で表すことにする。 $\alpha$  に関してつぎのエネルギーと平衡分布密度が導かれる。

$$\begin{aligned} \text{En}(\alpha) &= -\sum_i w_i s_i(\alpha) - \sum_{i,j} w_{ij} s_i(\alpha) s_j(\alpha), \\ p(\alpha) &= C \exp\{-\text{En}(\alpha)\}, \end{aligned} \quad (2.11)$$

この分布が平衡分布であることは、つぎのようにして確かめることができる。ボルツマン・マシンの状態が (2.11) の分布に従っていたとする。ここで  $s_i$  が (2.10) に従って状態を変化させたとする。このとき状態変化の後の確率密度  $p_i(\alpha)$  を求めると、

$$p_i(\alpha) = p(s_i | \alpha_i) p(\alpha_i), \quad (2.12)$$

となる。ここに  $\alpha_i$  は、 $i$  番目の素子を除いたボルツマン・マシンの状態を表し、 $p(\alpha_i)$  は、状態変化前の  $\alpha_i$  の分布 ( $\alpha_i$  は変化しないので、これは状態変化後の  $\alpha_i$  の分布でもある)、 $p(s_i | \alpha_i)$  は、 $\alpha_i$  によって決められる新しい  $s_i$  の分布である。よって、(2.10) より、

$$p(s_i | \alpha_i) = p(s_i) = D(\alpha_i) \exp\{w_i s_i + \sum_j w_{ij} s_j\}, \quad (2.13)$$

$$D(\alpha_i) = 1 / \int_{-1}^1 \exp\{w_i s_i + w_{ii} s_i^2 + \sum_{j \neq i} w_{ij} s_j\} p(\alpha_i) ds_i.$$

また、(2.11) より、

$$p(\alpha_i) = C_i \int_{-1}^1 \exp\{-\text{En}(\alpha)\} ds_i.$$

ここで、 $\text{En}(\alpha)$  を  $s_i$  に関係する部分とそうでない部分  $\text{En}(\alpha_i)$  に分けると、

$$\text{En}(\alpha) = -\{w_i s_i(\alpha) + \sum_j w_{ij} s_j(\alpha) s_i(\alpha)\} - \text{En}(\alpha_i). \quad (2.14)$$

これより、

$$\begin{aligned} p(\alpha_i) &= C \exp\{-\text{En}(\alpha_i)\} \int_{-1}^1 \exp\{w_i s_i + w_{ii} s_i^2 + \sum_{j \neq i} w_{ij} s_j\} p(\alpha_i) ds_i \\ &= C \exp\{-\text{En}(\alpha_i)\} / D(\alpha_i) \end{aligned} \quad (2.15)$$

(2.13), (2.15) を (2.12) に代入し、(2.14) を使うと、

$$p_i(\alpha) = p(\alpha),$$

を得る。よって、(2.11) は、平衡分布である。

式 (2.9) については、つぎのようになる。

$$w_i = (1/2) \log(p_i/p_{-i}),$$

$$w_{ii} = \log(\sqrt{(p_i p_{-i})}/p_0), \quad w_{ij} = \log(p_{ij} p_0 / p_i p_{-j}).$$

ここに、 $p_i(p_{-i})$  は、 $i$  番目の素子以外の全ての素子が 0 であるという条件の下で、 $i$  番目の素子が  $1(-1)$  となる条件付確率密度である。また、 $p_{ij}(p_{-ij})$  は、 $i$  番目

とj番目の素子以外の全ての素子が0であるという条件の下で、i番目とj番目の素子が1となる(i番目の素子が0、j番目の素子が1となる)条件付確率密度である。

その他の§2.3と§2.4の議論はすべて成り立つ。

2.5.2  $\Pi\Sigma$ 素子のボルツマン・マシンとk次のボルツマン・マシン  
 $\Pi\Sigma$ 素子とは、入力 $x_1, x_2, \dots, x_n$ にたいして出力、

$$y = f(\sum_{i,j} w_{ij} x_i x_j + \sum_i w_i x_i - \theta),$$

を出す素子である。これを用いて $\Pi\Sigma$ ボルツマン・マシンをつくることができる。つまり、i番目の素子への総入力を、

$$u_i = \sum_{j,k} w_{ijk} w_{jks} s_j s_k + \sum_j w_{ij} s_j + w_i,$$

とするのである。式(2.1)によって状態を更新していくとエネルギー、

$$En(\alpha) = -\sum_{i,j,k} w_{ijk} s_i s_j s_k - \sum_{i,j} w_{ij} s_i s_j - \sum_i w_i s_i(\alpha),$$

に関して式(2.2)が成り立つ。ただしこの式のなかの最初の $\Sigma$ では、 $i < j < k$ を満たすi, j, kの組み合わせについて和をとり、2番目の $\Sigma$ では、 $i < j$ を満たすi, jについて和をとる。

$\Pi\Sigma$ ボルツマン・マシンによって実現される分布のなす集合を $M'$ とすれば、 $M'$ は式(2.8)の4次以上の項を0と置いてできる部分空間である。この点を除けば§2.3の議論は全て成り立つ。

さらに次のように1次からn次(nは素子の数)までの任意のk次のボルツマン・マシンを定義することができる。i番目の素子への総入力は、

$$u_i = w^1_i + \sum w^2_{ij} s_j + \sum w^3_{ijk} s_j s_k + \dots + \sum w^k_{i_1 i_2 \dots i_{k-1}} s_{i_1} s_{i_2} \dots s_{i_{k-1}},$$

エネルギーは、

$$En(\alpha) = -\sum w^1_i s_i(\alpha) - \sum w^2_{ij} s_i(\alpha) s_j(\alpha)$$

$$- \sum w^3_{ijk} s_i(\alpha) s_j(\alpha) s_k(\alpha) - \dots - \sum w^k_{i_1 i_2 \dots i_{k-1}} s_{i_1}(\alpha) s_{i_2}(\alpha) \dots s_{i_{k-1}}(\alpha),$$

となる。ただし、入力を与える式のなかの $\Sigma$ では、添字 $j_1, j_2, \dots$ は何れもiに等しくなく、 $j_1 < j_2 < \dots$ を満たしているものについて和をとり、エネルギーを与える式のなかの和は、 $i_1 < i_2 < \dots$ を満たしているものについて和をとる。普通のボルツマン・マシンは2次のボルツマン・マシン、 $\Pi\Sigma$ ボルツマン・マシンは、3次のボルツマン・マシンである。k次のボルツマン・マシンによって実現される分布のなす集合を $M_k$ とすれば、 $M_k$ は式(2.8)のk+1次以上の項を0と置いてできる部分空間である。この点を除けば、やはり§2.3の議論は全て成り立つ。

### §2.6 H素子があるときの幾何学

H素子があるときの幾何学は、ふたとおりの考え方が可能である。一つはV素子群の状態のみの確率分布のつくる空間を考える場合、もうひとつは、H素子を含めたボルツマン・マシン全体の状態の確率分布の成す空間を考える場合である。先ず、V素子群の状態のみの確率分布のつくる空間を考える場合について述べる。

V素子群の素子の状態を $x_i$ 、H素子群の素子の状態を $y_j$ で表すことにする。平衡状態における確率分布は次の式で書ける。

$$p(\alpha) = C \exp\{\sum_{i,j} w_{ij} x_i x_j + \sum_i w_i x_i + \sum_{i,j} u_{ij} x_i y_j + \sum_i v_i y_i\}.$$

この式は $y_j$ を固定すると、 $w'_{ij} = w_{ij}$ 、 $w'_i = w_i + \sum_j u_{ij} y_j$ であるようなH素子のないボルツマン・マシンの平衡分布と見ることができる。これを、 $p(\gamma | \xi)$ と書き、この分布の作る空間を $M(\xi)$ と書くことにする。 $\alpha = (\gamma, \xi)$ 、 $\gamma$ はV素子群の状態、 $\xi$ はH素子群の状態をあらわす。

$$p(\gamma | \xi) = C(\xi) \exp\{\sum_{i,j} w_{ij} x_i x_j + \sum_i w_i x_i + \sum_{i,j} u_{ij} x_i y_j\},$$

$$M(\xi) = \{p(\gamma | \xi)\}.$$

個々の $M(\xi)$ は、1-flatである。V素子群の状態の分布は、



$$\begin{aligned}
 p(\tau) &= \sum_{\xi} p(\tau, \xi) \\
 &= \sum_{\xi} (C/C(\xi)) \exp(\sum_{i,j} v_{ij} y_{ij} + \sum_{i,j} v_{ij} y_{ij}) p(\tau | \xi).
 \end{aligned}$$

よって、 $p(\tau)$ は、 $M(\xi)$ に含まれる分布を重みをつけて足し合わせたものであることが分かる。すなわち、 $p(\tau)$ のつくる空間 $M$ は $\{M(\xi)\}$ の-1測地線による凸包の部分集合である。これは、もはや1-flatではなく、与えられた分布からの1-測地線による射影は一意には決まらない。

つぎに、H素子を含めたボルツマン・マシン全体の状態の確率分布の成す空間を考えてみよう。この場合は、H素子がない場合と同様ボルツマン・マシンの実現する平衡分布の集合 $M$ は、1-flatな部分空間となる。しかし、外から与えられる分布はV素子だけに関するものなので、空間内の1点を指定することができず、つぎのような確率分布の集合 $T(p(\tau))$ を指定することになる。

$$\begin{aligned}
 T(p(\tau)) &= \{p(\tau, \xi) \mid \sum_{\xi} p(\tau, \xi) = p(\tau)\} \\
 &= \{p(\tau, \xi) \mid \exists p(\xi | \tau): p(\tau, \xi) = p(\tau)p(\xi | \tau)\}.
 \end{aligned}$$

$p(\tau, \xi)$ と $p'(\tau, \xi)$ のKullback divergenceを計算すると、

$$\begin{aligned}
 D(p(\tau, \xi), p'(\tau, \xi)) &= \sum_{\tau, \xi} p(\tau, \xi) \log(p(\tau, \xi)/p'(\tau, \xi)) \\
 &= \sum_{\tau, \xi} p(\tau, \xi) \log p(\tau, \xi) - \sum_{\tau, \xi} p(\tau, \xi) \log p'(\tau, \xi) \\
 &= \sum_{\tau, \xi} p(\xi | \tau) p(\tau) \log p(\xi | \tau) p(\tau) \\
 &\quad - \sum_{\tau, \xi} p(\xi | \tau) p(\tau) \log p'(\xi | \tau) p'(\tau) \\
 &= \sum_{\tau} p(\tau) \log p(\tau) + \sum_{\tau} p(\tau) \sum_{\xi} p(\xi | \tau) \log p(\xi | \tau) \\
 &\quad - \sum_{\tau} p(\tau) \log p'(\tau) - \sum_{\tau} p(\tau) \sum_{\xi} p(\xi | \tau) \log p'(\xi | \tau) \\
 &= D(p(\tau), p'(\tau)) + \sum_{\tau} p(\tau) D(p(\xi | \tau), p'(\xi | \tau)).
 \end{aligned}$$

ところが、上式中の $p(\xi | \tau)$ は、実際には与えられないので、 $D(p(\tau, \xi), p'(\tau, \xi))$ を最小にする $p'(\tau, \xi)$ と $p(\xi | \tau)$ を求めると、

$$p(\xi | \tau) = p'(\xi | \tau).$$

と、 $D(p(\tau), p'(\tau))$ を最小にする $p'(\tau)$ を得ることになる。そこで、 $T(p(\tau))$ と $M$ を結ぶ1-測地線のうちで最短のものを求めれば、その $M$ 上の足が $p(\tau)$ を最もよく近似する $p'(\tau)$ となる。

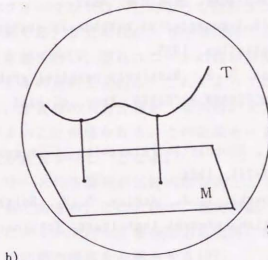
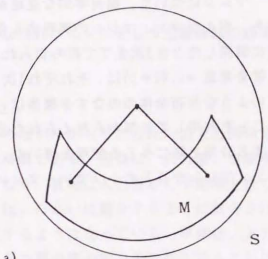


図2.5 H素子があるボルツマン・マシンの実現する分布。

a)はV素子群の状態のみの確率分布のつくる空間を考える場合、b)はボルツマン・マシン全体の状態の確率分布の成す空間を考える場合の図である。a)では、ボルツマン・マシンの実現する分布のなす空間 $M$ は1-flatでない。b)では、 $M$ は1-flatであるが、学習すべき分布が空間内の一点を定めない。

本章では、ボルツマン・マシンについて、幾何学的な見地からその性質を明らかにしてきた。すなわち、ボルツマン・マシンの実現する分布は、分布密度関数の対数を $s_i$ の多項式に展開したとき2次までで打ち切られるような分布である。素子のしきい値 $w_{ij}$ と結合荷重 $w_{ij}$  ( $i \neq j$ )は、それぞれ1次と2次の展開係数とみることができる。このような分布全体のなす多様体は、接続 $\Gamma^{(1)}$ に関してflatである。このことを利用して外部から与えられた分布にたいしてボルツマン・マシンが実現する分布を陽に与える定理を導いた。さらにこれらの議論が連続値型ボルツマン・マシンや $\Pi\Sigma$ ボルツマン・マシンでも成り立つことを示した。

#### 文献

1. 甘利 俊一, 神経回路網の数理. 産業図書, 昭52.
2. Amari, S., Differential-geometrical methods in statistics, Springer lecture notes in statistics, 1985.
3. Farman, S. E., Hinton, G. E., Massively parallel architecture for AI: NETL, THISTL, BOLTZMANN MACHINES. Proc. of AAAI, pp.109-113, 1983.
4. Geman, S., Geman, D., Stochastic relaxation of images. IEEE Trans. on PAMI, 6-6 pp.721-741, 1984.
5. Hinton, G. E., Sejnowski, T. J., Ackley, D. H., Boltzmann machines: constraint satisfaction networks that learn. Technical report CMU-CS-84-119, 1984.
6. Hopfield, J. J., Feinstein, D. F., Palmer, R. G., Unlearning has a stabilizing effect in collective memories. Nature 304 pp.158-159, 1983.

### 第3章 競合的な隠れユニットをもつ三層神経回路網の学習

#### §3.0 はじめに

ここで論ずるのは、外から与えられた入出力関係を学習することのできる、ある種の神経回路網である。回路は、入力層、中間層（隠れユニット群）、出力層の三層からなり、情報は入力層から中間層を経て出力層に伝えられる。中間層のユニットは、たがいに競合するように結合され、全体として最大値検出回路として機能するようになっている。学習は、入力層と中間層の間の結合、および中間層と入力層の間の結合を変化させることによっておこなわれる。中間層の層内結合は、変化しない。このモデル自体は新しいものではないが、理論的な研究は少ない。

この回路のパフォーマンスは、どのような隠れユニットが形成されるかにかかっている。本章では、与えられた入力分布に対する、隠れユニットの形成に関する理論的な考察を行い、隠れユニットの数が非常に多い場合の、隠れユニットの重みベクトルの分布を求める。これにより、出力の平均自乗誤差を最小にするためには、学習時の入力分布を、実用時の入力分布とは少し異なったものにしたほうがよいことが導かれる。この結果を一言で述べれば、“むずかしいことは、何度も練習すべし”となる。

フィードフォワードの多層神経回路の学習則としては、バック・プロパゲーション[10]が有名であるが、この方法は、現実の脳のモデルとしては、不自然であるし、さらにフィードバック結合のある回路にたいしては、結合荷重の変化を計算するために別の回路を必要とする[9]。

ここでは、与えられた入出力関係を学習することのできる、ある種の三層神経回路網について理論的解析を試みるが、中間層のユニット（隠れユニット）間に競合的なフィードバック結合を仮定し、中間層と出力層の間の結合は誤り訂正学習、入力層と中間層の間の結合はHebb学習によって変化するものとする。隠れユニット間の結合は固定する。

隠れユニット間の結合によるダイナミクスを記述するために、隠れユニットは連続時間モデルで扱う。出力層のユニットには、この必要が無いので、その入出力関係は、単なる関数として表す。入力層に一定の入力が与えられると、隠れユニットの出力は、始めのうちは変化しているが、やがて一定値に収束する。このとき、出力層のユニットの出力も一定値に収束しているが、これを、

与えられた入力にたいする、この回路の出力と考えることにする。

一般に、三層回路では、中間層にどのようなユニットが形成されるかが学習の成否を左右するし、またシミュレーションによって形成される隠れユニットの反応に意味づけが可能な場合もあり[8]、非常に興味深い。ユニットの特性が線形の場合には主成分分析と関係した隠れユニットが形成されることが知られているが[2, 3, 4, 5]、非線形ユニットの場合、それを理論的に予測することは、学習の多安定性のため一般には難しい。

本章で扱うモデルは、それ自体新しいものではないが[6]、筆者は隠れユニットの数が非常に多い極限で、学習の結果、形成される隠れユニットの結合荷重ベクトルの分布を理論的に求めることができた。

本章では、まず、この回路の中間層を成す基本競合系について、簡単に解説し、つぎに学習則について述べ、隠れユニットの学習則がポテンシャルを持つことを示す。そして、隠れユニットの数が非常に多い極限で、このポテンシャルを極小にする荷重ベクトルの分布として、学習の結果得られる荷重ベクトルの分布を導く。さらにこの結果を用いて、出力の平均自乗誤差を最小にするための教育法(学習時の入力の与え方)について考察する。

### §3.1 基本競合系

本モデルの中間層は、 $M$ 個のユニットからなる基本競合系[1]をなすような内部結合をもち、最大値検出器として働く。まずこれについて述べよう。基本競合系の各ユニットは、ポテンシャルと呼ばれる内部状態をもっている。そして、ポテンシャルはひとつの実数で表される。全部で $M$ 個のユニットを考え、そのうちの $i$ 番目のユニットのポテンシャルを $u_i$ と書くことにする。ユニットの出力は、 $1[u_i]$ で与えられる。ここで、 $1[u]$ は、

$$1[u] = \begin{cases} 1 & (u > 0), \\ 0 & (u \leq 0), \end{cases} \quad (3.1)$$

なる関数である。各ユニットは、自分を含む中間層内のすべてのユニットに抑制性の信号を送っている。また、自分自身にはこれに加えて興奮性の信号を送っている。 $u_i$ のダイナミクスは、つぎの式で表されるものとする。

$$\dot{u}_i(t) = -u_i(t) + c_1 1[u_i(t)] - c_2 \sum_{j=1}^M 1[u_j(t)] + s_i, \quad (3.2)$$

$$i = 1, 2, \dots, M.$$

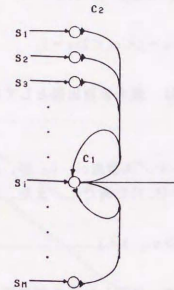


図3.1 基本競合系

図が複雑になり過ぎるので $i$ 番目のユニットから他のユニットへの結合だけを示してあるが、実際はユニット間の結合は完全に対称的であり、全てのユニットから他のユニットへこれと同じだけの結合がある。

ここに $c_1$ と $c_2$ はそれぞれ、興奮性と抑制性のシナプス強度を表す正数、 $s_i$ は $i$ 番目のユニットが入力層から受け取る信号で、

$$-1 \leq s_i \leq 1, \quad (3.3)$$

を満たすものと仮定する。シナプス強度 $c_1, c_2$ が、

$$2 < c_2, c_2 + 1 < c_1 < 2c_2 - 1, \quad (3.4)$$

を満たしているとき、この回路は、最大値検出器としてはたらく。

#### 定理 3.1

基本競合系(3.2)において、シナプス強度 $c_1, c_2$ が、条件(3.4)を満たし、入力 $s_i, i=1, 2, \dots, M$ が、条件(3.3)を満たし、また、ある $s_j$ について、

$$s_j > s_i, \quad i \neq j,$$

が成り立つとする。さらに全ての $u_i$ に等しい正の初期値 $u_0$ を与えて、入力 $s_i$ を固定したとする。このとき $s_i, i=1, 2, \dots, M$ によらずに定まるある時刻までに、 $j$ 番目のユニットだけが、出力1を出し、他の出力は0になる。

#### 証明

次の二つの補題による。

#### 補題 3.1

$t=0$ においてすべての $u_i$ が等しい初期値を持つならば、 $t>0$ の任意の $t$ において $u_i(t)$ の大小関係は $s_i$ のそれに一致する。すなわち

$$\forall t > 0: \quad u_i(t) \geq u_j(t) \Leftrightarrow s_i \geq s_j$$

#### 補題 3.2

定理3.1の条件が成り立つとき、時刻 $t$ で出力1を出しているユニットの数を $n(t)$ とすると、つぎの二つが成立する。

- i)  $n(t) \geq 2$ のとき、すべての非負の $u_i$ は、減少し、 $s_i, i=1, 2, \dots, M$ によらずに定まるある時刻までに $u_i(t) < 2$ となる。

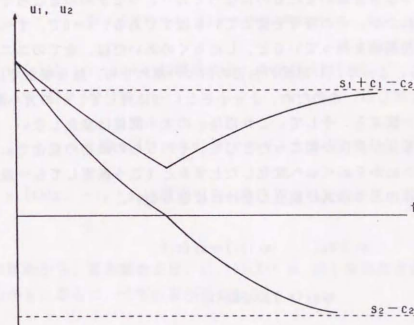


図 3.2 ユニットが2個で、 $s_1 > s_2$ の場合の式(3.2)の解

式(3.2)の解は指数曲線をつなぎ合わせたものになっており、つなぎめのところでは $u_i(t)$ のうちの何れかが符号を変えている。また、十分小さな $t > 0$ に対して、 $u_i$ の大小関係は $s_i$ の大小関係と一致する。そして、これ以後 $u_i$ の大小関係は変化しない。 $u_i > 0$ である素子が2個以上の場合は、すべての非負の $u_i$ は減少し、やがて、そのような素子はただひとつになる。

ii)  $n(t)=1$ のとき、ただ一つの正の $u_i$ は、ある正の値に漸近し、他の $u_i$ は、それぞれ、ある負の値に漸近する。

#### 補題 3.1 の証明

$u_i(t)$ のうちの何れかが正から負へ、あるいは、負から正へと符号を変えないかぎり、式(3.2)の右辺の第2項以下は定数である。よって、式(3.2)の解は指数曲線をつなぎ合わせたものになっており、つなぎめのところでは $u_i(t)$ のうちの何れかが、その符号を変えているはずである。  $t=0$ で、すべての $u_i$ が等しい正の初期値を持っていると、しばらくのあいだは、全てのユニットの出力は1である。よって、(3.2)式の右辺の四つの項のうち、第3項までは、すべての $i$ について等しい。このため、十分小さな $t>0$ に対して、 $u_i$ の大小関係は $s_i$ の大小関係と一致する。そして、これ以後 $u_i$ の大小関係は変化しない。なぜなら、大小関係の変化が何度か起こったとして、そのうちの最初の変化で $u_1$ と $u_2$ の大小関係が $u_1 > u_2$ から $u_1 < u_2$ へ変化したとすると(こう仮定しても一般性は失われない)、次の三つの式が成立しなければならない。

$$\begin{aligned} s_1 > s_2, \quad u_1(t_1) &= u_2(t_1), \\ \frac{d}{dt} u_1(t_1) &\leq \frac{d}{dt} u_2(t_1). \end{aligned} \quad (3.5)$$

ここに、 $t_1$ は、最初の大小関係の変化が起こった時刻である。ところが、これらの式の最初の二つと(3.2)式から、

$$\frac{d}{dt} u_1(t_1) > \frac{d}{dt} u_2(t_1),$$

が導かれる。これは明らかに(3.5)に矛盾するので、 $u_i(t)$ ,  $i=1, \dots, M$ の大小関係は変化しないことがわかる。 q. e. d.

#### 補題 3.2 の証明

$n(t)$ を使って(3.2)式を書き換えると、

$$\frac{d}{dt} u_i(t) = -u_i(t) + c_1 I[u_i(t)] - c_2 n(t) + s_i, \quad i=1, 2, \dots, M,$$

となる。これと、 $s_i$ ,  $c_1$ ,  $c_2$ に関する条件(3.3), (3.4)より

i)  $n(t) \geq 2$ のとき、 $u_i(t) \geq 0$ ならば、

$$\frac{d}{dt} u_i(t) \leq c_1 - 2c_2 + 1 < 0,$$

となり、 $u_i(t)$ は減少する。また、 $s_i$ ,  $i=1, 2, \dots, M$ によらずに定まる時刻  $t = u_2 / (-c_1 + 2c_2 - 1)$ までに $u_i(t) < 2$ となることが分かる。

ii)  $n(t)=1$ のとき、 $u_i(t) > 0$ ならば、

$$\frac{d}{dt} u_i(t) = -u_i(t) + c_1 - c_2 + s_i,$$

となり、 $u_i(t)$ は、 $c_1 - c_2 + s_i$ に漸近する。条件(3.3)(3.4)より、これは正である。また、 $u_i(t) \leq 0$ ならば、

$$\frac{d}{dt} u_i(t) = -u_i(t) - c_2 + s_i,$$

となり、 $u_i(t)$ は、 $-c_2 + s_i$ に漸近する。条件(3.3)(3.4)より、これは負である。 q. e. d.

以上の解析から、基本競合系は、 $s_i$ ,  $i=1 \dots m$ のうちの最大値を検出することがわかる。さらに、つぎの系が導かれる。

#### 系 3.1

シナプス強度 $c_1$ ,  $c_2$ が、条件(3.4)を満たし、 $u_i$ の初期値 $u_i(0)$ ,  $i=1, 2, \dots, M$ のうちのある $u_j(0)$ について、

$$u_i(0) > u_j(0), \quad i \neq j,$$

が成り立つとする。全ての入力 $s_i$ を0に固定し、(3.2)式のダイナミクスを働かせると十分な時間の後、 $j$ 番目のユニットだけが、出力1を出し、他の出力は0になる。

この系は、基本競合系が、多安定な系であることを表している。

#### § 3.2 モデルと学習則

本章で扱う三層回路の入力層には、 $L+1$ 個のユニットがあり、ここに外部か

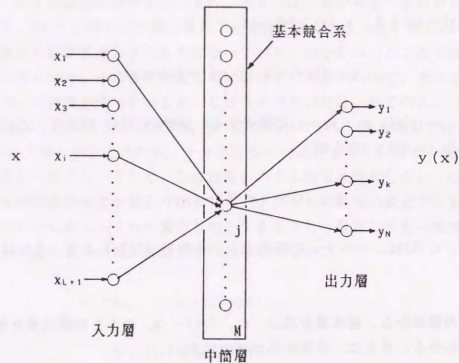


図3.3 競合的な隠れユニットを持つ三層回路網

中間層のユニットは図3.1と同じ相互結合を持ち基本競合系を成している。この部分の結合は固定しており、学習は入力層から中間層への結合と中間層から出力層への結合を変化させることによっておこなわれる。

らの入力  $x = (x_1, x_2, \dots, x_i, x_{i+1})^T$  が与えられる。入力ベクトルのユークリッドノルムは、常に1であると仮定する。つまり、 $x \in S^L$  ( $S^L$  は  $R^{L+1}$  中の単位球面) である。入力層の  $j$  番目のユニットから中間層の  $i$  番目のユニットへのシナプスの強度を、 $w^1_{ij}$  とし、これを  $j$  についてまとめて  $w^1_i = (w^1_{i1}, w^1_{i2}, \dots, w^1_{iL}, w^1_{i(L+1)})^T$  と表す。すなわち中間層のユニットの入力は、

$$s_i = (x, w^1_i), \quad i=1, 2, \dots, M, \quad (3.6)$$

で与えられる。回路への入力同様、シナプスベクトル  $w^1_i$  も  $S^L$  に制限されると仮定する。これにより、式(3.3)の条件が満たされる。

出力層は、 $N$ 個の線形な特性を持ったユニットから成っており、その出力つまり回路の出力は、つぎのように与えられる。

$$y(t) = (y_1(t), y_2(t), y_3(t), \dots, y_N(t))^T, \quad (3.7)$$

$$y_i(t) = \sum_j w^2_{ij} I[u_j(t)], \quad i=1, 2, \dots, N.$$

以上、式(3.1)、(3.2)、(3.6)、(3.7)で、この回路網の入出力関係が完全に記述された。すべての隠れユニットの内部状態  $u_i$  を、等しい正の値にリセットしたのち、入力層をある  $x \in S^L$  に固定してしばらくすると、中間層のユニットのなかで最大の入力を受け取るものだけが1を出力するようになる。そのユニットを  $k(x)$  番目と書くことにすれば、式(3.7)より、その時点で回路の出力は  $y_i(t) = w^2_{i, k(x)}$ 、 $i=1, 2, \dots, N$  となっている。これを入力  $x$  に対するこの回路の出力  $y_i(x)$  であると考えことにする。

入力の空間  $S^L$  を、 $k(x)$  によって、そのなかでは、 $k(x)$  が一定であるような部分集合

$$V_i = \{x \mid k(x) = i\}$$

$$= \{x \mid (x, w^1_j) > (x, w^1_{j'}), \quad j=1, 2, \dots, i-1, i+1, \dots, M\},$$

に ( $V_i$  の境界は除いて) 分割すると、この回路の入出力関係は、

$$y_i = \begin{cases} w_{2,1}^i, & x \in V_1, \\ w_{2,2}^i, & x \in V_2, \\ w_{2,3}^i, & x \in V_3, \\ \vdots & \vdots \\ w_{2,n}^i, & x \in V_n, \end{cases} \quad i=1, 2, \dots, N. \quad (3.8)$$

という階段関数になる。この分割は、 $\{w_{1,1}^i, w_{1,2}^i, \dots, w_{1,n}^i\}$ によって、 $S^L$ 上に生成されるポロノイ・ダイアグラムである。

学習は、隠れユニットを等しい初期値にリセットし、回路網に入力を与え、シナプス強度  $w_{1,j}^i, w_{2,j}^i$  を変化させることを繰り返しておこなわれる。中間層内部の結合は固定する。学習の目的は与えられた出力関数  $f: S^L \rightarrow R^N$ ,  $f(x) = (f_1(x), f_2(x), \dots, f_n(x))^T$  をできるだけ正確に近似することである。中間層と出力層の間の結合は誤り訂正学習によって変化させる。

$$\begin{aligned} w_{2,j}^i &:= w_{2,j}^i + e_2 \{f_j(x) - y_j(x)\} 1[u_i] \\ &= w_{2,j}^i + e_2 \{f_j(x) - w_{2,j}^i\} 1[u_i], \end{aligned} \quad (3.9)$$

$$i=1, 2, \dots, N, \quad j=1, 2, \dots, M,$$

$$0 < e_2 < 1,$$

ここで  $:=$  は更新を表す。

入力層と中間層の間の結合  $w_{1,j}^i$  の学習はHebb学習によるが、 $w_{1,j}^i$  は、 $S^L$ 上に制限されているので、学習を与える式は次のようになる。

$$w_{1,j}^i := (w_{1,j}^i + e_1 1[u_i] x_j) / \|w_{1,j}^i + e_1 1[u_i] x_j\|, \quad (3.10)$$

$$i=1, 2, \dots, M,$$

$$0 < e_1 < 1,$$

### §3.3 隠れユニットの分布

回路網への、入力の頻度分布が、 $p(x)$ ,  $x \in S^L$  であるとき、学習の結果どのような結合が形成されるかを考えよう。バックプロパゲーションと違って、中間層と出力層の間の結合  $w_{2,j}^i$  は中間層のユニットの活動に影響を与えない

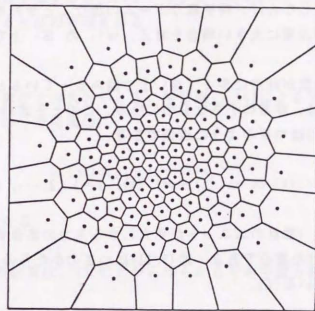


図3.4  $L=2$ の場合のポロノイ・ダイアグラム  
(鈴木教夫博士の御厚意による)

空間をその中に散らばる可算個の母点によって、各母点の周りの「なわばり」(ポロノイ領域)に分割したものがポロノイ・ダイアグラムである。空間内の各点はその点に最も近い母点のポロノイ領域に属する。もともとポロノイ・ダイアグラムはユークリッド空間の中で考えるのが普通であり、ここに示した例もそれであるが、ここでは球面  $S^L$  上のポロノイ・ダイアグラムを考える。しかし、母点の数が非常に多く、各々のポロノイ領域が非常に小さい場合を考えるので、局所的にはポロノイ・ダイアグラムは接平面上のそれで十分近似される。

ので、先ず入力層と中間層の間の結合  $w^1_i$  の形成だけを考える。これが決まれば、 $w^2_j$  の形成は、後で述べるように大変簡単に分析できる。

直観的に言えば、 $M$ 個のベクトル  $w^1_i$  は互いに反発しながら  $p(x)$  の大きい方へ動くので、もし  $S^1$  上にいくつかの  $p(x)$  のピークがあり、隠れユニットの数がそれとほぼ同じならば、椅子取りゲームの様なことがおきる。ここでは隠れユニットの数が非常に大きい場合を考え、 $w^1_i$  の  $S^1$  上での分布  $\rho(x)$  を考えることにする。

隠れユニットの数が非常に多く、 $S^1$  上で混み合っているとき、個々のポロノイ領域  $V_i(t)$  は、非常に小さなものになる。このとき式(3.10)から導かれる平均学習方程式[1]はつぎのように近似できる。

$$\dot{w}^1_i(t) \propto m_i(t) - w^1_i(t), \quad i=1, 2, \dots, M. \quad (3.11)$$

ここに  $m_i(t)$  は、 $i$  番目のユニットが発火しうる入力の集合すなわち  $V_i(t)$  の、 $p(x)$  で重みを付けた重心である。式(3.11)はつぎのかたちのポテンシャルをもつことが分かっている[7]。

$$E(w^1_1, w^1_2, \dots, w^1_M) = \sum_i \int_{V_i} \|x - w^1_i\|^2 p(x) dx, \quad (3.12)$$

それでは  $E$  を最小にする  $w^1_i$  の分布とはなにか、隠れユニットの数が非常に多いとき、局所的には  $p(x)$  は一定で、 $w^1_i$  の配列は局所的には一様な最密構造を実現しているものと考えられる。任意の点  $x_0$  付近のポロノイ領域の直径を  $a(x_0)$  とすると、 $a(x_0) \propto \rho(x_0)^{-1/L}$ 。一つの領域内で、 $p(x)$  は  $p(x_0)$  で十分よく近似される。よって、この付近での一つの領域内での  $\|x - w^1_i\|^2$  の積分  $e(a(x_0))$  は、 $y = (x - w^1_i) / a(x_0)$  と変数変換することにより、

$$\begin{aligned} e(a(x_0)) &= \int_{V_i} \|x - w^1_i\|^2 p(x) dx \\ &\approx \int_{|y| \leq a(x_0)/a(x_0)} \|y\|^2 p(x_0) a(x_0)^L dy \\ &= p(x_0) a(x_0)^{L+2} e(1) \\ &\propto p(x_0) \rho(x_0)^{-1(L+2)/L}. \end{aligned} \quad (3.13)$$

ここに、 $V_i/a(x_0)$  は、 $\{y \mid |a(x_0)y + w^1_i| \in V_i\}$  によって定義される直径1の領域である。また、領域  $V_i$  が十分小さく、その内部での  $p(x)$  の変化が無視できるため、領域内で  $p(x) = p(x_0)$  が成り立つことを使っている。

単位体積あたりのポロノイ領域の数が  $\rho$  だから、

$$\begin{aligned} E &\approx \int_{S^1} \rho(x) e(a(x)) dx \\ &\approx \int_{S^1} p(x) \rho(x)^{-2(L+2)/L} dx \end{aligned} \quad (3.14)$$

となる。これを最小にする  $\rho(x)$  を、 $\int_{S^1} \rho(x) dx = 1$  という条件下で変分法をつかって求めると次の定理を得る。

定理 3.2

中間層の素子の数が十分大きいとき、学習則(3.10)による学習が完了したときの中間層の素子の分布密度は、

$$\rho(x) \propto p(x)^{L/(L+2)}, \quad (3.15)$$

によって近似される。

$w^2_j$  の学習の結果は、(3.9)式から得られる平均学習方程式、

$$\dot{w}^2_j = \int_{V_j} \{f_j(x) - w^2_j\} p(x) dx, \quad (3.16)$$

の平衡状態として、つぎのように得られる。

$$\begin{aligned} w^2_j &= \int_{V_j} f_j(x) p(x) dx / \int_{V_j} p(x) dx \\ &\approx \int_{V_j} \{f_j(m_j) + \nabla f_j(m_j) \cdot (x - m_j)\} p(x) dx / \int_{V_j} p(x) dx \\ &\approx f_j(m_j) \end{aligned} \quad (3.17)$$

ここに  $m_j$  は、 $V_j$  の重心 ( $L=2$  ならば正六角形の中心) である。

### §3.4 平均自乗誤差を最小にする隠れユニットの分布密度

前節ではHebb学習によって形成される隠れユニットの分布を求めたが、これとは別に、この回路で実現される出力関数  $y(x; w^1, w^2)$  と、与えられた関数  $f(x)$  の平均自乗誤差 (MSE) を最小にするためには隠れユニットは、どのように分布すべきかを考えよう。MSEはつぎのように書ける。

$$\begin{aligned} \text{MSE} &= \int_{S^1} \|f(x) - y(x; w^1, w^2)\|^2 p(x) dx \\ &= \sum_{j,j'} \text{MSE}_{j,j'} \end{aligned} \quad (3.18)$$



$$\begin{aligned}
\text{MSE}_{i,j} &= \int_{V_j} (f_i(x) - w_{i,j}^2)^2 p(x) dx \\
&= p(m_j) \int_{V_j} \nabla f_i(m_j) \cdot (x - m_j)^2 dx \\
&= p(m_j) \int_{V_j} \nabla_i \cdot (x - m_j) (x - m_j)^T \nabla_i dx \\
&= p(m_j) \nabla_i C_j \nabla_i^T. \tag{3.19}
\end{aligned}$$

ここに、 $\nabla_i = \nabla f_i(m_j)$  (横ベクトル)、 $C_j = \int_{V_j} (x - m_j)(x - m_j)^T dx$  である。

本節でも、隠れユニットは、局所的には最密構造になっているという条件下で MSE を最小にする分布を考えることにする。つまり、 $\nabla f_i(m_j)$  と直交する方向に薄く延びたボロノイ領域を作るようなことは考えない(そうすれば同じユニットの密度で MSE を減らすことができる)。なぜなら、その様な配列は Hebb 学習によってつくることができないからである。しかし、最密構造のボロノイ領域は球ではないので、これをどの方向に向けるかという問題は残る(つまり、 $L=2$  の場合ならば、正六角形の角を  $\nabla f_i(m_j)$  に関してどちらに向けるかという問題)。これに関してはつぎの補題がある。

### 補題 3.3

ボロノイ領域が最密構造になっているとき、 $\nabla_i C_j \nabla_i^T$  は、 $\nabla_i$  の方向によらない。

#### 証明

$C_j$  は対称行列なので、適当な直交行列  $O$  を用いて、 $C_j = O^T \Lambda O$ 、 $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_L)$  と分解できる。また、 $V_j$  の対称性から、いくつかの直交行列  $O_0$  に対して、 $O_0^T C_j O_0 = C_j$  となる。これより、 $Q = O O_0 O^T = (q_i)$  とすれば、

$$\Lambda = Q^T \Lambda Q. \tag{3.20}$$

ここで、一般性を失うことなく、 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$  と仮定する。式(3.20)の(1,1)要素に着目すると、

$$\lambda_1 = \sum_k \lambda_k q_{k1}^2. \tag{3.21}$$

$Q$  も直交行列だから、 $\sum_k q_{k1}^2 = 1$ 。よって、 $q_{11} \neq 0$  であるような  $k$  に対して、

$\lambda_k = \lambda_1$  となる。すべての  $k$  に対して適当な  $O_0$  がとれるので、結局  $\Lambda = \lambda_1 I$ 、よって  $O_j = \lambda_1 I$  を得る。したがって  $\nabla_i C_j \nabla_i^T = \lambda_1 \|\nabla_i\|^2$ 。 q. e. d.

この補題により MSE を最小にする隠れユニットの配置問題は、 $f_i$  の変化の二次以上のオーダーを無視すれば、密度  $\rho(x)$  のみを問題にすればよいことがわかる。

$C_j$  の定義より式(3.19)から、 $\lambda_1$  はボロノイ領域  $V_j$  の直径の  $L+2$  乗に比例する量である。よって式(3.13)と同様に、

$$\text{MSE}_{i,j} \propto p(m_j) \|\nabla_i(m_j)\|^2 \rho(m_j)^{-(L+2)/L}. \tag{3.22}$$

を得る。これより MSE は、

$$\begin{aligned}
\text{MSE} &= \sum_{i,j} \text{MSE}_{i,j} \\
&= \int_{\mathcal{S}} p(x) \sum_i \|\nabla_i(x)\|^2 \rho(x)^{-(L+2)/L} \rho(x) dx \\
&\propto \int_{\mathcal{S}} z(x) \rho(x)^{-2/L} p(x) dx. \tag{3.23}
\end{aligned}$$

ここに  $z(x) = \sum_i \|\nabla_i(x)\|^2 = \sum_i \|\nabla f_i(x)\|^2$  は、 $x \in \mathcal{S}$  付近で、関数  $f_i(x)$  が、どのくらい速く変化しているかを表している。これはこの付近での、この関数の学習のむずかしさを表すひとつの尺度と考えていいだろう。

式(3.23)を最小にする  $\rho$  を、式(3.15)と同様に求めることができる。

#### 定理 3.5

中間層の素子の数が十分大きいとき、平均自乗誤差を最小にする中間層の素子の分布密度は、次の式で近似される。ただし、中間層の素子は微視的には最密構造を作っているとする。

$$\rho(x) \propto \{z(x)p(x)\}^{L/(L+2)}. \tag{3.24}$$

式(3.24)は、式(3.15)の  $p(x)$  を  $z(x)p(x)$  で置き換えたものにほかならない。従ってつぎのことがいえる。ある入力分布  $p(x)$  に対して MSE を最小にしたのなら学習の際は入力分布として  $p(x)$  の代わりに、 $z(x)p(x)$  使うべきである。これを一言でいえば、“むずかしいことは何度も練習すべし”ということである。

### §3.5 誤差について

前節で取り扱ったモデルは与えられた非線形関数を階段関数で近似するものであるから、その誤差の解析は容易である。

まず学習の際の入力の提示確率として  $p(x)$  をもちいた場合は、(3.15)より

$$\rho(x) = CMp(x)^{1/(L+2)},$$

となる。ここに  $M$  は隠れユニットの総数、 $C$  は、

$$C = 1 / \int p(x)^{1/(L+2)} dx,$$

である。よって、 $x$  付近でのポロノイ領域の直径  $a(x)$  は、

$$\begin{aligned} a(x) &= (\rho(x)v)^{-1/L} \\ &= (vCM)^{-1/L} p(x)^{-1/(L+2)}. \end{aligned}$$

で与えられる。ここに  $v$  は  $L$  次元の最密配列を母点とするのポロノイ領域 ( $L=2$  ならば正六角形) で直径が1であるものの体積である。すると、学習すべき非線形関数とこのモデルの実現する関数の差は、点  $x$  付近のポロノイ領域の境界では、

$$\begin{aligned} e_i(x) &\approx a(x) \|v_i(x)\| / 2 \\ &\approx (vCM)^{-1/L} p(x)^{-1/(L+2)} \|v_i(x)\| / 2. \end{aligned}$$

となる。ただしこれは関数の第  $i$  成分に関する誤差である。

学習の際の入力の提示確率として  $z(x)p(x)$  をもちいた場合も同様に、

$$e'_i(x) \approx (vC'M)^{-1/L} \{z(x)p(x)\}^{-1/(L+2)} \|v_i(x)\| / 2.$$

ただし、

$$C' \approx 1 / \int \{z(x)p(x)\}^{1/(L+2)} dx,$$

と求めることができる。

いずれの場合も誤差は  $M^{-1/L}$  のオーダーであるが、 $x$  によって変化する部分を比較すると、学習の際の入力の提示確率として  $z(x)p(x)$  をもちいた場合は、

$\|v_i(x)\|$  が大きいところでは  $z(x)^{-1/(L+2)}$  は小さくなる傾向があるので誤差の変化は  $e_i(x)$  より  $e'_i(x)$  のほうがなだらかになる。特に出力層が1個のユニットからなる場合は、 $z(x) = \|v_i(x)\|$  だから、

$$e'_i(x) \approx (vCM)^{-1/L} p(x)^{-1/(L+2)} \|v_i(x)\|^{(L+1)/(L+2)} / 2,$$

となる。

### §3.6 バック・プロパゲーションを用いた場合

バック・プロパゲーションは、本来フィードバック結合のない神経回路の学習法として提案されたものであり、一般にフィードバック結合のある神経回路の場合に、拡張しようとする、結合強度の変化量を局所的かつ並列的に計算することができない。しかし本章で取り扱ったモデルでは、その困難を回避することができる。

このモデルにバック・プロパゲーションを適用すると、まず中間層と出力層の間の結合に関しては、誤り訂正学習であるから、(3.9)と同じになる。問題は入力層と中間層の間の結合  $w^1_j$  である。

ある入力  $x$  が回路に与えられたとき、その入力に対する回路の出力を入力  $x$  を正解に近付けるように結合を変化させて行くのがバック・プロパゲーションの基本的な考え方である。 $x$  が何れかのポロノイ領域の内点であるとき、 $w^1_j$  に微少な変更を加えても  $x$  が属するポロノイ領域は変わらず、よって、出力に変化はない。出力に変化があるのは  $x$  が二つのポロノイ領域の境界上にあるときだけである。このとき、その二つのポロノイ領域の母点(中心の点)を  $w^1_i$ 、 $w^1_j$  とすると、中間層の  $i$  番目と  $j$  番目の素子への入力  $x$  が等しくなっている。すなわち、

$$(x, w^1_i) = (x, w^1_j), \quad (3.25)$$

がなりたつ。もし  $x$  が、 $w^1_i$  の周りのポロノイ領域に入れば出力は  $w^2_i = (w^2_{i1}, w^2_{i2}, \dots, w^2_{in})^T$ 、 $w^1_j$  の周りのポロノイ領域に入れば出力は  $w^2_j = (w^2_{j1}, w^2_{j2}, \dots, w^2_{jn})^T$  になるのであるから、 $w^2_i$  と  $w^2_j$  のうち、正解  $f(x)$  に近い方の出力を出すようにするには、

$$r = \|x - w^2_j\|^2 - \|x - w^2_i\|^2,$$

を使って,

$$w^1_i := w^1_i + \epsilon_i r x / \|w^1_i + \epsilon_i r x\|, \quad (3.26)$$

$$w^1_i := w^1_i - \epsilon_i r x / \|w^1_i - \epsilon_i r x\|.$$

とすればよい (図3.5). これを条件(3.25)が成立する場合にのみ実行するのである.

いうまでもなく, 入力  $x$  がちょうど二つのボロノイ領域の境界上にくる確率は0である. よって, ここで述べた学習をそのまま実行したのではいつまでたっても回路は変化しない. そこで実際には, 条件(3.25)は次のように変更しなければならない.

$$\|(x, w^1_i) - (x, w^1_j)\| < d. \quad (3.27)$$

この場合  $(x, w^1_i)$  と  $(x, w^1_j)$  は,  $(x, w^1_1), (x, w^1_2), \dots, (x, w^1_n)$  のうちの最大のみたつ. また  $d$  は適当な正の数である. こうすると  $x$  がボロノイ領域の境界にある程度近い場合に  $w^1_i$  を変化させることになる.

また,  $w^1_i$  を変化させる場合とそうでない場合の間を連続的につないで,

$$w^1_i := w^1_i + \epsilon_i r g x / \|w^1_i + \epsilon_i r g x\|, \quad (3.28)$$

$$w^1_i := w^1_i - \epsilon_i r g x / \|w^1_i - \epsilon_i r g x\|,$$

としてもよい. ここに,

$$g \triangleq \phi (\|(x, w^1_i) - (x, w^1_j)\| / d), \quad (3.29)$$

$\phi(x)$  は, たとえば  $\exp\{-x^2\}$  のように  $|x| \rightarrow \infty$  で0に収束する対称な単調減少関数である. また,

$$\phi(x) \triangleq \begin{cases} 1, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

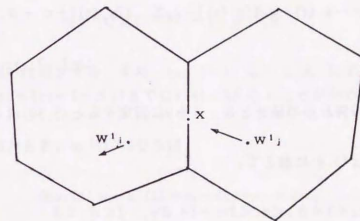


図3.5 学習則(3.26)の意味

入力  $x$  が二つのボロノイ領域  $w^1_i, w^1_j$  の境界上 (もしくは境界のごく近く) にきた場合, 正解  $f(x)$  と  $w^1_i, w^1_j$  を較べて近い方の出力が出るように母点  $w^1_i, w^1_j$  を動かす.

とすれば, (3.26), (3.27)と同じになる.

(3.27), (3.29)において,  $d$ は定数とせず, 学習の初期では, ある程度大きくしておき, 次第に0に近付けようにしたほうがよいだろう.

条件(3.27)が成り立つかどうかを判定するには, 中間層の素子の中で最大の入力を受け取っているものだけでなく, その次に大きな入力を受けている素子を見つけなくてはならない. それには, 次のようにすればよい.

まず, 中間層の全ての素子に対して共通に, 新たな制御用の入力  $\delta$  を設ける. この入力は, 各素子に対して等しいシナプス荷重1で結合している.

$$\begin{aligned} \dot{u}_i(t) = -u_i(t) + c_1 1[u_i(t)] - c_2 \sum_{j=1}^M 1[u_j(t)] + s_i + \delta, \quad (3.30) \\ i = 1, 2, \dots, M. \end{aligned}$$

$\delta$  は  $\delta_1 > 0$  と 0 の何れかの値をとる.  $\delta$  を 0 に固定すると (3.30) は (3.2) に一致する.

$c_1, c_2, \delta_1$  は (3.4) に加えて,

$$2c_2 + 1 - \delta_1 < c_1 < 3c_2 - 1 + \delta_1, \quad 2 < \delta_1 < 3 \quad (3.31)$$

を満たすようにとる. これを満たす  $c_1, c_2, \delta_1$  が存在することは容易に確かめられる. 実際例えば,  $c_1 = 4.7, c_2 = 3, \delta_1 = 2.5$  とすればよい. (3.29) のダイナミクスについて, 補題 3.2 と同様に, 次の補題が証明できる.

#### 補題 3.4

基本競合系 (3.30) において,  $\delta = \delta_1$  とする. シナプス強度  $c_1, c_2$  が, 条件 (3.4), (3.31) を満たし, 入力  $s_i, i = 1, 2, \dots, M$  が, 条件 (3.3) を満たし, また, ある  $s_j, s_k$  について,

$$s_j > s_i, \quad s_k > s_i, \quad i \neq j, k$$

が成り立つとする. さらに全ての  $u_i$  に等しい正の初期値を与えて, 入力  $s_i$  を固定し,  $\delta$  も  $\delta_1$  に固定したとする. このとき,

- i)  $n(t) \geq 3$  のとき, 全ての  $u_i$  は, 減少し,  $s_i, i = 1, 2, \dots, M$ , によらずに定まるある時刻までに  $u_i(t) < 3$  となる
- ii)  $n(t) = 2$  のとき, 二つの正の  $u_i$  は, ある正の値に漸近し, 他の  $u_i$  は,

それぞれ, ある負の値に漸近する.

証明

(3.30) において,  $\delta = \delta_1$  とし,  $n(t)$  を使って書き直すと,

$$\dot{u}_i(t) = -u_i(t) + c_1 1[u_i(t)] - c_2 n(t) + s_i + \delta_1.$$

これと,  $s_i, c_1, c_2$  に関する条件 (3.3), (3.31) より

i)  $n(t) \geq 3$  のとき,  $u_i(t) \geq 0$  ならば,

$$\dot{u}_i(t) \leq c_1 - 3c_2 + 1 + \delta_1 < 0,$$

となり,  $u_i(t)$  は減少する. また,  $s_i, i = 1, 2, \dots, M$ , によらずに定まる時刻  $t = u_0 / (-c_1 + 3c_2 - 1 - \delta_1)$  までに  $n(t) < 3$  となることが分かる.

ii)  $n(t) = 2$  のとき,  $u_i(t) > 0$  ならば,

$$\dot{u}_i(t) = -u_i(t) + c_1 - 2c_2 + s_i + \delta_1,$$

となり,  $u_i(t)$  は,  $c_1 - 2c_2 + s_i + \delta_1$  に漸近する. 条件 (3.3), (3.31) より, これは正である. また,  $u_i(t) \leq 0$  ならば,

$$\dot{u}_i(t) = -u_i(t) - 2c_2 + s_i + \delta_1,$$

となり,  $u_i(t)$  は,  $-2c_2 + s_i + \delta_1$  に漸近する. 条件 (3.3), (3.4), (3.31) より, これは負である. q. e. d.

また補題 3.1 は基本競合系 (3.29) に対しても成り立つ. 補題 3.1 と補題 3.4 から, 定理 3.1 と同様に次の定理を得る.

#### 定理 3.6

基本競合系 (3.30) において, 補題 3.4 の条件が成り立つとすると,  $s_i, i = 1, 2, \dots, M$ , によらずに定まるある時刻までに, 最大の入力を受け取っているものと, その次に大きな入力を受けている素子の二つが, 出力1を出し, 他の出力は0になる.

入力  $x$  が与えられたら、しばらくの間  $\delta = \delta_1$  としておき、中間層の素子のなかで最大の入力を受け取っているものと、その次に大きな入力を受けている素子の二つが、出力 1 を出し、他の出力は 0 になるまで待つ。そのうち、 $\delta = 0$  として、最大の入力を受け取っているものだけが残るのを待たせよ。

### S3.7 モデルの単純化

本章では、入力  $x$  と  $w^i$  の絶対値を 1 に制限したが、この制限を用いないモデルを作ることもできる。

入力層に、 $L$  個のユニットがあり、ここに入力  $x = (x_1, x_2, \dots, x_L)^T$  が与えられるとする。入力ベクトルのユークリッドノルムを 1 であると仮定しないかわりに、次元が  $L+1$  からひとつ減って  $L$  になったことに注意されたい。これに伴って、中間層の素子の荷重ベクトル  $w^i$  の次元もひとつ減る。中間層のユニットの入力には、(3.6) のかわりに、

$$s_i = h(\|x - w^i\|), \quad i = 1, 2, \dots, M. \quad (3.32)$$

を使う。ここで、 $h(x)$  は、 $-1 < h(x) < 1$  を満たす、単調減少関数である。学習則 (3.10) は、

$$w^i_{j+1} := (w^i_j + \epsilon_i [u_j(x - w^i_j)]), \quad i = 1, 2, \dots, M. \quad (3.33)$$

$$0 < \epsilon_i < 1,$$

と変更する。(3.26)、(3.28) も同様である。

このモデルに関して本章で述べられたことは全て成り立つ。

最後に、本章では、入力層は  $S^L$  全体に分布していると仮定したが  $S^L$  上の次元の低い、なめらかな部分多様体上に分布していたとしても、同様な議論が成り立つ。ただしその場合、 $L$  は、その部分多様体の次元となる。それでは、入力が  $M$  次元 ( $K \leq L-1$ ) のなめらかな多様体に沿った非常に薄い  $K+1$  次元の多様体上に分布しているとき、 $K$  を部分多様体の次元と解釈すべきなのか、それとも  $K+1$  を部分多様体の次元と解釈すべきなのか。この判断は中間層の素子の数とも関係してくる。ポロノイ領域を  $M$  次元多様体上で考えたときの直径に較べて、この  $M+1$  次元多様体の厚さが十分小さいときは、 $K$  を部分多様体の次元と解釈すべ

きである。逆に直径の方が厚さと較べて十分小さいときは、 $K+1$  を部分多様体の次元と解釈すべきである。

### 文献

1. 甘利俊一, 神経回路網の数理. 産業図書, 昭52.
2. 麻生英樹, バックアロバゲーション. コンピュータローラ, 23, pp.53-60, 昭63.
3. 村上研二, 泉田正則, 相原恒博, 二段階連想方式とその分散形連想記憶の記憶領域縮小への応用. 信学論(A), J67-A, 9, pp.912-919, 昭59.
4. Baldi, P., Hornik, K., Neural networks and principal analysis: learning from examples without local minima. Neural Networks, 2, pp.53-58, 1989.
5. Gallinari, P., Thiria, S., Soulie, F. F., Multi-layer perceptrons and Data Analysis. Proc. IEEE International Conference on Neural Net-Works, San Diego, California, 11, pp.391-399, 1988.
6. Hecht-Nielsen, R., Applications of counterpropagation networks. Neural Networks, 1, pp.131-139, 1988.
7. Iri, M., Murota, K., Ohya, T., A fast voronoi diagram algorithm with applications to geographical optimization problems. Proc. 11th IFIP Conference on System Modelling and Optimization 1983, Copenhagen, Lecture Notes in Control and Information Science 59, System Modelling and Optimization, ed. P. Thoft-Christensen, pp. 273-288, Springer-Verlag, Berlin, 1984.
8. Lehky, S., Sejinowski, T., Network model of shape-from-shading: neural function arises from both receptive and projective fields. Nature London, 333, pp.452-454, 1988.
9. Pineda, F. J., Generalization of back-propagation to recurrent neural networks. Physical Review Letters, 59, pp.2229-2232, 1987.
10. Rumelhart, D. E., McClelland, J. L. et al., Parallel distributed processing, 1, 2, The MIT Press, Cambridge, Massachusetts, 1986.

## §4.0 はじめに

われわれの網膜上に与えられた視覚情報は外側膝状体 (LGBまたはLGN) の細胞を経由して大脳皮質後部の視覚領の第17野に送られる。外側膝状体も大脳皮質も層状構造をなしており、何枚かの2次元の神経場と考えてよい。網膜の出力細胞は神経節細胞と呼ばれるが、1個の神経節細胞は外側膝状体の比較的狭い範囲の細胞に結合している(投射するという)が、網膜上で隣合う神経節細胞が投射する領域は、外側膝状体上でもやはり隣合っている。すなわち、網膜から外側膝状体への投射はトポロジカルなマッピングになっている。外側膝状体から第17野への投射も同様である。したがって、網膜と第17野の細胞の対応関係もやはりトポロジカルである。ただし、視野の中心付近の網膜では周りと較べて視細胞のサイズが小さく、単位面積あたりの細胞数が多くなっている。この部分は、第17野では広い面積を占めている。このため網膜と第17野の細胞の対応関係は、トポロジカルではあるが、歪んだものになっている。魚類や両棲類など大脳皮質の未発達な脊椎動物では、視覚の中樞は大脳視覚領ではなく中脳の上丘と呼ばれる部位であるが、これらの動物でも網膜と上丘の間に同様の連続的な結合関係が観察される。

このように、二つの神経場間のトポロジカルな投射のことをトポグラフィあるいはトポグラフィック・マッピングとよんでいる。視覚系に見られる、網膜と関係したトポグラフィック・マッピングを、特にレティノトピと呼んでいる。レティノトピは、外側膝状体と第17野だけでなく、視覚領のさらに高次の部分でもある程度保存されている。

脳のなかには、レティノトピに限らず、さまざまなトポグラフィック・マッピングの存在が知られている。たとえば、運動野と体性感覚野には、それぞれ全身の皮膚とのトポグラフィがあって、手のように複雑な運動をする部位は運動野で広い範囲を占めており、舌のように感覚の細やかな部位は感覚野で広い範囲を占めている。

第17野のレティノトピについては、以下に述べるように、さらに細かい構造が知られている[6]。第17野には、網膜上の特定の位置(受容野)に特定の傾きの線分あるいは明暗の境界(エッジ)が現れたときに反応する単純型細胞と、網膜上のある領域内に特定の傾きの線分あるいは明暗の境界(エッジ)が現れたときに反応する複雑型細胞があるが、これらの細胞は、同じような方

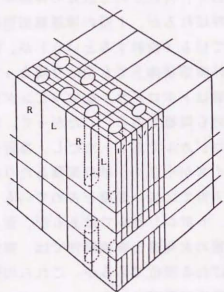


図4.1 視覚領のコラム構造とハイパー・コラム構造の模式図

コラムの特性には選択方向の他に上図左側面に示された眼優位性がある。そして、幾つかのコラムにまたがって、色選択性を持つブロップと呼ばれる領域がある（図中の円柱）。

向の線分やエッジに反応するものが皮質上でも隣合うように並んでいる。したがって、第17野の表面に一本の直線を引いて、それに沿って細胞の反応を見て行くと、細胞の受容野の位置と反応する線分またはエッジの方向の両方が変化していく。しかし受容野の位置の変化は、反応する線分またはエッジの方向に較べて、遙かにゆっくりしたものになっている。また、受容野の位置にかかわらず、ある方向の線分またはエッジに反応するすべての細胞の分布を調べると、ちょうど指紋のようにもなっていることもわかっている。この「紋」の縞の幅すなわち方向の変化の「波長」は、サルの場合数百 $\mu$ である。

さらに細かく見ると、細胞の受容野の位置と反応する線分またはエッジの方向の変化は連続的なものでなく、ところどころで小さなジャンプを繰り返す階段関数のようなものになっている。これは、大脳皮質に広く存在するコラム構造のひとつである。大脳皮質はコラムと呼ばれる小領域に分割され、ひとつのコラムがひとつの機能単位として働いている。第17野の場合、同じコラムに含まれる細胞は、単純型、複雑型など細胞の種類は違っても、ほぼ同じ位置の受容野と反応方向をもっている（図4.1）。

第17野上で、直径数百 $\mu$ の範囲内に存在するコラムの集合を考えると、そのなかには、視野上の一点に提示されたすべての方向の線分またはエッジに反応する細胞が揃っている。これを一つの機能単位と考えることもできるとことから、これをハイパー・コラムと呼ぶ。

視覚領に見られるこのようなモザイク状の複雑な構造は、2次元の大脳皮質に、2次元の視野とその各点に与えられたさまざまな情報をうまく配置する役割を果たしていると考えられる。

それでは、このような構造は、なにによって作られるのか。遺伝子によってあらかじめ決まっているとする説と、生後与えられる視覚情報に応じて自己組織によってつくられるとする説とがあって、未だに完全な決着をみてはいない。しかし、脳のどの部分のどの種類の細胞がどの部分のどの種類の細胞に投射するかなどの大まかなところは遺伝子によって決定され、神経回路の細部は自己組織によってつくられるとみるのが妥当であろう。

WillshawとMalsburg[10]は、レティノトピーの形成のモデルを発表した。彼らのモデルの神経場は、入力層、出力層ともに直線上に並んだ神経細胞群である。入力層の細胞は外部から与えられた入力によってのみ発火し、入力層の細胞間の結合や出力層からのフィード・バック結合はない。入力層に隣合ったひとかたまりの細胞が同時に発火するようなパターンが与えられる。出力層では近くの細胞同士は互いに興奮性の結合で結ばれ、少し離れた細胞同士は逆に抑制性の結合で結ばれているため、一つの細胞が発火すると、その周囲のある範

皿の細胞がまとまって発火する。そして、入力層と出力層の間の結合はHebb則によって形成される。ランダムな結合から出発したシミュレーションの結果、入力層から出力層への連続的な結合が形成されることが示された。

これに先立って、Malsburg[9]は、単純型細胞に相当する方位選択細胞の場の形成モデルも発表している。このモデルの神経場は入力層、出力層ともに平面状に並んだ神経細胞群である。入力層に与えられる入力は9種類の方向の異なった直線パターンで、方向の似通ったパターンは、共通に発火する細胞の数が多くなっている。したがって、これらのパターンの集合はT<sub>1</sub>(円環)の構造を持っている。その他はレティノトピーのモデルとほぼ同じである。やはりランダムな結合から出発したシミュレーションの結果、出力層において近くにある細胞同士は似たような方向の入力に強く反応するようになった。

Amariら[5,9]は、WillshawとMalsburgのレティノトピーの形成のモデルとほぼ同様のモデルを解析し、1次元の入力層の一定の長さの区間に含まれる細胞を同時に発火させて学習を進めるとき、入力層の神経場と、1次元の出力層の神経場との間の、連続なマッピングが安定に保持される条件や、出力層における分解能などを理論的に求めた。また、連続なマッピングが不安定になる場合には、コラム状の微細構造が形成されることをシミュレーションによって示した。Kohonenも同様のモデルに関して様々なシミュレーションをおこなっている[7]。Kohonenのモデルについては、§4.2で詳しく述べる。これらのモデルは、脳内に普遍的にみられるトポグラフィの形成原理に関する仮説となっている。

以上の四つのモデルの共通の特徴を、次の三つにまとめることができるだろう。

- 1) 入力層に与えられるパターンの集合(信号空間)はパターン間の内積の大きさによって定義されるトポロジーを持っている。
- 2) 出力層では近くの細胞同士は互いに興奮性の結合で結ばれ、少し離れた細胞同士は逆に抑制性の結合で結ばれている。このため一つの細胞が発火すると、その周囲にある範囲の細胞がまとまって発火する。出力層はこの発火パターンによって定義されるトポロジーを持っている。
- 3) 入力層と出力層の間の結合はHebb則によって形成される。

もちろん、入力層に与えられるパターンの集合も出力層の細胞の集合も有限集合であるから、ここでいう“トポロジー”は厳密に言えば比喩である。

この三つの特徴にもうひとつ付け加えるとすれば、これらのモデルでは、いずれも入力信号の次元が出力層の次元と同じかあるいはそれ以下である。しかし、例えば網膜のあちこちに様々な方位の線分が写るような状況を考えて、

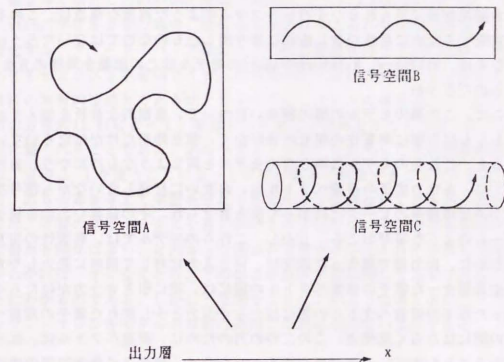


図4.2 信号空間の次元が出力層の次元より高い場合に予想されるマッピング

出力層の細胞にそれを最もよく興奮させる信号を対応させる。出力層から、信号空間への写像を考え、出力層が1次元で信号空間が2次元の場合に予想されるマッピングをこの写像の像によって示した。Aは信号空間が広い平面の場合、Bは細長い帯状領域の場合、Cは細長い円柱の側面である場合である。



入力信号は、 $D \times T^1$  (円板  $\times$  円環) すなわち 3 次元の構造を持っており、これに対して脳内の神経場は 2 次元である。このように、入力信号の信号空間のほうが出来層より次元が高い場合には、位相幾何学の教えるところによれば、入力信号空間から出力層への 1 : 1 の双方向に連続なマッピングは存在しえない。大脳視覚領に見られるハイパー・コラムのような複雑な構造は、この不可能を克服するために自然が苦し紛れに作り出したものなのではないだろうか。それならば、Willshaw, Malsburg や Amari のモデルはこの困難を同様の方法で回避するのだろうか。

ここで、この種のモデルの解の振舞いについて、直観的な分析を加えておこう。もしも出力層に興奮性の相互結合がなく、相互抑制だけがはたらいいたとすると、これらのモデルは第 3 章のモデルと同じようなものになり、前述したように、各出力素子の荷重ベクトルは、お互いに反発し合いながら信号空間のなかの出現確率のピークに向かって引き寄せられ、その結果いわゆる椅子取りゲームのような事がおこる。しかし、これらのモデルでは、興奮性の相互結合のために、出力層で隣合った素子は、同じ入力に対して同時に発火しやすく、その結果隣合った素子の荷重ベクトルの間には、逆に引き合う力がはたらく。隣合った素子の荷重ベクトルの間にはたらく引力と少し離れた素子の荷重ベクトルの間にはたらく反発力、この二つの力のために、荷重ベクトルは、ばらばらになることもなく、一点に集中することもなく、順序よく信号空間の中に拡がっていくのである。したがって、もし信号空間の次元が出力層の次元よりも高い場合は、信号空間内の荷重ベクトル群の作る曲面は余分の次元を埋めるためにひだを作ってしまうことが予想される。信号空間の余った次元の方向の幅が厚いときは様々な不規則なうねりかたが可能だが、その方向の幅が薄いときは、うねりかたが、比較的規則的なものに限られるのではないだろうか (図 4.2)。

次節 § 4.1 では、この問題にボルツマン・マシンの理論を応用し、入力層が 2 次元で出力層が 1 次元である場合に、ハイパー・コラム状の構造ができることを示す。さらに § 4.2 では、Kohonen のモデルに現れる単純連続解の安定性を調べることにより、入力信号の信号空間のほうに、出力層より次元が高い場合にハイパー・コラム状の構造ができるための条件を求める。また、現在までのところでは Amari のモデル以外のモデルでコラム状の構造ができた例はないが、コラム状の微細構造の形成されるための条件を解析して、Kohonen のモデルでは、コラム構造はできないことを示す。さらに、モデルをより自然なものに改良することによってコラム状の構造を作ることができることを示す。

#### § 4.1 ボルツマン・マシンを応用したトポグラフィック・マッピング形成のモデル

##### 4.1.0 ボルツマン・マシンをモデルに適用することの意味

第 2 章で紹介した Hinton の定理 (定理 2.1) によればボルツマン・マシンに Hinton の学習をさせた場合、結合は V 素子に与えられた分布を再現するように変化する。フィード・バック結合をもつ学習神経回路モデルの場合、神経興奮のダイナミクスと学習のダイナミクスの関わり合いの解析が神経回路モデルの解析の本質的な部分であるが、ボルツマン・マシンの場合は、それがたった 1 本の方程式 (2.4) によって、簡潔かつ非常に一般的なかたちで表されているのは実に驚くべきことである。ボルツマン・マシンを応用した学習神経回路モデルは、定理 2.1 を使うことによって非常に簡単に解析することができる。すなわち、ある分布を与えて学習させたときどのような結合ができるかを知りたいときは、学習のダイナミクスを追いかけるのではなく、その分布を再現するような結合を探せばよい。どのような結合が V 素子に与えられた分布を再現するかがわかれば、すくなくとも、その結合状態が学習に関して安定であることが直ちにわかるのである。しかし不思議なことにこのような考え方になって、ボルツマン・マシンをモデルの解析に應用した例は未だないようである。

##### 4.1.1 ボルツマン神経場

ここではボルツマン・マシンによる神経場を考えるが、その基礎となるのが、第 2 章の例 2.2 である。それによると、 $n$  個の V 素子だけからなるボルツマン・マシンにおいて、その内の  $k$  個 ( $0 \leq k \leq n$ ) が発火するパターン ( $nC_k$  個ある) を等確率で出現させるような分布は

$$w_i = r(k - 0.5),$$

$$w_{ij} = -r, i \neq j,$$

として  $r \rightarrow \infty$  の極限で実現できる。

さて、 $n$  個の素子を考え、この結合によって、そのうちの  $k$  個だけが興奮するように相互抑制性結合を定める。つぎに、これを環状に並べ、隣合った素子だけを、ある程度の強さの興奮性結合でつなぐと、結果として、Malsburg, Amari などの仮定した結合と同じものができ (図 4.3)。

$$w_{ij} = r(k - 0.5),$$

$$w_{ij} = -r, \quad i \neq j, \quad j \pm 1 \pmod{n}$$

$$w_{ij} = -r(1+b), \quad i = j \pm 1 \pmod{n}$$

$b$ が十分小さければ、隣合った素子どうしの結合に付け加えられた興奮性結合  $rb$ がエネルギーに与える影響も小さく、興奮する素子の数が  $k$ 個のときのエネルギーがそれ以外の数の素子が興奮する場合のエネルギーより低いという性質は保存される。しかし興奮する素子の数の等しいパターン間のエネルギーの間には、興奮性結合の影響で微妙な差が生ずる。興奮性結合によるエネルギーの減少は、発火パターンに含まれる隣合ったペアの数に比例するから、 $b$ を十分小さく取れば、 $r$ 無限大の極限で、連続した  $k$ 個の素子の発火だけが起きるようにできる。これは、甘利の神経モデル [1.5, 9]における局在興奮 (local excitation) に相当している。

#### 4.1.2 1次元から1次元へのトポグラフィック・マッピング

ボルツマン・マシンの学習則は、 $V$ 素子群に関しては、教師付学習である。したがって、これをトポグラフィック・マッピングの形成モデルに応用しようとする時、出力層を  $V$ 素子群と考えることができない。そこで、入力層を  $V$ 素子群とし、出力層を  $H$ 素子群とする。通常のトポグラフィック・マッピングでは、入力層から出力層に向けてのみ結合ができるが、ボルツマン・マシンでは、すべての結合は対称なので、ここでは、入力層と出力層の間に対称な結合が形成されるモデルを考える。

前節で考えた  $n$ 個の素子からなる環状のボルツマン神経場を考え、隣合った2個の素子だけが、発火できるように結合を定める。環状の場を考えるのは、境界の影響を除外するためである。これが出力層すなわち  $H$ 素子群である。入力層すなわち  $V$ 素子群も同じく環状に並んだ  $n$ 個の素子からできている。 $V$ 素子群の細胞は、すべて等しく高いしきい値をもっているとする。 $V$ 素子群内部には結合は考えない。また  $H$ 素子群内部の結合も固定され、学習は  $V$ 、 $H$ 素子群の間の結合でのみ起こるものとする。

このボルツマン・マシンの  $V$ 素子群に隣合った2個の素子が発火するパターンを等確率で提示して Hinton の学習則に従って学習と反学習をくりかえすと、ボルツマン・マシンは、その自由な動作状態において、 $V$ 素子群に隣合った2個の素子の発火を等確率で再現するようになるはずである。では、それを可能

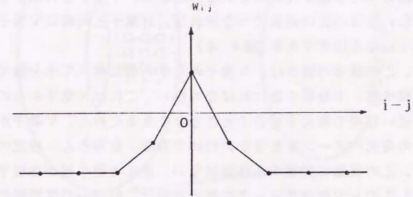


図4.3 ボルツマン神経場の内部結合

この図では、 $w_{ij}$ のところにしきい値  $w_i$ を示してある。

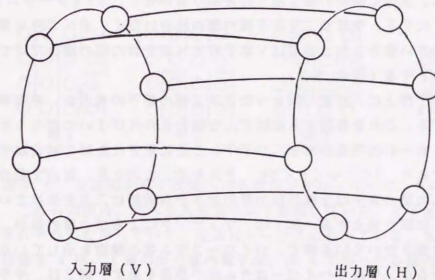


図4.4 ボルツマン神経場における  
1次元から1次元へのトポグラフィック・マッピング

入力層と出力層の細胞を一つずつ順序よくつなげば、入力層に与えられた分布を実現できる。

にする結合は、どのようなものなのだろうか。V素子は、高いしきい値をもっているので、H素子からの正の入力がなければ、ほとんど発火できない。H素子群は、隣合った2個が発火しているのだから、V素子とH素子を順序よく1対1に等しい強さの正の結合でつなげれば、H素子と同時にV素子も同じ発火をするようになるはずである(図4.4)。

ただし、この結合の強さは、V素子のしきい値に較べて十分強く、H素子間の結合に較べて、十分弱くなければならない。これはH素子からの入力がV素子を1に近い確率で発火させることができるためと、V素子からの入力がH素子群の発火パターンを乱さないためである。もちろん、任意の初期値から出発して、この状態に到達する保証はない。多くの極小値が存在するはずである。しかしこのトポグラフィックな結合状態は、Hintonの学習則に関して安定である。

#### 4.1.3 2次元から1次元へのトポグラフィック・マッピング

この場合も、H素子群の構造は、1次元から1次元への場合と同じである。ただし、素子の数は $m \times n$ 個であるとする。一方、V素子群の方は、 $m \times n$ 個の長方形に並べ、さらに向かい合う辺と辺を貼り合わせて、 $T^2$ (トーラス)の構造をもつようにする。やはり、V素子群内部の結合はなく、すべてのV素子は、等しく高いしきい値をもち、学習はV素子群とH素子群の間の結合だけでおこなわれるものとする(図4.5)。

さて、V素子群上に、任意の隣合った $2 \times 2$ 個の素子の発火が、等確率で提示されたとする。これを再現するにはどんな結合を作ればよいのだろうか。図4.6に $m = n = 4$ の場合の解を三つ示す。ここに示された解の結合はすべて同じ強さ $v$ であり、 $1 < v < \theta < 2v$ をみたす。このとき、自由な動作状態において、入力層の素子は2個の出力層の素子から同時に入力を受けているとき、そのときに限り発火する。解1は、マッピングによる出力層の像が、入力層に螺旋状に巻き付いている解で、ハイパーコラム状の構造を示している。解2は少し異なるが、やはりハイパーコラム状の構造を示す。解3は、やや不規則な解である。 $m$ と $n$ が大きくなるにつれて、このような不規則な解の数が増えてくる。しかし、 $m$ 、 $n$ 共に小さい場合や、 $n$ だけが大きい、細長い領域の場合は、不規則な解は多くない。

本節では、学習神経回路網における神経興奮のダイナミクスと学習のダイナミクスの関わりあい、ボルツマン・マシンをつかって説明し、さらにその応用として、トポグラフィック・マッピングの形成モデルを構成した。このモデ

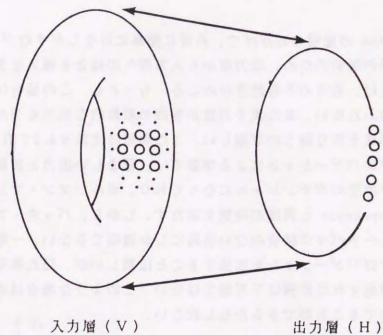


図4.5 2次元の入力層と1次元の出力層の素子の配列

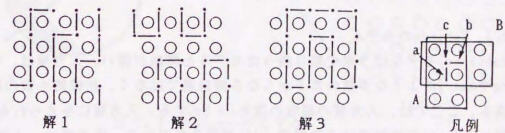


図4.6 2次元から1次元へのトポロジカル・マッピングの解の例

これらの図は、すべて入力層の素子を表している。上下と左右の辺は、それぞれつながっていて、全体はトーラスになっている。凡例中、例えば線分 a は、1個の出力層の素子が、枠 A で囲った6個の素子と結合していることを示す。線分 b は、出力層で、それに隣合った素子が、枠 B で囲まれた6個の素子と結合していることを示す。これらの二つの素子が同時に発火すると、点 c の周りの4個の素子が発火する。

ルでは、Hinton の定理のおかげで、非常に簡単にはなしがすむが、そのかわり、結合の対称性の制約のため、出力層から入力層への結合を導入せざるをえず、モデルとしては、若干の不自然さがのこる。もっとも、この結合は、反学習のときにしか使われぬ。また素子の数が有限の離散的な場のモデルなので、コラム構造の形成を取り扱うのが難しい。この問題は次節 § 4.2 で取り扱う。

バック・プロパゲーションによる学習では、望ましい出力と回路の出力の平均自乗誤差が学習のポテンシャルになっており、ボルツマン・マシンにおける Kullback divergence と同様の役割を果たす。しかし、バック・プロパゲーションは、フィードバック結合のない回路にしか適用できない。一般の神経回路にバック・プロパゲーションを拡張することは難しいが、隠れ素子群にある種の一様性を仮定すれば拡張は不可能ではない。このような場合は本節と同様の議論を組み立てることができるかもしれない。

#### § 4.2 Kohonen のモデルにおけるコラム構造とハイパー・コラム構造の形成

本節では、Kohonen のモデルの単純連続解の安定性を解析することにより、このモデルにおける、コラム構造とハイパー・コラム構造の形成の可能性を調べる。そのために、まず、Kohonen のモデルの簡単な解説をおこなう。

##### 4.2.1 Kohonen のモデル

Kohonen のモデルは 3 章で取り扱ったモデルと関係が深い。入力層は、Malsburg や Amari のような多数の素子からなる神経場ではなく、比較的小数の細胞から成る。ここでは、入力層の細胞の数を  $l+1$  とする。入力層に与えられる入力信号  $y \in R^{l+1}$  の絶対値は 3 章同様、1 に制限されている。出力層は、3 章で取り扱ったモデルの中間層にトポロジーをいれたものと見なすことができる。素子は 2 次元に並べられ、そのおのおのが荷重ベクトル  $w_{ij}$  を持っている (図 4.7)。荷重ベクトルの絶対値も、3 章のモデルと同様、すべて 1 に制限されている。

$$\|y\| = 1, \quad \|w_{ij}\| = 1, \quad i, j = 1, 2, \dots, M. \quad (4.1)$$

出力層の各素子の受け取る総入力も 3 章同様、入力信号と荷重ベクトルの内積  $(y, w_{ij})$  で与えられる。各素子にはそれぞれの近傍  $N(i, j)$  が定められている。通常は、その素子自身とそれに隣合う 4 個または 8 個の素子を近傍と定める。すなわち、

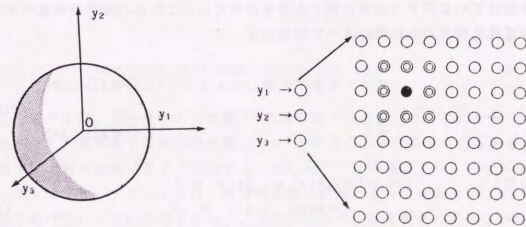


図 4.7 Kohonen のモデル

入力信号によって一つの素子 (●) が発火すると、周囲の 8 個の素子 (◎) も同時に発火し、Hebb 則にしたがって学習する。

$$N_5(i, j) \equiv \{(i, j), (i \pm 1, j), (i, j \pm 1)\}, \quad (4.2)$$

$$N_0(i, j) \equiv N_5(i, j) + \{(i \pm 1, j+1), (i \pm 1, j-1)\},$$

である。出力層が1次元の場合は、その素子の両側の何個かの素子を近傍とする。

$$N_D(i) \equiv \{i, i \pm 1, i \pm 2, \dots, i \pm D\}, \quad D=1, 2, \dots, \quad (4.3)$$

入力層に入力信号 $y$ が与えられると、出力層の細胞のうち、最も大きな入力を受けている素子の近傍に属する素子が発火し、これらの細胞の荷重ベクトルが正規化付きのHebb則によって変化する。

$$w_{ij} := \left\{ \begin{array}{l} w_{ij} + \epsilon y / \|w_{ij} + \epsilon y\|, \quad (i, j) \in N(i^*, j^*) \\ w_{ij}, \quad (i, j) \notin N(i^*, j^*) \end{array} \right. \quad (4.4)$$

ただし、
$$(i^*, j^*) \equiv (i^*(y, w), j^*(y, w)) = \operatorname{argmax}_{(i, j)} \{y \cdot w_{ij}\} \quad (4.5)$$

Kohonenのモデルは、それ以前のモデルと次の2つの点で異なっている。第一の点は入力層が比較的小数の素子から成っている点である。入力信号は、 $L+1$ 次元空間内の $L$ 次元球面 $S^L$ または、その中のさらに次元の低い部分空間から選ばれて入力層に与えられる。この信号空間の次元がMalsburgやAmariのモデルにおける入力層の次元に相当している。入力のノルムが1に制限されているので、他のモデルと同様に信号空間が内積で定義されるトポロジーを持つようになっているが、入力層の素子の数が少ないのでシミュレーションが容易である。

第2の点は、出力層における相互結合とそれによるダイナミクスを省略して、その代わりに各点の周りの孤立局在興奮に相当する近傍を導入した点である。これも、このモデルのシミュレーションを容易にしている。出力層のトポロジーはこの近傍によって定義されている。

#### 4.2.2 Kohonenのモデルの簡単化と連続化

Kohonenのモデルの数学的取り扱いを容易にするために§3.7で3章のモデル

に対しておこなったと同じ簡単化をおこなう。

まず、入力層の素子の数を一つ減らして $L$ 個とし、その代わり入力信号のノルムを1に制限することはしない。これにともなって、出力層の素子の荷重ベクトルも $L$ 次元となり、ノルムに関する条件もなくなる。

入力層に入力 $y$ が与えられると、出力層の素子のなかで、これに最も近い荷重ベクトルを持った素子の近傍が発火し、これらの素子の荷重ベクトルがHebb則に従って変化する。元のモデルと違って正規化はおこなわない。次節以後で取り扱うのは出力層の次元が1次元のときだけなので、その場合の式を書いておく。

$$w_i := \left\{ \begin{array}{l} w_i + \epsilon (y - w_i), \quad |i - i^*| \leq D, \\ w_i, \quad |i - i^*| > D. \end{array} \right. \quad (4.6)$$

ただし、
$$i^* \equiv i^*(y, w) = \operatorname{argmin}_i \{\|y - w_i\|\}, \quad (4.7)$$

このモデルは、元のモデルを球面 $S^L$ の接平面上で近似したものになっている。したがって、元のモデルの信号空間が、 $S^L$ 上のごく小さな領域である場合は元のモデルの解の振舞いをよく近似する。またそうでない場合でも、トポグラフィック・マッピングのモデルは、局所的な相互作用によって全体のマッピングを作りあげていくものであるから、この近似はモデルの本質をのがしていない。もともと我々は、 $S^L$ のトポロジーの大域的な性質そのものには何の興味もないのである。

このモデルにおける信号空間の位相は、信号空間が $R^L$ 全体の場合は、通常の $R^L$ の位相が信号空間の位相となり、 $R^L$ の部分空間であるときは、通常の $R^L$ の位相から導かれる相対位相が信号空間の位相となる。

次に、無限個の素子から成る出力層を考えることにより、このモデルを連続化する。境界条件を考えないですむように無限に長い1次元の神経場を考える。場所 $x \in X$ にある素子の荷重ベクトルを $w(x) \equiv Y$ とする。場所 $x$ の近傍は、

$$N_D(x) \equiv [x - D, x + D], \quad D > 0, \quad (4.8)$$

となり、学習則は、

$$w(x) + \epsilon(y - w(x)), \quad |x - x^*| \leq D, \\ w(x) := \begin{cases} w(x), & |x - x^*| \leq D, \\ \end{cases} \quad (4.9)$$

$$\text{ただし,} \quad x^* = x^*(y, w) \triangleq \operatorname{argmin}_x \{ \|y - w(x)\| \}, \quad (4.10)$$

となる。

#### 4.2.3 Kohonenのモデルの単純連続解の安定性。

本節では前節の最後で導いた連続化されたKohonenのモデル(4.9), (4.10)を取り扱う。

##### 4.2.3.1 帯状2次元領域の信号空間から1次元の神経場へのトポグラフィック・マッピング

境界の影響を考えないで済むように、出力層は実数全体とする。入力層は2個の素子からなり、入力信号は信号空間  $R \times [-a, +a]$ ,  $a > 0$ , から、一様な確率で選ばれ、入力層に与えられる。信号空間の次元の方が入力層の次元より高い場合を考えるわけである(図4.8)。信号空間をYで表し、Yに属する信号を  $y = (y_1, y_2)^T$  で表す。

$$y \in Y, \quad y = (y_1, y_2)^T, \quad y_1 \in R, \quad y_2 \in [-a, +a], \quad a > 0.$$

出力層の神経場はXで表し、その一点をxで表す。

$$x \in X = R.$$

Kohonenのモデルにおけるトポグラフィック・マッピングは、wによって決まる、信号空間Yから神経場Xへの写像、

$$y \rightarrow x = x^*(y, w) \triangleq \operatorname{argmin}_x \{ \|y - w(x)\| \},$$

と見ることができるが、Yの次元の方がXの次元より高い場合には、通常この写像は、無限個のyを1個のxに対応させてしまうので、逆にXからYへの写像、

$$x \rightarrow w(x) = (w_1(x), w_2(x))^T \in Y,$$

を主に取り扱う。当然ながら、もしw(x)が単射であれば、この二つの写像のあいだには、

$$x^*(y, w(x)) = x,$$

なる関係が成り立つ。

場所xの素子に最大の入力を与えるような信号の集合を  $M(x, w) \subset Y$  と書くことにする。

$$M(x, w) = \{ y \mid x = x^*(y, w) \}.$$

場所xの素子は、xの近傍  $N_D(x) = [x-D, x+D]$  に属する素子に最大の入力を与えるような入力に対して発火する。その様な入力の集合を  $R(x, w) \subset Y$  と書くことにする。

$$R(x, w) = \{ y \mid x^*(y, w) \in N_D(x) \} \\ = \bigcap_{x \in N_D(x)} M(x, w).$$

$R(x, w)$ を素子xの受容野(receptive field)という。

曲線w(x)がなめらかで、そのうねりが十分小さい場合、すなわち  $w_2(x)$  と  $w_2'(x)$  が十分0に近い場合は  $M(x, w)$  は、点w(x)における曲線w(x)の法線とYとの共通部分となり、したがって  $R(x, w)$  はYのなかで、 $M(x+D, w)$  と  $M(x-D, w)$  すなわち点w(x-D)における法線と点w(x+D)における法線に挟まれた部分になる(図4.8)。ここに ' はxによる微分を表す。

入力信号がYから一様分布に従って選ばれるとすると、モデル(4.9), (4.10)の平均学習方程式はつぎようになる。

$$\begin{aligned} \dot{w}(x, t) &= \int_{R(x, w)} \{ (y - w(x, t)) / 2a \} dy \\ &= (1/2a) \int_{R(x, w)} y \, dy - (1/2a) \int_{R(x, w)} w(x) \, dy \\ &= (|R(x, w)| / 2a) g(x, w) - (|R(x, w)| / 2a) w(x) \end{aligned}$$

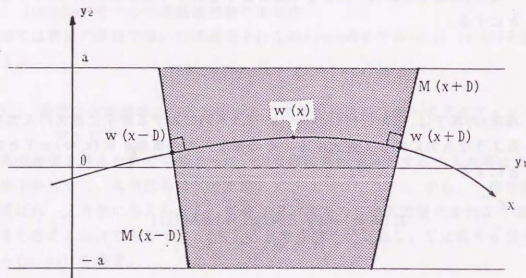


図4.8 帯状2次元領域から1次元の神経場へのマッピング

曲線  $w(x)$  のうねりが十分小さい場合、 $M(x, w)$  は、点  $w(x)$  における曲線  $w(x)$  の法線と  $Y$  との共通部分となり、受容野  $R(x, w)$  は  $Y$  のなかで、 $M(x+D, w)$  と  $M(x-D, w)$  すなわち点  $w(x-D)$  と点  $w(x+D)$  における法線に挟まれた部分になる。

$$= G(x, w) - S(x, w) w(x). \quad (4.11)$$

ここに、 $G(x, w) \triangleq (|R(x, w)| / 2a) g(x, w), \quad (4.12)$

$$S(x, w) \triangleq |R(x, w)| / 2a. \quad (4.13)$$

また、 $|R(x, w)|$ 、 $g(x, w)$  は、それぞれ  $R(x, w)$  の面積と重心である。

$$|R(x, w)| = \int_{R(x, w)} dy,$$

$$g(x, w) = \int_{R(x, w)} y dy / |R(x, w)|.$$

しかし、このままでは解  $w(x)$  が、 $x$  に関して非常に細かい変動を起こす可能性がある。一方もとの離散モデルでは  $w$  の変動に素子の間隔以下の周期の成分はない。そこで、変動の高周波数成分を抑えるため、 $w$  のダイナミクスに拡散を付け加える。

$$\partial_t w(x, t) = G(x, w) - S(x, w) w(x, t) + \sigma \Delta w(x, t), \quad (4.14)$$

$$\Delta = \partial^2 / \partial x^2, \quad \sigma > 0.$$

これらの量は、つぎの  $p(x, w) \triangleq (p, q, m, n)^T$  (' $\cdot$ ' は転置を表す) を使って表すことができる。

$$p(x, w) \triangleq w_1(x-D) + w_2(x-D) w_2'(x-D) / w_1(x-D),$$

$$q(x, w) \triangleq w_1(x+D) + w_2(x+D) w_2'(x+D) / w_1(x+D),$$

(4.15)

$$m(x, w) \triangleq -w_2'(x-D) / w_1'(x-D),$$

$$n(x, w) \triangleq -w_2'(x+D) / w_1'(x+D).$$

$1/m$  と  $p$  は、それぞれ点  $w(x-D)$  における曲線  $w(x)$  の法線の傾きと  $y_1$  軸切片、 $1/n$  と  $q$  は点  $w(x+D)$  における法線の傾きと  $y_1$  軸切片である。これらを使うと、

| R(x, w) | と g(x, w) は,

$$| R(x, w) | = 2a(q-p),$$

$$g(x, w) = \left( \frac{\{(q+p)/2\} + \{a^2(n^2-m^2)/12(q-p)\}}{a^2(n-m)/3(q-p)} \right),$$

と表される。この式の導出は付録 1 に示してある。これを使って (4.12), (4.13) の G と S を求めると,

$$S(x, w) = q-p, \quad (4.16)$$

$$G(x, w) = G(p(x, w)) = \left( \frac{\{(q^2-p^2)/2\} + \{a^2(n^2-m^2)/12\} + a^2(n-m)/3}{a^2(n-m)/3} \right), \quad (4.17)$$

となる。

式 (4.14) はつぎの自明な定常解を持つ。

$$w(x, t) = (x, 0)^T, \quad (4.18)$$

$$(m=n=0, p=x-D, q=x+D).$$

これを単純連続解と呼ぶことにする。この節の目的は、平均学習方程式 (4.14) の単純連続解 (4.18) の周りの変分を求め、解 (4.18) の安定性を調べることである。これを次のような手順で求める。まず、式 (4.15) から作用素行列 (dp/dw)。また (4.17) から行列 (dG/dp) を求め、これらに w(x) = (x, 0) を代入する。同様に、(4.16) から作用素行列 (dSw/dp) |\_{w=(x,0)} も求める。最後にこれらの値から変分を計算する。

解 (4.18) からの w の微小な変化を  $\delta w = (\delta w_1, \delta w_2)^T$  として、式 (4.15) より、作用素 (dp/dw) の各要素を求めると、まず  $(\partial p / \partial w_1)$  は、

$$\left( \frac{\partial p}{\partial w_1} \right) \delta w_1(x) = \delta w_1(x-D) - w_2(x-D) w_2'(x-D) \delta w_1'(x-D) / w_1'(x-D)^2,$$

となる。これに、w = (x, 0) を代入すると、

$$\left( \frac{\partial p}{\partial w_1} \right) |_{w=(x,0)} \delta w_1(x) = \delta w_1(x-D), \quad (4.19)$$

となる。同様に、

$$\left( \frac{\partial p}{\partial w_2} \right) \delta w_2(x) = \{ \delta w_2(x-D) w_2'(x-D) + w_2(x-D) \delta w_2'(x-D) \} / w_1'(x-D),$$

$$\left( \frac{\partial q}{\partial w_1} \right) \delta w_1(x) = \delta w_1(x+D) - w_2(x+D) w_2'(x+D) \delta w_1'(x+D) / w_1'(x+D)^2,$$

$$\left( \frac{\partial q}{\partial w_2} \right) \delta w_2(x) = \{ \delta w_2(x+D) w_2'(x+D) + w_2(x+D) \delta w_2'(x+D) \} / w_1'(x+D),$$

$$\left( \frac{\partial m}{\partial w_1} \right) \delta w_1(x) = w_2'(x-D) \delta w_1'(x-D) / w_1'(x-D)^2,$$

$$\left( \frac{\partial m}{\partial w_2} \right) \delta w_2(x) = -\delta w_2'(x-D) / w_1'(x-D)^2,$$

$$\left( \frac{\partial n}{\partial w_1} \right) \delta w_1(x) = w_2'(x+D) \delta w_1'(x+D) / w_1'(x+D)^2,$$

$$\left( \frac{\partial n}{\partial w_2} \right) \delta w_2(x) = -\delta w_2'(x+D) / w_1'(x+D)^2,$$

これらに、w = (x, 0) を代入すると、

$$\left( \frac{\partial p}{\partial w_2} \right) |_{w=(x,0)} \delta w_2(x) = 0, \quad (4.20)$$

$$\left( \frac{\partial q}{\partial w_1} \right) |_{w=(x,0)} \delta w_1(x) = \delta w_1(x+D), \quad (4.21)$$

$$\left( \frac{\partial q}{\partial w_2} \right) |_{w=(x,0)} \delta w_2(x) = 0, \quad (4.22)$$

$$\left( \frac{\partial m}{\partial w_1} \right) |_{w=(x,0)} \delta w_1(x) = 0,$$

$$\left( \frac{\partial m}{\partial w_2} \right) |_{w=(x,0)} \delta w_2(x) = -\delta w_2'(x-D),$$



$$(\partial n / \partial w_1) |_{w=(x, z)} \delta w_1(x) = 0,$$

$$(\partial n / \partial w_2) |_{w=(x, z)} \delta w_2(x) = -\delta w_2'(x+D),$$

を得る。これらをまとめると、

$$\begin{aligned} (dp/dw) |_{w=(x, z)} \delta w(x) \\ = (\delta w_1(x-D), \delta w_1(x+D), -\delta w_2'(x-D), -\delta w_2'(x+D))^T, \end{aligned} \quad (4.23)$$

となる。

つぎに、(4.17)から、行列  $(dG/dp) |_{w=(x, z)}$  を求めると、まず、

$$(dG/dp) = \begin{pmatrix} -p & q & -a^2 m/6 & -a^2 n/6 \\ 0 & 0 & -2a^2/3 & 2a^2/3 \end{pmatrix},$$

これより、

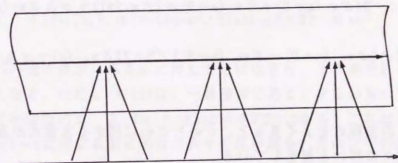
$$(dG/dp) |_{w=(x, z)} = \begin{pmatrix} -x+D & x+D & 0 & 0 \\ 0 & 0 & -2a^2/3 & 2a^2/3 \end{pmatrix}, \quad (4.24)$$

を得る。

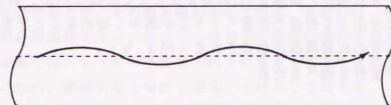
つぎに、(4.16), (4.19)~(4.22)から、 $(dSw/dp) |_{w=(x, z)}$  を求めると、

$$\begin{aligned} (q-p)w_1 \\ (dSw/dp) |_{w=(x, z)} \delta w = d/dw ( \quad ) |_{w=(x, z)} \delta w \\ (q-p)w_2 \\ (\delta w_1(x+D) - \delta w_1(x-D))x + 2D \delta w_1(x) \\ = ( \quad ) \\ \quad \quad \quad 2D \delta w_2(x) \end{aligned} \quad (4.25)$$

最後に、(4.23), (4.24), (4.25)から、 $\delta w$  を支配する変分方程式を求めると、



a)  $w_1$  に関する不安定性



b)  $w_2$  に関する不安定性

図4.9 2種類の不安定性

$w_2$  に関する不安定が生ずると、曲線  $w(x)$  が、 $y_2$  方向にうねり、ハイパー・コラム状の構造が形成され始める。 $w_1$  に関する不安定が生ずると、曲線上に等間隔に並んだ点に向かって付近の荷重ベクトルが集中し始め、コラム状の繊細構造が形成される

$$\frac{1}{2} \delta w(x, t) = \left\{ (dG/dp) (dp/dw) - (dSw/dp) \right\} |_{w=(x, a)} \delta w + \sigma \Delta \delta w(x)$$

$$= \left( \begin{array}{c} D(\delta w_1(x+D) + \delta w_1(x-D) - 2\delta w_1(x)) + \sigma \Delta \delta w_1(x) \\ -2a^2(\delta w_2'(x+D) - \delta w_2'(x-D)) / 3 - 2D\delta w_2'(x) + \sigma \Delta \delta w_2(x) \end{array} \right), \quad (4.26)$$

となる。この方程式をよく見ると、 $\delta w_1$ と $\delta w_2$ に関する方程式が互いに独立しており、別々に解ける形をしている。

これから、この方程式の安定性を調べるが、そのままに、それぞれの不安定性の持つ意味を述べておく。 $w_2$ に関する不安定性は、曲線 $w(x)$ が、 $y_2$ 方向にうねることを意味している。その場合は $x$ が、局所的には $y_2$ 方向の座標、大域的には $y_1$ 方向の座標となり、結果としてハイパー・コラム状の構造が形成される。 $w_1$ に関する不安定性は、曲線 $w(x)$ の形を変えることはないが、曲線上に等間隔に並んだ点に向かって付近の荷重ベクトルが集中し始めることを意味する。この結果コラム状の微相構造が形成される(図4.9)。

(4.26)の安定性を調べるため、

$$\delta w(x, t) = \exp\{\lambda t + i\omega x\} w_0, \quad (4.27)$$

$$w_0 = (w_{01}, w_{02})^T,$$

を(4.26)に代入すると、固有値問題

$$\left( \begin{array}{c} D(\exp\{i\omega D\} + \exp\{-i\omega D\} - 2) - \sigma \omega^2 \\ \{-2i\omega a^2/3\}(\exp\{i\omega D\} - \exp\{-i\omega D\}) - 2D - \sigma \omega^2 \end{array} \right) w_{01} \\ \left( \begin{array}{c} 0 \\ (4\omega a^2/3)\sin\omega D - 2D - \sigma \omega^2 \end{array} \right) w_{02} = 0$$

を得る。この固有値問題は自明に解ける。すなわち固有ベクトル $e_1 = (1, 0)$ 、 $e_2 = (0, 1)$ と、それぞれに対応した固有値 $\lambda_1$ 、 $\lambda_2$ を得る。

$$\lambda_1(\omega, D, \sigma) = 2D(\cos\omega D - 1) - \sigma \omega^2 \\ \lambda_2(\omega, a, D, \sigma) = (4\omega a^2/3)\sin\omega D - 2D - \sigma \omega^2 \quad (4.28)$$

この二つの固有値が、ある $\omega$ に対して正になると、その周波数の外乱に対して不安定となる。ただし(4.14)は、一樣構造である。すなわち $y_1$ 方向の平行移動に関して不変なので、 $\lambda_1(\omega, D, \sigma)$ は $\omega=0$ で0になる。これは、固有値 $\lambda_1(0, D, \sigma)=0$ に対する固有関数の表す外乱が解を $y_1$ 方向に平行移動させることを意味しているからである。(4.14)のすべての平衡解は、 $y_1$ 方向に平行移動に対して“無抵抗”であり、このような系における解の安定性は、位置を変化させながらもその波形のみ保持しようとする性質と考えなければならない。このような安定性を波形安定と呼ぶ。この場合、 $\lambda_1$ は $\omega=0$ のときを除いて他はいつも負、また $\lambda_2$ は任意の $\omega$ に対して負である。逆に、 $\lambda_1$ 、 $\lambda_2$ に関してこれらの条件が成立するとき、その解を線形安定と呼ぶ。正確には、線形安定性は波形安定性の必要条件であり、従って、線形不安定性は波形不安定性の十分条件である。本章では、もっぱら不安定性の方に焦点を当てて考察を進めるので、以後簡単のため、線形安定を単に“安定”と呼ぶことにする。

(4.28)の場合には、まず、 $\lambda_1$ 、 $\lambda_2$ は両方共 $\omega$ の偶関数なので、 $\omega \geq 0$ のときだけ考えればよい。 $\lambda_1$ は、パラメタ $a$ 、 $D$ 、 $\sigma$ をどんな値にとっても、 $\omega=0$ のときを除いて他はいつも負である。これは、コラム構造が形成されないことを意味している。 $\lambda_2$ は、

$$\sin\omega D > (3D/2a^2) \{ (1/\omega) + (\sigma\omega/2D) \} \quad (4.29)$$

のとき正になる。この式の右辺は、 $a > 0$ を大きくしていけば、0に近付き、 $a$ を0に近付ければ無限大に発散する。よって、 $a$ が十分小さいときは(4.18)は安定であるが、 $\sigma$ と $D$ を固定して、 $a$ をしだいに大きくしていくと、ある限界値 $a^*(D, \sigma)$ を越えたとき解(4.18)は不安定になり、うねり始める(図4.10)。 $a^*(D, \sigma)$ に関しては次の定理が得られる。

定理4.1

$$(3D/2a^2) \{ (1/\omega) + (\sigma\omega/2D) \}$$

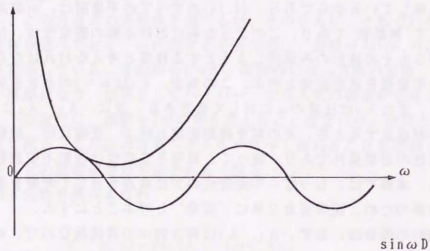


図 4.10 単純連続解の安定条件

$a$ が、だんだん大きくなり、単純連続解の安定性が保たれなくなるぎりぎりの値 $a^*$ に達したときのようすを示す。

$a^*(D, \sigma)$ に関して、次の不等式が成り立つ。

$$(3/2)^{1/2} (\sigma D)^{1/4} \geq a^*(D, \sigma).$$

等号が成立するのは、

$$\sqrt{(2D^3/\sigma)} = (2n+0.5)\pi, \quad n=0, 1, 2, \dots,$$

のときである。

証明 (4.29)式の右辺の $\omega$ に関する変化を見ると、 $\omega = \sqrt{(2D/\sigma)}$ のとき最小値 $(3/2a^2)\sqrt{(\sigma D)}$ をとることがわかる。この値が1より大きければ $\lambda_2$ はいつも負である。これより定理の不等式が導かれる。等号が成立するためには、最小値を与える $\omega$ で(4.29)式の左辺が1になっていなければならない。それが、定理で与えられている条件に他ならない。 q. e. d.

#### 4.2.3.2. 拡散を含まないモデル

ここで拡散を含まないモデルについて考えてみよう、前節で最後に外乱の高周波数成分を抑えるためにモデルに拡散項を付け加えたが、この影響は(4.28)の $\lambda_1, \lambda_2$ のなかの $-\sigma\omega^2$ に現れている。もし $\sigma=0$ とにおいて、拡散の効果を取り除くと、固有値は、

$$\lambda_1(\omega, D, 0) = 2D(\cos \omega D - 1),$$

(4.30)

$$\lambda_2(\omega, a, D, 0) = (4\omega a^2/3)\sin \omega D - 2D,$$

を得る。これらの固有値を解析することによって、出力層の素子の数を無限に多くしていった極限の状況を知ることができる。出力層の素子の数を増やしていくと幾らでも高い周波数の外乱を考えに入れることになるからである。

まず、 $\lambda_1$ は、すべての $\omega$ に対して0以下であるが、 $\omega = 2n\pi/D$ で0になり、これらの周波数の外乱に対しては不安定ぎりぎりであることがわかる。 $\lambda_2$ の方からは、どんなパラメタに対しても必ず不安定になることがわかる。これは、帯状信号空間がいくら細くても、出力層の素子の数をどんどん増やしていくと解(4.18)はいつか必ず不安定になり、ハイパー・コラム状の構造を作るようになることを示している。

#### 4.2.3.3 出力層にアナログ素子を使った場合

今まで扱ったKohonenのモデルでは、出力層の各素子に対してそれぞれの近傍が定義されており、ある素子が最大の入力を受け取ると、その素子の近傍内の素子が等しい強さで発火し、等しい強さで学習がおこなわれる。しかし、モデルをより現実に近いものにするためには、出力層の細胞の発火レベルはアナログ値をとると考えた方がよい。ここでは簡単のため、素子の発火レベルは、最大の入力を受け取った素子で1、その周囲の素子では、最大の入力を受け取った素子からの距離 $\xi$ に応じて決まる関数 $r(\xi)$ で与えられるとする。関数 $r(\xi)$ は、正数 $\xi$ に対して定義され、

$$r(\xi) = r(-\xi), \quad r(0) = 1,$$

を満たす関数である。通常は図4.1.1のa)のような、 $|\xi|$ に関して単調減少な関数を用いるが、b)のように発火領域の周りに負の抑制領域を持つものも考えることもできる。

この発火レベルに比例してHebb学習が起こると仮定すると、(4.11)の代わりに、

$$w(x) := w(x) + e r(x - x^*) (y - w(x)), \quad (4.31)$$

$$x^* \triangleq \arg \min_x \{ \|y - w(x)\| \}. \quad (4.10)$$

となる。 $r$ を、

$$r(\xi) = r_0(\xi) \triangleq \begin{cases} 1, & |\xi| \leq D, \\ 0, & |\xi| > D, \end{cases}$$

とすれば、(4.31)は(4.11)に一致する。

このモデルも(4.11)と同様に一様構造なので、その平均学習方程式は単純連統解(4.18)を持つ。

$r(\xi)$ をいろいろな $D$ の $r_D(x)$ の重ね合わせと考えれば、(4.31)の右辺は、いろいろな $D > 0$ に関して(4.11)の右辺を重み $-r'(D)$ をつけて積分したものであることがわかる。学習方程式から平均学習方程式を求め、その変分方程式に(4.27)

の形の固有関数を代入して2次元の固有値問題に帰着するオペレーションは線形である。また、すべての $D$ に関して、この2次元の固有値問題の固有関数は共通なので、固有値も各々の場合の重ね合わせである。よって、このモデルの単純連統解の安定性を支配する固有値は(4.28)の固有値を重み $-r'(D)$ をつけて $D$ で積分することによって得られる。

$$\begin{aligned} \lambda_1(\omega, r(x), \sigma) &= -\int_0^\infty \lambda_1(\omega, x, \sigma) r'(x) dx \\ &= -\int_0^\infty 2x(\cos \omega x - 1) r'(x) dx - \sigma \omega^2, \end{aligned} \quad (4.32)$$

$$\begin{aligned} \lambda_2(\omega, a, r(x), \sigma) &= -\int_0^\infty \lambda_2(\omega, a, x, \sigma) r'(x) dx \\ &= -\int_0^\infty \{(4\omega a^2/3) \sin \omega x - 2x\} r'(x) dx - \sigma \omega^2. \end{aligned}$$

この式から直ちにわかるのは、 $x = 2n\pi$ 以外のときに $r'(x) < 0$ である場合は、拡散のない場合でも、すべての $\omega$ に対して $\lambda_1 < 0$ となり、(4.30)のように $\omega = 2n\pi$ に対して不安定ぎりぎりというようなことがなくなるのである。さらに詳しく見るために二つの例に関する計算を示す。

#### 例4.1

$$r(\xi) = \begin{cases} 1 - |\xi|/D, & |\xi| \leq D, \\ 0, & |\xi| > D, \end{cases} \quad (4.33)$$

の場合に(4.32)を具体的に計算すると、

$$\begin{aligned} \lambda_1 &= (1/D) \int_0^\infty 2x(\cos \omega x - 1) dx - \sigma \omega^2 \\ &= \sqrt{(D^2 \omega^2 + 1) |\sin(D\omega + 1) - \sqrt{(D^2 \omega^2 + 1)}|} / D \omega^2 - \sigma \omega^2, \\ \lambda_2 &= (1/D) \int_0^\infty \{(4\omega a^2/3) \sin \omega D - 2D\} dx - \sigma \omega^2 \\ &= (4a^2/3D) (1 - \cos \omega D) - D - \sigma \omega^2, \end{aligned}$$

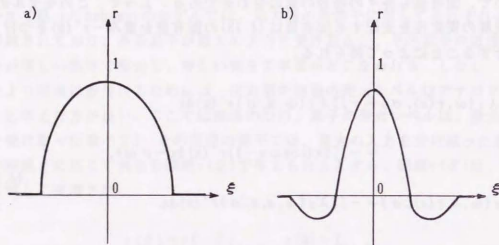


図 4.1.1 アナログ素子の発火パターンを決める関数  $r(\xi)$  の例

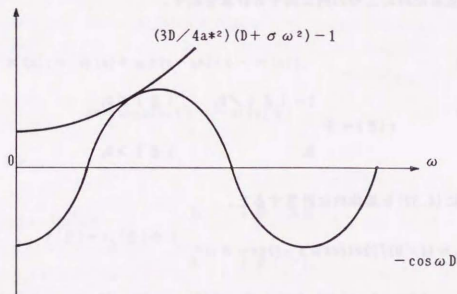


図 4.1.2 アナログ素子を使った場合の単純連続解の安定条件

$a^*$ が、だんだん大きくなり、単純連続解の安定性が保たれなくなるぎりぎりの値  $a^*$  に達したときのようすを示す。

となり、 $\lambda_1$  は、拡散のない場合 ( $\sigma=0$ ) を含めて、 $\omega=0$  のとき以外はいつも負である。

$\lambda_2$  が正になるのは、

$$-\cos \omega D > (3D/4a^2)(D + \sigma \omega^2) - 1,$$

が成立するときである。この式の左辺は  $\omega=0$  で最小値  $3D^2/4a^2 - 1$  をとる。一方右辺の最大値は 1 (図 4.1.2)。これから次の定理を得る。

定理 4.2

発火パターンが (4.33) で与えられるモデルの単純連続解が安定である限界の  $a$  の値  $a^*(D, \sigma)$  は、次の式を満たす。

$$a^*(D, \sigma) > \sqrt{(3/8)D}$$

01素子のモデルの場合、拡散を考えないときは  $y_2$  方向の不安定は必ず起こったが、この定理は、アナログ素子のモデルの場合、拡散を含まない場合でも、 $a$  が小さければ単純連続解は安定であることを示している。

$a$  が  $a^*(D, \sigma)$  を越えて大きくなると最初に不安定を起こす周波数  $\omega^*(D, \sigma)$  は、拡散のあるときは、

$$0 < \omega^*(D, \sigma) < \pi/D, \quad D > 0,$$

$$\omega^*(D, \sigma) \rightarrow \pi/D, \quad (\sigma \rightarrow 0),$$

であるが、 $\sigma=0$  のときは、 $\omega^* = (2n-1)\pi/D$ ,  $n=1, 2, \dots$  で同時に不安定になる。

例 4.2 (発火パターンが抑制領域を持つ場合)

$$1, \quad |\xi| \leq D,$$

$$r(\xi) = \begin{cases} -s, & D < |\xi| \leq 2D, \quad s > 0, \\ 0, & 2D < |\xi|, \end{cases} \quad (4.33)$$

$$0, \quad 2D < |\xi|,$$

の場合の固有値を計算すると、

$$\begin{aligned}\lambda_1 &= 2(1+s)D(\cos\omega D - 1) - 4sD(\cos 2\omega D - 1) - \sigma\omega^2 \\ &= 2D(\cos\omega D - 1)(1 - 3s - 4s\cos\omega D) - \sigma\omega^2, \\ \lambda_2 &= (4\omega a^2/3)\{(1+s)(\sin\omega D - 2D) - s(\sin 2\omega D - 4D)\} - \sigma\omega^2 \\ &= (4\omega a^2/3)\{(1+s)\sin\omega D - s\sin 2\omega D + 2D(s-1)\} - \sigma\omega^2,\end{aligned}$$

(4.28)の場合と同様に $\lambda_2$ は、任意のパラメタ $a, s, D, \sigma$ に関して、ある $\omega$ において正となり、ハイパー・コラムは必ず生成される。 $\lambda_1$ は(4.28)の場合と異なり、あるパラメタに対して正になる。特に、拡散のない場合は、

$$s > 1/7,$$

の場合に、コラム構造が形成される。

#### 4.2.3.4 円柱状3次元領域の信号空間から1次元の神経場へのトポグラフィック・マッピング

本節では、信号空間を円柱状の3次元領域

$$Y = R \times D^2(a), \quad D^2(a) \triangleq \{(y_2, y_3) \mid y_2^2 + y_3^2 \leq a\},$$

から、1次元の神経場 $X = R$ へのトポグラフィック・マッピングを4.2.3.1節と同様な方法で解析する。

入力層は3個の素子からなり、そこに与えられる入力 $y = (y_1, y_2, y_3)$ は、 $Y$ のなかから一様分布に従って、ランダムに選ばれる。

$$y_1 \in R, \quad (y_2, y_3) \in D^2(a).$$

(4.11)のモデルをこの場合に適用し、平均学習方程式を求めて、前節と同様、外乱の高周波数成分を抑えるために拡散項を加えると、

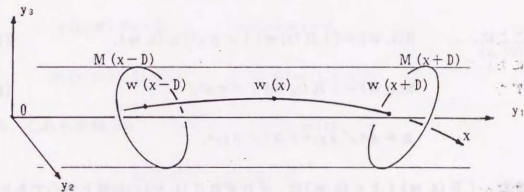


図4.13 円柱状3次元領域から1次元の神経場へのマッピング

この図は図4.13の3次元版である。曲線 $w(x)$ のうねりが十分小さければ、受容野 $R(x, w)$ は、点 $w(x-D)$ と点 $w(x+D)$ において曲線 $w(x)$ に直交する平面が $Y$ から切り取る立体となる。

$$\oint w(x, t) = \int_{R(x, w)} \{(y - w(x)) / \pi a^2\} dy + \sigma \Delta w(x)$$

$$= (1 / \pi a^2) \int_{R(x, w)} y dy - (1 / \pi a^2) \int_{R(x, w)} w(x) dy + \sigma \Delta w(x)$$

$$= (|R(x, w)| / \pi a^2) g(x, w) - (|R(x, w)| / \pi a^2) w(x) + \sigma \Delta w(x)$$

$$= G(x, w) - S(x, w) w(x) + \sigma \Delta w(x), \quad (4.34)$$

ここに,  $G(x, w) = (|R(x, w)| / \pi a^2) g(x, w), \quad (4.35)$

$$S(x, w) = |R(x, w)| / \pi a^2, \quad (4.36)$$

$$\Delta \triangleq \partial^2 / \partial y_1^2 + \partial^2 / \partial y_2^2,$$

また,  $|R(x, w)|$  と  $g(x, w)$  は, それぞれ  $R(x, w)$  の体積と重心である.

$$|R(x, w)| = \int_{R(x, w)} dy,$$

$$g(x, w) = \int_{R(x, w)} y dy / |R(x, w)|.$$

曲線  $w(x)$  のうねりが十分小さければ, 受容野  $R(x, w)$  は, 点  $w(x-D)$  と点  $w(x+D)$  において曲線  $w(x)$  に直交する平面が  $Y$  から切り取る立体である (図 4.13).

(4.35) は, つぎの自明な定常解を持つ.

$$w(x, t) = (x, 0, 0)^T. \quad (4.37)$$

$G(x, w)$  と  $S(x, w)$  は, つぎの  $p(x, w) \triangleq (p, q, m_2, m_3, n_2, n_3)^T$  を使って表すことができる.

$$p(x, w) \triangleq (w'(x-D), w(x-D)) / w_1'(x-D)$$

$$q(x, w) \triangleq (w'(x+D), w(x+D)) / w_1'(x+D) \quad (4.38)$$

$$m(x, w) \triangleq (1, m_2, m_3)^T \triangleq -w'(x-D) / w_1'(x-D),$$

$$n(x, w) \triangleq (1, n_2, n_3)^T \triangleq -w'(x+D) / w_1'(x+D).$$

$m(x, w)$  と  $p$  は, それぞれ点  $w(x-D)$  において曲線  $w(x)$  に直交する平面の  $y_1$  軸切片とその点における曲線の接ベクトル,  $n(x, w)$  と  $q$  は, それぞれ点  $w(x+D)$  において曲線  $w(x)$  に直交する平面の  $y_1$  軸切片とその点における曲線の接ベクトルである. 解 (4.37) の場合は

$$p(x, w) = x-D, \quad q(x, w) = x+D, \quad (4.39)$$

$$m(x, w) = (1, 0, 0)^T, \quad n(x, w) = (1, 0, 0)^T,$$

となる. これらを使うと,  $|R(x, w)|$  と  $g(x, w)$  は,

$$|R(x, w)| = \pi a^2 (q-p),$$

$$- (m_2^2 + m_3^2 - n_2^2 - n_3^2) / 2 \quad (4.40)$$

$$g(x, w) = r^2 / 4 (q-p) \left( \begin{array}{c} m_2 - n_2 \\ m_3 - n_3 \end{array} \right)$$

$$(q+p) / 2$$

$$+ \left( \begin{array}{c} 0 \\ 0 \end{array} \right),$$

$$0$$

と表される. この式の導出は付録 2 に示してある. これを使って, (4.35), (4.36) の  $S, G$  を求めると,

$$S(x, w) = q - p, \quad (4.41)$$

$$-(m_2^2 + m_3^2 - n_2^2 - n_3^2)/2$$

$$G(x, w) = r^2/4 ( \quad m_2 - n_2 \quad )$$

$$m_3 - n_3$$

$$(q^2 - p^2)/2$$

$$+ ( \quad 0 \quad ), \quad (4.42)$$

$$0$$

となる。

解(4.18)からの $w$ の微小な変化を $\delta w = (\delta w_1, \delta w_2)^T$ として、式(4.38)より、作用素 $(dp/dw)$ の各要素を求めると、まず $(\partial p/\partial w)$ は、

$$\begin{aligned} (\partial p/\partial w) \delta w(x) &= \{ (\delta w(x-D), w'(x-D)) \\ &\quad + (w'(x-D), \delta w(x-D)) w_1'(x-D) \\ &\quad - (w(x-D), w'(x-D)) \delta w_1'(x-D) \} / w_1'(x-D)^2, \end{aligned}$$

となる。これに、単純連続解 $w = (x, 0, 0)^T$ を代入すると、

$$(\partial p/\partial w) |_{w=(x, 0, 0)} \delta w(x) = \delta w_1(x-D), \quad (4.43)$$

となる。

同様に、

$$(\partial q/\partial w) |_{w=(x, 0, 0)} \delta w(x) = \delta w_1(x+D), \quad (4.44)$$

$$\begin{aligned} (\partial m_2/\partial w) \delta w(x) \\ = \{ w_2'(x-D) \delta w_1'(x-D) + \delta w_2'(x-D) w_1'(x-D) \} / w_1'(x-D)^2, \end{aligned}$$

これより、

$$(\partial m_2/\partial w) \delta w(x) |_{w=(x, 0, 0)} \delta w(x) = \delta w_2'(x-D), \quad (4.45)$$

を得る。以下同様に、

$$(\partial m_3/\partial w) \delta w(x) |_{w=(x, 0, 0)} \delta w(x) = \delta w_3'(x-D), \quad (4.46)$$

$$(\partial n_2/\partial w) \delta w(x) |_{w=(x, 0, 0)} \delta w(x) = \delta w_2'(x+D), \quad (4.47)$$

$$(\partial n_3/\partial w) \delta w(x) |_{w=(x, 0, 0)} \delta w(x) = \delta w_3'(x+D), \quad (4.48)$$

を得る。

つぎに、(4.42)から、行列 $(dG/dp) |_{w=(x, 0, 0)}$ を求めると、まず、

$$\begin{aligned} & \begin{matrix} -p & -q & -\mu m_2 & -\mu m_3 & \mu n_2 & \mu n_3 \end{matrix} \\ (dG/dp) &= \begin{pmatrix} 0 & 0 & -\mu & 0 & -\mu & 0 \\ 0 & 0 & 0 & \mu & 0 & \mu \end{pmatrix} \end{aligned}$$

これより、

$$\begin{aligned} & (dG/dp) |_{w=(x, 0, 0)} \\ & \begin{matrix} -x+D & x+D & -\mu m_2 & -\mu m_3 & \mu n_2 & \mu n_3 \end{matrix} \\ &= \begin{pmatrix} 0 & 0 & -\mu & 0 & -\mu & 0 \\ 0 & 0 & 0 & \mu & 0 & \mu \end{pmatrix}, \quad (4.49) \end{aligned}$$

を得る。ここに、 $\mu = a^2/4$ である。

つぎに、(4.41), (4.43), (4.44)から、 $(dSw/dp) |_{w=(x, 0, 0)}$ を求めると、



$$(dSw/dp) |_{w=(x, \theta, \theta)} \delta w(x) = d/dw ((q-p)w) |_{w=(x, \theta, \theta)} \delta w$$

$$(\delta w_1(x+D) - \delta w_1(x-D))x + 2D \delta w_1(x)$$

$$= ( \quad 2D \delta w_2(x) \quad ) ,$$

$$2D \delta w_3(x) \quad (4.50)$$

最後に、(4.43)~(4.48)、(4.49)、(4.50)から、 $\delta w$ を支配する変分方程式を求めると、

$$\delta \delta w(x, t)$$

$$D(\delta w_1(x+D) + \delta w_1(x-D) - 2\delta w_1(x)) + \sigma \Delta \delta w_1(x)$$

$$= ( - (a^2/4)(\delta w_2'(x+D) - \delta w_2'(x-D)) / 3 - 2D \delta w_2'(x) + \sigma \Delta \delta w_2(x) ) ,$$

$$- (a^2/4)(\delta w_3'(x+D) - \delta w_3'(x-D)) / 3 - 2D \delta w_3'(x) + \sigma \Delta \delta w_3(x) \quad (4.51)$$

となる。この方程式をよく見ると、3本の方程式はすべて互いに独立しており、 $\delta w_1$ に関する方程式は2次元の場合と同じ、 $\delta w_2$ 、 $\delta w_3$ に関する方程式も、一部の係数を除いて2次元の場合と同じである。よって以後の解析は、拡散項のない場合、アナログ素子の場合を含めて2次元の場合と同様である。

#### 4.2.3.5 発火パターンが変化するKohonenのモデル

##### 4.2.5.3.1 1次元から1次元への場合

前節までで取り扱ったKohonenのモデルでは、出力層に起きる発火パターンの波形 $r(\xi)$ は、最大入力を受け取っている素子の周りの素子への入力の大きさに依らず一定であった。しかし、この発火パターンは出力層内の相互結合によ

て保持される孤立局在興奮なのであるから、本来、入力によって変化しうものなのである。もっとも、これを変化させるようなモデルを作っても、それによって新たに説明できる現象がなければ、そのモデルはいたずらに複雑なだけであって、それをあらためて提案する意味はない。本節ではKohonenのモデルを、発火パターンが変化するよう改良し、このモデルでは、いままでのKohonenのモデルでは不可能だったコラム状の微細構造の形成が起きることを示す。

ここでコラム状の微細構造が形成されるメカニズムについて直観的に考えてみよう。2次元の信号空間から1次元の神経場へのトポグラフィック・マッピングの場合、コラム状の微細構造が形成されるためには、 $\delta y_i$ に関する方程式の固有値 $\lambda_i$ がある $\omega$ において正になり、図4.9のa)のような外乱が生じたとき、それがさらに助長されなければならない。図4.9のa)の場合、信号空間 $Y$ において信号 $y_i$ は荷重ベクトル $w$ が集中している部分にあったとしよう。この信号が与えられると、最大入力を受け取る $x^*(y_i, w)$ 付近の素子は $w(x^*)$ に近い荷重ベクトルを持っているのであるから、それらの素子も場所 $x^*$ の素子と殆ど同じ入力を受けている。いいかえれば、信号 $y_i$ は $x^*$ を中心とした広い範囲の素子に大きな入力を与えている。このような場合の孤立局在興奮は、範囲も広く興奮度も強いはずである。すると、そのとき発火した素子の荷重ベクトルは、Hebb学習によって、より強く $y_i$ に引き寄せられ、こうして $y_i$ 付近にはますます荷重ベクトルが集中することになる。

Amaríのモデルは、孤立局在興奮を保持する神経興奮のダイナミクスをそのままモデルに組み込んでいるので、モデルは複雑だが、 $y_i$ のような入力に対する孤立局在興奮の幅が広がる効果を取り入れることに成功している。ただし数学的解析を可能にするため、素子の出力を1または0の2値とするので、興奮の強さの変化はモデルに取り入れられていない。一方Kohonenのモデルでは、孤立局在興奮のパターンを一種類に限ったために、モデルは簡単になったが、コラム構造を形成することができなくなってしまった。

そこで、Kohonenのモデルの簡単さを損なわないようにしながら、興奮パターンの変化を取り入れた新しいモデルを提案する。このモデルでは、孤立局在興奮を表す興奮パターンの幅は変化しないが興奮度が変化する。信号 $y$ に対し、荷重ベクトル $w$ を持つ素子は、3章の(3.32)式と同じく、入力 $h(\|y - w\|)$ を受け取る。ここに $h$ は、非負の実数 $x$ に対して定義された、

$$h(0) = 1, \quad h'(x) < 0, \quad \text{for } \forall x > 0, \quad (4.52)$$

を満たす関数である(図4.14)。 $h(\|y_1 - y_2\|)$ は、 $y_1$ と名付けられた信

号から最も大きな入力を受け取る細胞が、 $y_2$ と名付けられた信号からどの程度の入力を受け取るかを定義する関数である。従って、信号と重みベクトルが球面上に分布し、入力が両者の内積で与えられるもとのKohonenのモデルにかえて考えると、 $y_1$ と名付けられた信号と $y_2$ と名付けられた信号の内積あるいは相関にあたる量である。興奮の領域は従来と同じく、

$$N_0(x) = \{\xi \in X \mid \|\xi - x\| \leq D\}, \quad (4.53)$$

とする。入力 $y$ に対しては場所、

$$\begin{aligned} x^*(y, w) &\triangleq \operatorname{argmax}_x \{h(\|y - w\|)\}, \\ &= \operatorname{argmin}_x \{\|y - w(x)\|\}, \end{aligned}$$

の素子の近傍 $N_0(x^*(y, w))$ が発火する。これも今までのモデルと同じである。

近傍内の全ての素子は等しい強さで発火するが、その強さは近傍内の素子の受け取る入力 $h$ の総和によって決まると考える、すなわち発火の強さ $A(y, w)$ は

$$A(y, w) \triangleq F(\int_{N_0(x^*(y, w))} h(\|y - w(x)\|) dx) \quad (4.54)$$

によって与えられる。ここに $F$ は、

$$F(u) \geq 0, \quad F'(u) \geq 0,$$

を満たす関数である(図4.1.4)。 $F(u) \equiv 1$ とすると、モデル(4.9)と一致する。信号 $y$ に対する学習は、 $y$ の引き起こす発火領域に属する素子の荷重ベクトルに、その発火の強さに比例した変化を起こす。

$$\begin{aligned} w(x) + \epsilon A(y, w)(y - w(x)), \quad |x - x^*| \leq D, \\ w(x) := \begin{cases} w(x), & |x - x^*| > D, \end{cases} \end{aligned} \quad (4.55)$$

$$x^* = x^*(y, w) \triangleq \operatorname{argmin}_x \{\|y - w(x)\|\}. \quad (4.56)$$

以後、4.2.3.5節では、コラム構造の形成に焦点をあてるので、さしあたって

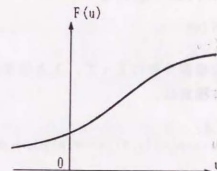
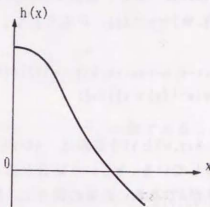


図4.1.4 出力層の細胞への入力 $h$ と発火の強さ $F$ (本文参照)

信号空間は多次元である必要はない。そこで、信号空間はRであるとして話を進める。便宜上  $h(-y) \triangleq h(y)$  ( $y < 0$ ) により、 $h$ をR全域で定義された偶関数であるとする。考慮する解  $w(x): R \rightarrow R$ を、連続で単調増加かつ上にも下にも有界でないものに限ると、 $x^*(y, w) = w^{-1}(y)$ 、が成り立つ。よってAは、

$$\begin{aligned} A(y, w) &= F \left( \int_{w^{-1}(y)-D}^{w^{-1}(y)+D} h(y-w(v)) dv \right) \\ &= F \left( \int_0^D h(y-w(w^{-1}(y)+v)) dv \right), \end{aligned} \quad (4.57)$$

と書き表すことができる。 $A(y, w)$ という表現は、Aがyという実数とwという関数によって決まることを意味している。wという結合があるところへ、信号yがきたときにおこる興奮の強さがAである。計算の都合上、関数Bを、

$$B(x, w) \triangleq F \left( \int_0^D h(w(x)-w(x+v)) dv \right), \quad (4.58)$$

と定義して、

$$A(y, w) = B(w^{-1}(y), w), \quad (4.59)$$

と表しておく。

信号空間Y = R上の一様な確率分布によって、入力信号yがランダムに選ばれるとき、(4.55)の平均学習方程式は、

$$\dot{w}(x, t) = \int_{w(x-D)}^{w(x+D)} A(y, w) (y-w(x)) dy + \sigma \Delta w(x) \quad (4.60)$$

となる。今までと同様に拡散項をつけておく。

wの変分を  $\delta w$  として、この式の変分方程式を求めると、

$$\begin{aligned} \dot{\delta w}(x, t) &= A(w(x+D), w) (w(x+D) - w(x)) \delta w(x+D) \\ &\quad - A(w(x-D), w) (w(x-D) - w(x)) \delta w(x-D) \\ &\quad - \int_{w(x-D)}^{w(x+D)} A(y, w) \delta w(x) dy \\ &\quad + \int_{w(x-D)}^{w(x+D)} (\partial A(y, w) / \partial w) \delta w \cdot (y-w(x)) dy \\ &\quad + \sigma \Delta \delta w(x), \end{aligned} \quad (4.61)$$

となる。この式の右辺の4番目の項の  $(\partial A(y, w) / \partial w)$  は  $A(y, w)$  を関数wで偏微分したものであり、作用素として次の  $\delta w$  に作用している。ここに単純連続解

$w(x, t) = I(x) = x$  ( $I$ は恒等関数をあらわす) を代入して、単純連続解の安定性を調べたい。そのためにまず、 $A(y, I)$ と求めると、(4.58)、(4.59)より、

$$\begin{aligned} A(y, I) &= B(y, I) \\ &= F \left( \int_0^D h(y - (y+v)) dv \right) \\ &= F \left( \int_0^D h(v) dv \right), \\ &= F(2 \int_0^D h(v) dv), \end{aligned}$$

を得る。これは、 $B(x, I)$ が、xによらない数であることを示している。これを  $\alpha(D)$  とおく。

$$\alpha(D) = B(x, I) = F(2 \int_0^D h(v) dv). \quad (4.62)$$

これより、

$$\partial B(x, I) / \partial x = 0, \quad (4.63)$$

を得る。つぎに、 $(\partial A(y, w) / \partial w)$  を求めると、(4.59)より、

$$\begin{aligned} (\partial A(y, w) / \partial w) \delta w &= (dB(w^{-1}(y), w) / dw) \delta w \\ &= \partial B(w^{-1}(y), w) / \partial x \delta(w^{-1}(y)) \\ &\quad + (\partial B(w^{-1}(y), w) / \partial w) \delta w. \end{aligned} \quad (4.64)$$

ここで  $(\partial B(w^{-1}(y), w) / \partial w)$  は作用素である。(4.58)から、これを求めると、

$$(\partial B(w^{-1}(y), w) / \partial w) \delta w = F' \cdot \int_0^D h' \cdot (\delta w(x) - \delta w(x+v)) dv,$$

となる。ただし、ここで、 $F'$ 、 $h'$ はそれぞれ、

$$F' = F' \left( \int_0^D h(w(x) - w(x+v)) dv \right),$$

$$h' = h'(w(x) - w(x+v)),$$

である。ここに、 $w = I$ を代入して、 $(\partial B(w^{-1}(y), I) / \partial w) \delta w$ を求めると、まず  $F'$ は、

$$F' = F' \left( \int_0^D h(v) dv \right) \\ = F' \left( 2 \int_0^D h(v) dv \right).$$

これは、 $x$ によらない定数なので  $\beta(D)$  と書くことにする。

$$\beta(D) = F' \left( 2 \int_0^D h(v) dv \right).$$

また、 $w=1$ のとき  $h'$  は、

$$h' = h'(-v) = -h'(v),$$

である。よって、

$$(\partial B(w^{-1}(y), I) / \partial w) \delta w \\ = -\beta(D) \int_0^D \int_0^D h'(v) (\delta w(x) - \delta w(x+v)) dv, \quad (4.65)$$

を得る。(4.64)に  $w=1$ を代入して、(4.63)、(4.65)をつかうと、

$$(\partial A(y, I) / \partial w) \delta w = \beta(D) \int_0^D -h'(v) (\delta w(x) - \delta w(x+v)) dv, \quad (4.66)$$

を得る。ここで、いよいよ(4.61)に  $w=1$ を代入して、(4.62)、(4.66)をつかうと、 $w=1$ の周りの変分方程式、

$$\begin{aligned} \delta \delta w(x, t) = & \alpha(D) D (\delta w(x+D) + \delta w(x-D) - 2\delta w(x)) \\ & + \beta(D) \int_0^D \int_0^D -h'(v) (\delta w(x+y) - \delta w(x+y+v)) y dv dy \\ & + \sigma \Delta \delta w(x), \end{aligned} \quad (4.67)$$

を得る。これに  $\delta w(x, t) = \exp\{\lambda t + i\omega x\}$ を代入すると、

$$\begin{aligned} \lambda(\omega, D, \sigma) = & 2\alpha(D) D (\cos \omega D - 1) \\ & + 4\beta(D) (\sin \omega D / \omega^2 - D \cos \omega D / \omega) \int_0^D -h'(v) \sin \omega v dv \\ & - \sigma \omega^2, \end{aligned} \quad (4.68)$$

と固有値が得られる。

前述のように、 $F(u) \equiv 1$ とすれば本節のモデルは、通常のモデル(4.9)に一致

するが、その場合は、 $\alpha(D) \equiv 1$ 、 $\beta(D) \equiv 0$ となり、ここで得られた固有値は(4.28)の  $\lambda_1(\omega, D, \sigma)$  に一致する。

一つの例について計算してみよう。

#### 例4.3

$F(u) = \max(0, u)$ 、 $h(v) = \tau(D^2 - v^2)/2$ 、 $\tau > 0$ とする。このとき、

$$\alpha(D) = 2\tau D^3/3, \quad \beta(D) = 1,$$

$$\begin{aligned} \lambda(\omega, D, \sigma) = & (4\tau D^4/3) (\cos \omega D - 1) \\ & + 4\tau (\sin \omega D / \omega^2 - D \cos \omega D / \omega)^2 - \sigma \omega^2 \end{aligned} \quad (4.69)$$

$$= -\sigma \omega^2 - (1/30)\tau \omega^4 D^6 + O(\omega^8),$$

となる。この固有値は、 $\lambda(2\pi/D, D, \sigma) = \tau D^4 / \pi^2 - 4\sigma \pi^2 / D^2$ となるので、拡散がなければ、 $D > 0$ 、 $\tau > 0$ の値に依らず、いつも不安定であり、コラム構造が形成される。拡散のある場合は、

$$\tau > 4\sigma \pi^4 / D^6, \quad (4.70)$$

が、不安定のためのひとつの十分条件になる。

#### 4.2.3.5.2 2次元から2次元への場合

前節で示した、発火パターンが変化するKohonenのモデルにおける単純連続解の不安定性を信号空間と神経場の両方が2次元の場合に付いても示しておく。

$X = Y = \mathbb{R}^2$ 、とし、神経場上の  $x = (x_1, x_2)$  が最大の入力を受け取ったとき、同時に発火する領域  $N_0(x)$  は、 $x$  を中心とする半径  $D$  の円内とする。

$$N_0(x) \triangleq \{x' \in \mathbb{R}^2 \mid \|x - x'\| < D\}. \quad (4.71)$$

これを  $B(D, x)$ 、 $B(D, 0)$  を  $B(D)$  と書くことにする。その他は、1次元の場合と同じである。

$x$  の受容野  $R(x, w) = w(N_0(x))$  は、単純連続解  $w(x) = 1(x) = x$  の場合、 $N_0(x)$  と同じになるが、結合状態がこの解から僅かにずれると、受容野の形も変化する。この形を、 $x$  から見た  $\theta$  方向の境界までの距離  $r(\theta, x, w)$  で表すこ

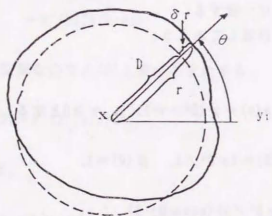
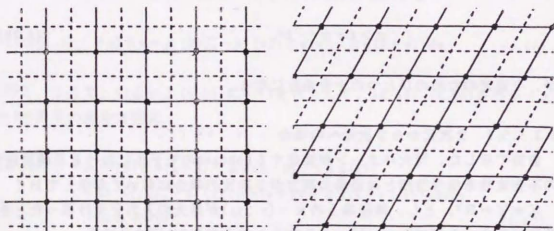


図4.15 受容野の変化の表しかた

変化した受容野の形を、 $x$ から見た $\theta$ 方向の境界までの距離 $r$ で表すことにする。



a)

b)

図4.16 二つの固有関数の重ね合わせによって得られる変形

a) 互いに直交する二つの $\omega$ について(4.73)を重ね合わせた場合、  
b) は、互いに $60^\circ$ の角度ををなす二つの $\omega$ に対して重ね合わせた場合、  
図中 $w$ は破線から離れ、実線に向かって集まるので、その結果黒丸に最も $w$ が集中する。

とにする(図4.15)。すなわち $\theta$ 方向の単位ベクトルを $e(\theta) = (\cos \theta, \sin \theta)^T$ と書くことにすれば、

$$r(\theta, x, w) = \max\{t > 0 \mid x + te(\theta) \in R(x, w)\},$$

となる。特に、 $r(\theta, x, 1)$ は、 $\theta, x$ に関わらず $D$ である。また、 $w$ が、1から僅かに変化して、 $w = 1 + \delta w$ となったときの $r$ の変化 $\delta r$ は、 $x + De(\theta)$ における $w$ の変化の $e(\theta)$ 方向の成分であるから、

$$\delta r(\theta, x) = (\delta w(x + De(\theta)), e(\theta)), \quad (4.72)$$

と与えられる。 $\delta w$ に関する変分方程式は、1次元の場合と同様に考えると、この $\delta r$ を使って、

$$\begin{aligned} \Delta \delta w(x, t) &= D^2 \alpha(D) \int_{\theta} \delta r(\theta, x) e(\theta) d\theta \\ &\quad - \pi D^2 \alpha(D) \delta w(x, t) \\ &\quad + \beta(D) \int_{B(D, x)} \int_{B(D, x)} -h'(\|v\|) \cdot \\ &\quad \quad (v / \|v\|, \delta w(x+y) - \delta w(x+y+v)) y dv dy \\ &\quad - \sigma \Delta \delta w(x, t), \end{aligned}$$

となる。ここに、

$$\alpha(D) \triangleq F(\int_{B(D, 0)} h(\|v\|) dv), \quad \beta(D) \triangleq F^2(\int_{B(D, 0)} h(\|v\|) dv),$$

である。

つぎに、この式に

$$\delta w(x, t) = w \exp\{\lambda t + i(\omega, x)\}, \quad (4.73)$$

を代入するのであるが、この系の等方性より、 $\omega$ が $e(0)$ 方向を向いていると仮定してもやはり一般性を失わない。そこで、 $\omega = (\omega_1, 0)$ と書くことにしよう。すると、固有値問題

$$\lambda w_a = (L - u u_a^T - \sigma \omega_1^2) w_a, \quad (4.74)$$

を得る。ここに、

$$L \triangleq D^2 \alpha (D) \left\{ 2 \int_0^{\pi} \cos(D\omega_1 \cos \theta) \left( \begin{array}{c} \cos^2 \theta \\ 0 \\ 0 \end{array} \right) d\theta - \pi I \right\},$$

$$u \triangleq \int_0^{\pi} dr \int_0^{\pi} d\theta r \sin(\omega_1 r \cos \theta) \left( \begin{array}{c} \cos \theta \\ 0 \\ 0 \end{array} \right), \quad (4.75)$$

$$u_h \triangleq \int_0^{\pi} dr \int_0^{\pi} d\theta -h'(r) \sin(\omega_1 r \cos \theta) \left( \begin{array}{c} \cos \theta \\ 0 \\ 0 \end{array} \right),$$

である。

(4.74)は固有ベクトル  $e(0) = (1, 0)^T$  と  $e(\pi) = (0, 1)^T$  を持つが、このうち、 $e(\pi)$ の固有値が必ず負になることは容易に分かる。 $e(0)$ の固有値に関してはこのままでは、よく分からないので、1次元のときとおなじく例4.3について考えてみることにしよう。この場合は、 $h'(r) = -\gamma r$ なので、

$$u_h = \gamma u,$$

となり、固有値問題(4.74)は、

$$\lambda w_a = (L - \gamma u u^T - \sigma \omega_1^2 I) w_a, \quad (4.76)$$

となる。この式より、 $\gamma$ が大きいとき、 $e(0)$ の固有値は $\gamma$ に近いことがわかる。従ってその実部は正である。いまは(4.73)において $\omega$ が $e(0)$ の方向を向いていると仮定しているが、一般の $\omega$ に対しては、 $\gamma$ が大きいとき、 $\omega$ 方向の固有ベクトルが正の実部を持つことが分かる。

$\omega$ と固有値 $w_a$ が等しいとき、(4.73)は、 $\omega$ に直交する等間隔の平行線群に向かって $w$ が集中するような変形を意味している。このような変形を、互いに直交する二つの $\omega$ について考え、それらを重ね合わせると、正方格子上の各点に向かって $w$ が集中するような変形が得られる。また、互いに $60^\circ$ の角度をなす二つの $\omega$ に対する変形を重ね合わせると、六方格子上の各点に向かって $w$ が

集中するような変形が得られる(図4.16)。実際に生じる不安定はこのようなものだと思われる。

#### 4.2.4 Amariのモデルとの比較

4.2.3.5節で、Kohonenのモデルを改良すればコラム状の微細構造が生ずるようになることを示した。ここで例4.3に関する理論的な考察から分かったことは、コラム状の微細構造が生ずるような不安定の鍵を握っているのは、パラメタ $\gamma$ であるということである。そして $\gamma$ は、信号空間の信号間の相関に相当する関数 $h$ が、信号空間にとられた座標の変化によってどのくらい速く減衰するかを表すパラメタであった。得られた結果を定性的に述べれば、 $h$ の広がり狭いときに単純連続解は不安定になり、コラム状の微細構造が形成されるということになる。

ここで、コラム状の微細構造の形成が理論と数値実験で示されているAmariのモデルと本論文で得られた結果を比較しておこう。Amariのモデル[5,9]は、コラム状の微細構造が自己組織化によって形成されること示した最初の、また本論文以前の唯一のモデルである。

##### 4.2.4.1 Amariのモデル

Amariのモデルは、入力神経場 $Y$ 、出力神経場 $X$ 、抑制型の神経細胞群 $I$ の三つから成っている。文献[5,9]では、 $X$ 、 $Y$ が一次元の閉区間である場合が取り扱われているが、ここでは $X=Y=R$ の場合を考えよう。これは、一つは簡単のため、もう一つは単純連続解(このモデルにおける単純連続解の正確な定義はあとで述べる)における $X$ と $Y$ の対応の比率を変化させたいからである。 $Y$ に与えられる興奮パターンは、定まった興奮パターン $a$ を様々な場所に平行移動させたものに限られる。例えば、 $y'$ の周りの興奮ならば $a(y-y')$ である。これは一山の正の偶関数である。 $Y$ 上の興奮は $Y$ から $X$ への結合 $w(x, y)$ によって、 $X$ に興奮を引き起こす。 $X$ 上の細胞は、 $Y$ からの入力のほかに $I$ からの抑制性的の入力 $w_0$ と、 $X$ 内部の相互結合 $w_X(x, x') = w_X(x-x')$ による入力を受けている。 $w_X(x)$ は偶関数で、 $|x|$ が小さいとき正、やや大きくなると負に転じる。 $X$ 上の細胞は、第3章のモデルのように、一個の実数 $u(x, t)$ で表される内部状態を持つ。 $u(x, t)$ のダイナミクスは次の式で与えられる。

$$\tau^{-1} u(x, t) = -u(x, t) + \int w_X(x-x') I[u(x', t)] dx' + \int w(x, y) a(y-y') dy - w_0 - h. \quad (4.77)$$

$y'$  を固定してしばらくすると、 $u(x, t)$  は、 $y'$ 、 $w$ 、 $w_0$  によって決まる平衡状態に収束する。式 (4.77) は一般には多安定であるが、通常  $\int w(x, y) a(y - y') dy - w_0$  を最大にする  $x$  の付近に、 $u > 0$  となる興奮領域を生ずる平衡状態がある。これを  $u'(x, y', w, w_0)$  と書くことにする。 $w$  が単純連続解に近い場合はこれ以外には平衡状態はない。Kohonen のモデルにおける  $x$  の近傍  $N_0(x)$  に対応するのは、 $u > 0$  となる興奮領域である。この定義から明らかなように、この興奮領域は Kohonen のモデルの  $N_0(x)$  と異なり、 $y'$ 、 $w$ 、 $w_0$  によってその広さが変化する。これがコラム状の微細構造を形成する原因だと考えられる。

$w$  と  $w_0$  は、Hebb 学習によって変化するが、それには  $u'(x, y', w, w_0)$  が用いられる。すなわち、 $w$  と  $w_0$  の変化は  $u(x, t)$  の変化に較べて遅かにゆっくりしており、 $Y$  に与えられる入力位置  $y'(t)$  も、 $u(x, t)$  が  $u'(x, y', w, w_0)$  のそばに十分長く留まっていられるようしばらくは一定の値に留まったのち他の値に変わる場合を考え、いわゆる断熱近似を用いるのである。

$$\begin{aligned} \tau \frac{d}{dt} w(x, y, t) &= -w(x, y, t) + ba(y - y'(t)) I[u'(x, y'(t), w, w_0)], \\ \tau \frac{d}{dt} w_0(x, t) &= -w_0(x, t) + b' I[u'(x, y'(t), w, w_0)], \end{aligned} \quad (4.78)$$

$w_x$  は変化しない。

$y'(t)$  が、 $Y$  上で一様かつランダムに選ばれるとして (4.78) から平均学習方程式を作ると、次のような 1-パラメタの単純連続解の族があることが容易に確かめられる。すなわち、 $Y$  に対する入力  $a(y - y')$  に対して上の興奮領域が  $[y'/p - D(p), y'/p + D(p)]$  となるような解である。 $u'(x, y', w, w_0)$ 、 $w(x, y)$ 、 $w_0(x)$  の具体的な形はこの条件から一意に定まる。 $p$  と  $D(p)$  の関係は次の式により定まる。

$$H(D, p) \triangleq K(2Dp) - W(2D) - h = 0 \quad (4.79)$$

ここに、

$$\begin{aligned} W(y) &\triangleq -\int_0^1 w_x(x) dx, & K(u) &\triangleq \int_0^1 k(y) dy, \\ k(y) &\triangleq b \int a(y') a(y' + y) dy' - b' \end{aligned}$$

である。

一つの  $p$  に対し (4.79) は、2 個の  $D$  を解としてあたえるが、小さい方の解は不安定な興奮を表し、大きい方の解が安定な興奮を表している。よって  $D(p)$  としては大きい方をとらなければならない。この  $D(p)$  に関しては (4.79) に加えて、

$$\partial H(D, p) / \partial D < 0, \quad (4.80)$$

が成立する [5]。

これらの単純連続解は、 $k(2Dp)$  が負のとき安定、正のとき不安定であることが知られている [9]。これは、 $k$  の広がり大きいほど不安定になり易いことを意味している。これを前節で取り扱った、改良された Kohonen のモデルに対応させて考えると、Amari のモデルの  $k$  が、改良された Kohonen のモデルの  $h$  に、ほぼ対応している。ところが、改良された Kohonen のモデルの場合は、 $h$  の広がり小さいほど不安定になり易い。

この定性的な食い違いはどこから生ずるのであろうか。4.2.3.5 節の最初で、荷重ベクトル  $w(x)$  が信号空間のある点  $y$  の周りに集中し始めたとき、信号  $y$  が  $X$  に引き起こす興奮パターンが強くなればコラム構造ができるはずであると述べた。Amari のモデルの場合、 $X$  における活動度を 1 からの入力が制御しており、この強さが学習によって変化する。このため、 $X$  における興奮は  $a$  の広がり大きいときでも、いわば、ぎりぎりのところで保持されている。このため、荷重ベクトルが信号空間のある点の周りに集中し始めたときの興奮の変化は、改良された Kohonen のモデルほど明らかではない。そこで (4.79) を使って、これを計算してみよう。 $p$  は、 $X$  の細胞の荷重ベクトルの  $Y$  への集中度の逆数と考えられる。そこで  $dD(p)/dp$  を求めてみると、

$$\begin{aligned} dD(p)/dp &= -(\partial H / \partial p) / (\partial H / \partial D) \\ &= -2DK'(2Dp) / (\partial H / \partial D) \\ &= -2Dk(2Dp) / (\partial H / \partial D), \end{aligned}$$

となる。よって、 $dD(p)/dp$  が正のときに (4.80) より  $k(2Dp)$  も正となり、単純連続解は不安定となる。つまり、Amari のモデルの場合は、本論文で扱ったモデルと異なり、 $X$  の細胞の荷重の集中が興奮領域の幅を狭める効果が働くときにコラム構造が形成されるということが分かった。

これは直観的には理解しにくいことである。おそらく Amari のモデルでは、コラム構造の形成に関して、抑制性のシナプスが非常に重要な役割を果たしてい

ることがこの違いをうむと思われる。

#### 4.2.5 Malsburgのモデルとの比較

本節では発火領域の幅が変化しうるモデルのも一つの例であるMalsburgのモデルに関する理論的な解析をおこなう。KohonenのモデルとMalsburgのモデルの違いは、発火領域を $N_0(x^*)$ のような予め決まったパターンで与えるか、神経場のダイナミクス、

$$\tau' \frac{\partial}{\partial t} u(x, t) = -u(x, t) + \int w(x-x') I[u(x', t)] dx' + \int h(w(x)-y) dy, \quad (4.81)$$

によって決定するかという一点である。このダイナミクスは、Amariのモデルの(4.77)と同じである。入力信号 $y$ を固定したときの平衡解 $u'(x, y, w)$ が正の部分だけで学習がおきる。 $X=Y=R$ ,  $w: X \rightarrow Y$ とする。

$$w(x) := w(x) + \epsilon I[u'(x, y, w)](y - w(x)), \quad (4.82)$$

場所 $x$ の細胞の受容野 $R(x, w) \triangleq \{y \in Y \mid u'(x, y, w) \geq 0\}$ を $[r_1(x, w), r_2(x, w)]$ ,  $r_1(x, w) < r_2(x, w)$ と書くことにすると、(4.82)の平均学習方程式は、

$$\frac{\partial}{\partial t} w(x, t) = (r_2 - r_1) \{ (r_2 + r_1) / 2 - w(x, t) \} + \epsilon \Delta w(x, t), \quad (4.83)$$

となる。今までどおり拡散項を付け加えておく。これから単純連続解、

$$w(x) = x, \quad r_1 = x - D, \quad r_2 = x + D,$$

の周りの変分方程式を求めると、

$$w(x) = x + \epsilon \delta w(x, t), \quad r_1 = x - D + \epsilon \delta r_1(x, w), \quad r_2 = x + D + \epsilon \delta r_2(x, w), \quad (4.84)$$

$$\frac{\partial}{\partial t} \delta w(x, t) = 2D \{ (\delta r_2 + \delta r_1) / 2 - \delta w(x, t) \}, \quad (4.85)$$

を得る。

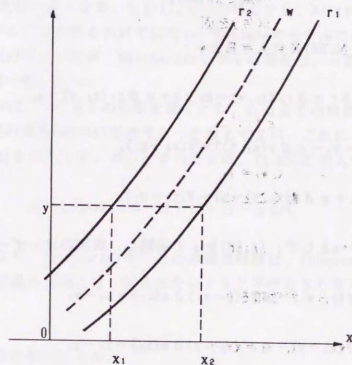


図4.17 荷重ベクトル $w$ と $r_1, r_2$



いま、信号 $y$ に対して神経場の領域 $[x_1, x_2]$ 、 $x_1 < x_2$ 、が発火したとする。つまり、

$$r_2(x_1) = r_1(x_2) = y, \quad (4.86)$$

とする(図4.17)。このとき、

$$h(y - w(x_1)) = h(y - w(x_2)) = W(x_1 - x_2), \quad (4.85)$$

が成立する[1]。これに単純連続解を代入すると、

$$h(D) = W(2D), \quad (4.86)$$

を得る。また、(4.84)を代入すると、

$$x_1 + D + \epsilon \delta r_2(x_1, w) = x_2 - D + \epsilon \delta r_1(x_2, w) = y, \quad (4.87)$$

$$h(y - x_1 - \epsilon \delta w(x_1, t)) = W(x_2 - x_1), \quad (4.88)$$

$$h(x_2 + \epsilon \delta w(x_2, t) - y) = W(x_2 - x_1),$$

を得る。 $J \triangleq W^{-1} \cdot h$ として、(4.88)を $\epsilon$ で展開し、高次のオーダーを無視すると、

$$J(y - x_1) - \epsilon J'(y - x_1) \delta w(x_1) = x_2 - x_1$$

$$J(x_2 - y) - \epsilon J'(x_2 - y) \delta w(x_2) = x_2 - x_1$$

これに、(4.87)から得られる、

$$x_2 - x_1 = 2D + \epsilon (\delta r_2(x_1) - \delta r_1(x_2)),$$

$$y - x_1 = D + \epsilon \delta r_2(x_1), \quad x_2 - y = D - \epsilon \delta r_1(x_2)$$

を代入して展開し、(4.86)をつかおうと、

$$\begin{pmatrix} 1 & -1+c \\ -1+c & 1 \end{pmatrix} \begin{pmatrix} \delta r_1(x_2) \\ \delta r_2(x_1) \end{pmatrix} = c \begin{pmatrix} \delta w(x_1) \\ \delta w(x_2) \end{pmatrix}, \quad (4.89)$$

最後に、(4.87)より、 $x_2 = x_1 + 2D + 0(\epsilon)$ なので、(4.89)において、 $0(\epsilon)$ を無視して、 $x_1$ を $x$ と書き直せば、

$$\begin{pmatrix} -1 & 1+c \\ 1+c & -1 \end{pmatrix} \begin{pmatrix} \delta r_1(x+2D) \\ \delta r_2(x) \end{pmatrix} = c \begin{pmatrix} \delta w(x) \\ \delta w(x+2D) \end{pmatrix}, \quad (4.89)$$

を得る。ここに、 $c \triangleq -J'(D) = h'(D) / w_x(2D) > 0$ であり、 $c$ が小さいほど、発火領域の幅の変化が大きくなる。しかし、 $c=0$ の場合でも、Malsburgのモデルは、Kohonenのモデルに一致するわけではない。Kohonenのモデルでは、発火領域の中心は常に $x^*(y)$ であったが、Malsburgのモデルの場合は、一般には発火領域の中心と $x^*(y)$ は一致しない。

(4.87)において、 $y$ が全ての実数を動くとき、 $x_1$ は全ての実数を動くので、(4.89)は、任意の実数 $x$ について成り立つと考えてよい。これに $\delta w(x, t) = \exp(i\omega x - \lambda t)$ を代入して、 $\delta r_1$ と $\delta r_2$ を求め、(4.85)に代入すると、

$$\lambda = 2D(\cos 2\omega D - 1) / (2+c) - \sigma \omega^2, \quad (4.90)$$

を得る。これより、Malsburgのモデルの単純連続解は、Kohonenのモデルと同様、コラム構造の形成の方向には、拡散のないときに不安定ぎりぎりであることが分かる。

#### 4.2.6 相互結合のないモデル。

発火強度が変化するように改良されたKohonenのモデルでは、その変化率がいかに小さくても、単純連続解を不安定にするためには十分であった。しかしMalsburgのモデルは、発火領域の幅が変化する性質を持っているにもかかわらず、そして、確かにその傾向が強まれば強まるほど固有値 $\lambda$ は大きくなるのだが、不安定性をおこせない。その理由を理解するために、つぎのようなモデルを考えてみよう。

$$w(x) := w(x) + \epsilon [u(x, y, w)](y - w(x)),$$

$$u(x, y, w) = h(w(x) - y) - \theta.$$

このモデルは、神経場のなかに相互結合を持たないでトポグラフィック・マッピングのモデルとしては trivial であり、単純連続解の安定性解析をおこなえば、拡散を付け加えなければ、全ての  $\omega$  に対して  $\lambda = 0$  という結果を得る。いふならば、これこそが本当に「不安定ぎりぎり」のモデルなのである。しかし、このモデルは、 $w(x)$  の集中にともなって発火領域が広くなるという性質を持っている。従って、このモデルより発火領域の広がり方が弱いモデルでは、コラム形成がおきないと考えられる。

このモデルでは受容野の広さが一定に保たれていることから考えると、もしも、あるモデルにおいて、 $w(x)$  の集中にともなって発火領域が広くなり、その結果受容野が広くなれば、コラム構造ができるのではないだろうか。すなわち、単純連続解  $w(x) = px$  における発火領域の幅を  $2D(p)$  とするとき、

$$d\{pD(p)\} / dp < 0$$

が成立すれば、コラム構造が形成されると思われる。

これを一般的な枠組みで証明することはできないが、Kohonen のモデルで確かめてみよう。(4.85)より、 $J(pD) = 2D$  を得る。 $H(p, D) = J(pD) - 2D = 0$  において、まず  $dD/dp$  を求めると、

$$\begin{aligned} dD/dp &= -(\partial H / \partial p) / (\partial H / \partial D), \\ &= -DJ'(pD) / (pJ'(pD) - 2). \end{aligned}$$

これより、

$$d\{pD(p)\} / dp = 2D / (2 - pJ'(pD)).$$

$J'(pD)$  は負なので、 $d\{pD(p)\} / dp$  は正である。

本章では、トポグラフィック・マッピングの形成モデルにおけるコラム構造とハイパーコラム構造の形成に焦点をあて、まずボルツマン・マシンの理論を応用したトポグラフィック・マッピングの形成モデルを構成し、ハイパーコラム構造を表す解が安定であることを示した。つぎにKohonenのモデルの単純連続解の不安定性としてこれらの構造の形成される条件をもとめた。その結果ハイ

パー・コラム構造はどのモデルでも神経場の細胞が十分多ければ必ずおこること。コラム構造の形成は通常のKohonenのモデルではおきないがモデルを改良すればおきるようにできることがわかった。またMalsburgのモデルでは、コラム構造は形成されないことが分かった。

コラム構造ができるかどうかという重要な点でモデル間に定性的な不一致があるということは看過できない事実である。脳内のコラム構造が、Hebb型の自己組織化で形成されるのかどうかを考えるにあたって、この不一致の原因を明らかにしておくことは、何にもまして重要である。また応用の観点からこの問題を眺めると、神経場をシミュレートするような機構で信号空間のトポロジを抽出しようとする場合、コラム構造の形成は避けたい場合が多いと思われる。よってコラム構造の形成条件は応用上も重要である。

本章で得られた結果、コラム構造の形成を促す要因として、

- 1) 信号空間における荷重ベクトルの集中が神経場の発火パターンを増強すること。ただし、興奮領域の拡大よりは興奮度の上昇のほうが有効らしい。
- 2) 神経場に生ずる興奮パターンの周囲に負の抑制領域が存在すること。
- 3) 神経場全体が抑制入力を受けており、そのシナプスの学習によって神経場の興奮度が制御されていること。

の三つが考えられることが分かった。3)の仮定は神経場のしきい値が変化するという仮定と同値である。これらの性質は、現実の神経場のモデルとして不自然なものではなく、特に1)と2)は極めて自然で、トポグラフィック・マッピング形成モデルに共通な内部結合を持った神経場がもし現実であれば、程度差こそあれ、殆ど避けがたく生ずる性質であるといえる。

#### 文献

1. 甘利俊一, 神経回路網の数理, 産業図書, 1978.
2. 倉田耕治, 甘利俊一, 確率的に動作する自己組織神経回路網について, 信学技報, MBE85-104, 1985.
3. 倉田耕治, 神経回路モデルとしてのボルツマン・マシン, 数理科学, 289, pp. 23-28, 1987.
4. Ackley, D. H., Hinton, G. E., Sejnowski, T. J., A learning algorithm for Boltzmann machine, Cognitive Science, 9, pp. 147-169, 1985.
5. Amari, S., Topographic organization of nerve fields, Bull. Math. Biol., 42, pp. 339-364, 1980.

6. Hubel, D. H., Wiesel, T. N., Functional architecture of macaque monkey visual cortex. Proc. R. Soc. B198, pp.1-59, 1977.
7. Kohonen, T. Self-organized formation of topologically correct feature maps. Biol. Cybern., 43, pp.59-69, 1982.
8. Malsburg, C. von der., Self-organization of orientation sensitive cells in the striate cortex. Kybernetik, 14, pp. 85-100, 1973.
9. Takeuchi, A., Amari, S., Formation of topographic maps and columnar micro structure. Biol. Cybern., 35, pp.63-72, 1981.
10. Willshaw, D. J., Malsburg, C. von der., How patterned neural connections can be set up by self-organization. Proc. R.Soc. B194 pp. 431-445, 1976.

あとがき

本論文では、情報の表現形式という側面から脳をながめ、それに関する一連のモデルにたいして数理工学的な解析をおこなった。各章で明らかにされたことがらは、まえがきでまとめておいたので、それを改めてくり返すことはしない。

現在、神経回路網の分野では現実問題への応用が盛んなため、扱われるモデルと脳との関連が次第に薄れていく傾向がある。バック・プロパゲーションも脳の学習モデルとしては、かなり不自然であるし、我々が巡回セールスマン問題を解くとき、Hopfieldの回路におけるようなプロセスが脳の中で起こっているとはとうてい考えられない。

応用の立場に立てば、この傾向を一概に悪いものと決めつけることはできないが、本論文では生物学的常識に反するモデルは取り扱わなかった。よって、ここに登場するモデルは、脳のモデルとしての権利を放棄していないものばかりである。

しかし、かといってここで述べられたモデルが、脳のどの部分にあるのかと問われれば、いまのところ確証をもって答えられるものはない。第1章のランダム対称結合が海馬とその周辺にあるとするのが森田のモデルであるが、これは今のところ、興味深い作業仮説であるに留まっているし、第3章のHebb学習をおこなう競合的な細胞群も今のところ見つからない。第4章のトポグラフィック・マッピングの形成モデルは感覚系のモデルであるが、未だに実験的には確認されていない。

しかし、まえがきでも述べたように、数理工学的な脳の研究の目的は生理学者に仮説を提供することなのである。いつの日にか、ここで述べられたモデルのうちのいずれか一つでも、パーセプトロンの行った道を辿り、ここでおこなわれた解析が、脳の情報処理の本質を幾分かでも捕らえていることが明らかになれば、著者にとってこれに過ぎる幸せはない。

謝辞

本論文をまとめるにあたり、院生、助手時代を通じて9年間の長きわたり、最後まで私を見捨てることなく御指導下さった甘利俊一教授に深く感謝します。赤穂昭太郎君には第1章の数値計算を手伝っていただきました。ここに感謝します。

付録1

帯状の信号空間における受容野  $R(x, w)$  の面積  $|R(x, w)|$  と重心  $g(x, w)$  を求める。  $R(x, w)$  は、  $y_1 y_2$ -平面において、  $(p, 0)$  を通り傾き  $1/m$  の直線、  $(q, 0)$  を通り傾き  $1/n$  の直線、および2直線  $y_2 = \pm a$  の計4本の直線に囲まれた台形の領域である(図A.1)。まず面積は、台形  $GHJI$  = 長方形  $FBC E$  - 三角形  $AFG$  + 三角形  $ABH$  + 三角形  $DEI$  - 三角形  $DCJ$  = 長方形  $FBC E$  より、簡単に

$$|R(x, w)| = 2a(q-p)$$

と求められる。

重心も同様に長方形  $FBC E$  と四つの三角形について計算したものを足し引きして求める。とりあえず  $m > 0$ 、 $n > 0$  の場合を考えると、それぞれの領域の重心と面積は表A.1の様に求められる。この表では、引かなければならない部分の面積は負にしてある。これらの重心をそれぞれの面積の重みを付けて平均すると、台形  $GHJI$  の重心

$$g(x, w) = \left( \begin{array}{c} ((q+p)/2) + \{a^2(n^2 - m^2)/12(q-p)\} \\ a^2(n-m)/3(q-p) \end{array} \right),$$

を得る。ここでは  $m$  と  $n$  が両方とも正であるとして計算したが、この結果は  $m$  と  $n$  の符号がどの様な組合せの場合でも成り立つ。

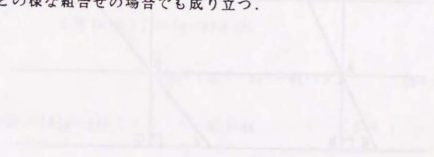


表 A. 1

領域	重心	面積
長方形 FBCE	$((p+q)/2, 0)^T$	$2a(q-p)$
三角形 AFG	$(p+am/3, 2a/3)^T$	$-a^2m/2$
三角形 ABH	$(p-am/3, -2a/3)^T$	$a^2m/2$
三角形 DEI	$(q+an/3, 2a/3)^T$	$a^2n/2$
三角形 DCJ	$(q-an/3, -2a/3)^T$	$-a^2n/2$

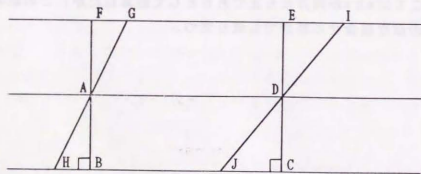


図 A. 1  $R(x, w)$  の面積と重心 (2次元の場合)

付録 2

円柱状の信号空間における受容野  $R(x, w)$  の体積  $|R(x, w)|$  と重心  $g(x, w)$  を求める。基本的な計算法は 2次元の場合と同じである。すなわち  $R(x, w)$  は、 $y_1, y_2, y_3$ -空間において、 $y_1$  軸を中心とした半径  $a$  の無限円柱領域  $C = \{(y_1, y_2, y_3)^T \mid y_2^2 + y_3^2 \leq a^2\}$  から、 $(p, 0, 0)^T$  を通りベクトル  $m = (m_1, m_2, m_3)^T$  に直交する平面と、 $(q, 0, 0)^T$  を通りベクトル  $n = (n_1, n_2, n_3)^T$  に直交する平面によって切り取られる立体である (図 A. 2)。ベクトル  $m$  は、この立体の内部を向いており、ベクトル  $n$  は、外部を向いている。この立体の体積と重心を 2次元の時と同様に 5個の立体の足し引きによって求めるのだが、そのうちの 4個は、円柱の底部から、底面の直径を通る平面で切り取られる形の立体である。この形の立体の体積と重心を求めると、つぎのようになる。すなわち、 $C$  から角度  $\theta$  をなす 2平面  $y_1 = 0, y_1 = y_2 \tan \theta$  によって切り取られる立体 (図 A. 3) の体積と重心は、それぞれ  $(2/3)r^2 \tan \theta, ((3/32)\pi r \tan \theta, (3/16)\pi r, 0)^T$  である。これを用いると 5個の立体の体積と重心は、それぞれ表 A. 2 の様に求められる。表 A. 2 において、

$$t(m) = \sqrt{(m_2^2 + m_3^2)}, \quad t(n) = \sqrt{(n_2^2 + n_3^2)},$$

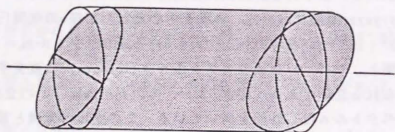
$$g(m) = (3/16)\pi a (t(m)/2, -m_2/t(m), -m_3/t(m))^T,$$

$$g(n) = (3/16)\pi a (t(n)/2, -n_2/t(n), -n_3/t(n))^T,$$

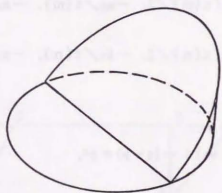
である。これより、

$$|R(x, w)| = (q-p)\pi a^2, \\ - (m_2^2 + m_3^2 - n_2^2 - n_3^2) / 2 \quad (p+q) / 2 \\ g(x, w) = a^2 / \{4(q-p)\} ( \begin{matrix} m_2 - n_2 \\ m_3 - n_3 \end{matrix} ) + ( \begin{matrix} 0 \\ 0 \end{matrix} )$$

を得る。


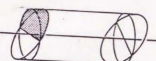
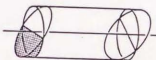
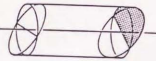
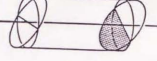


図A.2  $R(x, w)$ の形(3次元の場合)



図A.3 円柱の底部から、底面の直径を通る平面で切り取られる形の立体

表A.2

傾城	重心	体積
	$((p+q)/2, 0, 0)^T$	$\pi a^2 (q-p)$
	$(p, 0, 0)^T + g(m)$	$-(2/3)a^3 t(m)$
	$(p, 0, 0)^T - g(m)$	$(2/3)a^3 t(m)$
	$(q, 0, 0)^T + g(n)$	$(2/3)a^3 t(n)$
	$(q, 0, 0)^T - g(n)$	$-(2/3)a^3 t(n)$

