TR－O－0018　　　　　　　　98

リミットサイクルを埋め込んだ、非対称な結合行列を持つ、神経回路網の記憶想起特性

森 裕平、 Peter Davis、 奈良 重俊

1990年1月19日

Abstract

記憶させたいパターンをグループ分けし、各グループ内で隣り合うパターンの直積を順番にとり、それをサイクル状に足し合わせたシナプス結合行列を持つような神経回路モデルの記憶想起特性を調べた。その結果、このモデルは、良く知られた自己相関でシナプス結合を作るモデルよりも、ノイズ除去能力が良いことがわかった。この理由は、擬似アトラクターへ陥ることが避けられるためである。

目次

1 はじめに

複数のパターンを神経回路に記憶させておき、ノイズの混ざった入力パターンをもとに、記憶させたパターンのひとつを思い出させる問題は、古い歴史をもつ問題である [1, 2, 3, 4, 5]. とりわけ、記憶させたいパターンの自己相関を対称行列に記憶させた神経回路の、記憶容量や、統計的な記憶想起特性が詳しく調べられた [6][and references therein]。Hopfield は [5]、自己相関を記憶させた対称行列で表される神経回路について、記憶想起過程を記述するために、エネルギー的な Lyapunov 関数と統計物理学的な方法を導入した。こういった対称行列型の神経回路では、想起する過程で擬似アトラクター (local minimum, spurious minimum, spurious attracter) に陥るため、思いだしの特性があまり良くならない。したがって、擬似アトラクターと、記憶させたパターンに対応するグローバルミニマムとをどうやって区別するかが問題である。擬似アトラクターができる原因とその性格について、これまで、多くの研究が行なわれてきた [6]。たとえば擬似アトラクターの量と温度との関係について議論された [7, 8, 9]。温度の上昇に伴って、若干のノイズを系に加えることが、擬似アトラクターから抜け出すことに役立つことがわかり、アニールして序序に温度を下げていく方法が、より深いミニマムへ行く方法として提案された [10]。

他方、非対称な行列をシナプス結合として持つ神経回路では、一般には Lyapunov 関数が定義できず、特定の一点のアトラクターへ収束するという保証はない。従ってもっと複雑な動力学的振るまいが起こる可能性がある。自己相関対称型の神経回路モデルに、少しだけランダムに非対称性を加えた時の効果は、ノイズを系に加えたときと同様であることが調べられた [11]。非対称神経回路に対する他の研究は、異なる見方からも行なわれている。Shinomoto[12] は、個々のシナプスが興奮型か抑制型のどちらか一方であるといった仮定から、非対称なシナプス結合行列をつくりモデル計算したところ、ほとんどの擬似アトラクターが除去でき、系全体は、記憶させたパターンのどれかひとつに収束するか、あるいは模様が一様な "わからない" を意味する状態に陥るかのどちらかであった。Parisi[13] は、一時的に行列の非対称性を大きくすることが記憶想起仮定に効果的であると主張した。対称な部分と非対称な部分が組み合わされた行列では、対称な部分はある特定の定常状態への収束をつかさどり、非対称な部分はパターン間の遷移を引き起こすといったモデルが提案され、一連のパターンがつぎつぎに移り変わる状態を作り出すものとして使われた。非対称なシナプス結合を一時的に利用するようなモデルは多い。たとえば、シナプス結合の動力学を扱ったもの [14]、時間をずらせてパターン遷移させるもの [15, 16, 17]、あるいは遷移を起こさせるために温度によるノイズを利用するもの [18] など [19, 20, 21] がある。これらのモデルは、パターン間の遷移やその遷移がサイクル状に回ることなど、動力学的に興味深い事柄にふれている。

本研究では、パターンを行列に記憶させる際に、自己相関ではなく、隣りのパターンとの相関を行列に足し込む方法でつくる非対称行列型神経回路モデルについて、記憶想起の特性が良くなるかどうかを調べた。

2　リミットサイクル型モデル

ここでは、発火パターンが次のようにくりかえし同期的に変化していくモデルを扱う。

$$S_i(t+1) = sign(\sum_{j=1}^{N} J_{ij} S_j(t)) \tag{1}$$

パターンの各要素は、

$$S_i = (1 \ or \ -1), \qquad i = 1 \sim N \tag{2}$$

N は神経細胞の数である。神経回路のなかに埋め込むべきそれぞれのパターンベクトルを $\xi^{\mu,\nu}$ とすると、シナプス結合を示す行列 J のなかには、隣り合うパターンベクトル $\xi^{\mu,\nu+1}$ と $\xi^{\mu,\nu}$ の直積を足し込んでいく。

$$J = \sum_{\mu=1}^{K} \sum_{\nu=1}^{L} \xi^{\mu,\nu+1} \otimes \xi^{\mu,\nu} \tag{3}$$

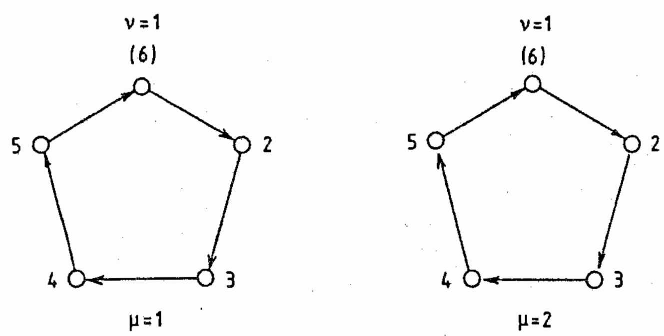

図1. J 行列のなかに埋め込むサイクルの例。μ はサイクル番号であり、ν はパターン番号である。直積をとるペアーを矢印で示す。

図1でそれぞれの丸は各パターンであり、隣との直積を矢印のようにとり、サイクル状に加える。式3で、μ はサイクルのラベルであり、ν はパターンの番号であり、また L は1サイクルあたりのパターンの数、K はサイクルの数である。全パターン数は $M = KL$ である。式1の行列 J のうち、対角要素は $J_{ii} = 0$ とおいた。ここで、$L = 1$ の場合は、自己相関型神経回路 (Hopfield 型) が同期的に変化していく場合であり、行列 J は対称である。この場合、神経回路の系は、ある一点に収束するか、あるいはまれに周期的な2点に収束する [22]。$L > 1$ のときは、行列 J は非対称であり、複雑な軌道を描く。たとえば、サイクルに収束したり、カオスになるなど。

このモデルにおいて、記憶させたパターンのどれかひとつにノイズを加えたものを初期パターンとして、式1に従って繰り返し、目的のパターンに収束させる。うまくいけば、記憶したパターンをサイクル状に渡り歩くものに収束する。ここで、もしノイズがなくなれば記憶想起は成功である。しかし、もし擬似アトラクターに収束したり、記

憶したパターンをサイクル状に渡り歩くものには収束できても位相がずれて収束したり、めざしたサイクル以外のサイクルに収束した場合は、記憶想起は失敗である。計算結果で、ノイズを含む擬似アトラクターを観察したところ、ほとんどが周期 L になり、まれに L の整数倍になり、また、さらにまれに非常に長い周期になっていた。本報告ではこの擬似アトラクターの性格には詳しくふれず、記憶想起過程を統計的に見ることにする。

3 モデルの記憶想起特性

記憶想起過程におけるノイズ除去の様子を、本モデルのリミットサイクルの場合と、よく知られた自己相関型モデル (Hopfield 型) とで比較する。メモリーパターンをランダムに選ぶと、ノイズ除去能力は、$\alpha = M/N$ に依存するものとして知られている [23]。自己相関型モデルの場合、α が一定値 $\alpha = 3/40$ を越えると、擬似アトラクターが発生し、記憶想起が成功する割合が極端に減る [23]。我々はこの α の値のときに、本モデルであるリミットサイクル型と、自己相関型モデルとでパーフォーマンスを比較した。ここでは、$N = 400$ ニューロンのなかに $M = 30$ パターンを記憶した。M 個のパターンは乱数で選んだ。従って、それらの間の重なり（内積）は、統計的にみると、0 のまわりに Gaussian 型の分布をしている。

計算は次のように行なった。記憶させたパターンのひとつ（$\xi^{R,r}$, つまり、パターン番号 $\nu = r$, サイクル番号 $\mu = R$）をターゲットパターンに取り、ターゲットパターンと初期パターンの重なりが $q(0)$ になるようにノイズを加えて、初期パターン $\mathrm{S}(0)$ を選ぶ。初期重なりは、$q(0) = N^{-1}\mathrm{S}(0)\cdot\xi^{R,r} = q_0$ と表せる。初期パターンと任意の他のパターン $(\mu, \nu) \neq (R, r)$ との重なりは、$O(N^{-1/2})$ である。我々の計算では、初期状態から出発し、式 1 に従ってパターンを順次変化させ、重なりの値 $q(t) = N^{-1}\mathrm{S}(t)\cdot\xi^{R,r+t}$ をその都度観察し、パターンがサイクルに収束するか、あるいは時間ステップが $t = 300$ を越えるかするまで追ってみて、このことを繰り返した。周期 L のサイクルに収束した場合、L 種類の重なりの平均値を、改めてその収束の重なりと呼ぶことにした。実際の計算結果では、収束が完全に成功する場合はステップ数は少なく、$t = 3 \sim 9$ ステップである。また、この α 値では周期が L を越えるものに収束する場合は極めて少ない。たとえば、初期重なり 0.2 で出発した場合、記憶したパターンのサイクルに収束する場合が 76%, 周期 L の擬似アトラクターが 21%, より長いサイクルは 3% である。サイクル内のパターンの数を変化させ、それぞれについて 5000 回ずつ走らせた場合の、収束重なり値の平均値を、図 2 に示す。この図から、メモリーサイクルの長さ L が長いほど、記憶想起パーフォーマンスが良いことがわかる。次に、記憶想起パフォーマンスが良くなる理由を詳しくみるため、初期重なり 0.2 から出発したときの収束重なり値をヒストグラムにプロットしたものを、図 3 に示す。自己相関型（すなわち $L = 1$）の場合、84% が目的のパターンとは異なるパターンに収束している。これらはほとんど擬似アトラクターに収束し、収束重なりは 0.4 付近に分布している。

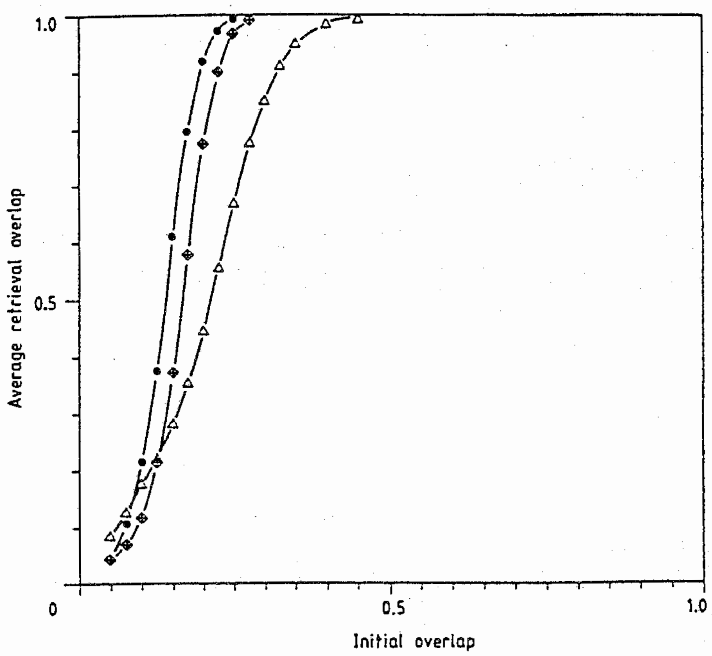

図2．それぞれの初期重なりに対する収束重なり値。収束重なり値は 5000 回走らせて平均してある。△: 自己相関の場合; ⊕: 3 パターン × 10 サイクルの場合; ●: 10 パターン × 3 サイクルの場合。

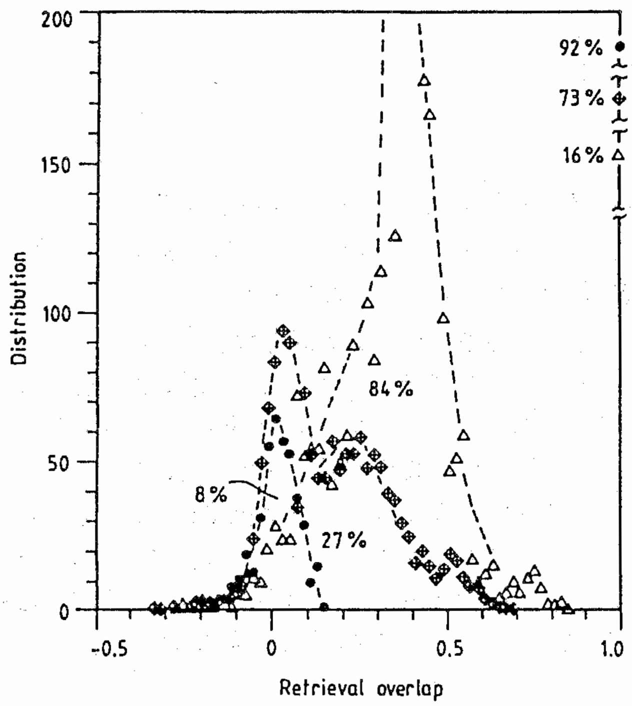

図3．初期重なりが 0.2 のときの収束重なり値のヒストグラム。シンボルの意味は図2と同じ。

$L = 3, L = 10$ のサイクルの場合を比較して見ると、サイクルの長さが長くなるほど収束が成功する割合が増え、擬似アトラクターへ行く割合は減っている。また、サイクル

が長いほど、擬似アトラクターの重なりが小さくなるという特徴があり、たとえば $L=10$ の場合、完璧な記憶想起をしない場合はわずか 8% であるが、収束重なりが 0 のまわりに分布している。

4 自己平均化効果と記憶想起特性

サイクルの長さが長くなるとともに、なぜ記憶想起特性が良くなるのであろうか。この理由として、一種のシャッフリング効果（自己平均化効果）が考えられる。

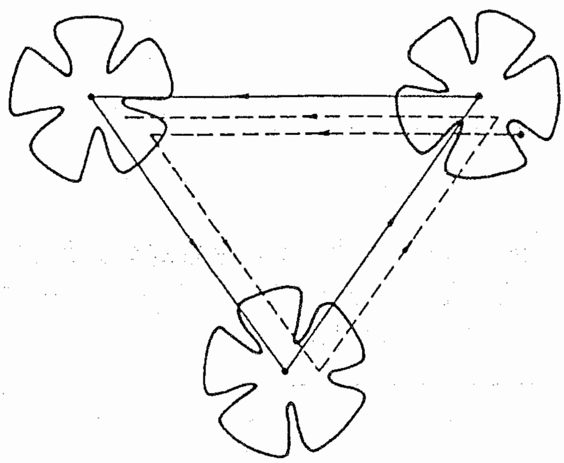

図 4． サイクル型モデルの記憶想起過程におけるパターンシャッフリングの概念図。閉じた領域は自己相関モデルのときの引き込み領域。実線は収束したときの軌道で、破線は収束にいたる途中の軌道である。

この概念を図 4 に示す。この図で、囲まれた領域は、自己相関型モデルにおいて記憶したパターン（中心の黒点）へ引き込まれる領域である [24]。この領域は一般に複雑な形をしている。そして、自己相関型モデルの場合、この領域の外側から出発すると、たとえ記憶したパターンに近いものであっても、記憶想起は成功せず、擬似アトラクターへトラップされてしまう。しかし、本モデルのようにサイクル型の場合は、ある特定のパターンの近くにトラップされることとなく、次々に他のパターンの近くを訪問する。そしてその都度記憶したパターンに近づくことができる。この効果は、自己相関型のモデルでよく使われた、温度ノイズの効果と似ている。この効果は、もうすこし詳しく次のよ

うに説明できる。収束途中段階での重なり値は、次式で表される。式 4 は自己相関型の場合、式 5 はサイクル型の場合である。自己相関型の場合、ターゲットパターンを ξ^R とすると、

$$\begin{aligned} q(R:t+1) &= N^{-1}(\xi^R \cdot \mathrm{S}(t+1)) \\ &= N^{-1}\sum_{i=1}^{N} sign\{(\xi^R \cdot \mathrm{S}(t)) + \sum_{\mu\neq R}\xi_i^R\xi_i^\mu(\xi^\mu \cdot \mathrm{S}(t))\} \end{aligned} \tag{4}$$

サイクル型の場合、ターゲットパターンは時間によって異なる。時間 t のときのターゲットを $\xi^{R,k}$ とすると、

$$\begin{aligned} q(R,r:t+1) &= N^{-1}(\xi^{R,k+1} \cdot \mathrm{S}(t+1)) \\ &= N^{-1}\sum_{i=1}^{N} sign\{(\xi^{R,k} \cdot \mathrm{S}(t)) + \sum_{(\mu,\nu)\neq(R,k)}\xi_i^{R,k+1}\xi_i^{\mu,\nu+1}(\xi^{\mu,\nu} \cdot \mathrm{S}(t))\} \end{aligned} \tag{5}$$

中括弧内の第一項は、ターゲットパターンによって引かれる力である。重なり値が大きいほどこの項は支配的である。中括弧内の第二項は、第一項の効果を引き離そうとする力である。自己相関型の場合、式 4 の第二項に注目すると、ある特定の i に対して、いくつかのパターンにおいて、$\xi_i^R\xi_i^\mu$ と $(\xi^\mu \cdot \mathrm{S}(t))$ が異符号をもつため、第一項の $(\xi^R \cdot \mathrm{S}(t))$ の値が小さいと、第二項が第一項の影響をひっくり返してしまう。擬似アトラクターの近くでは、毎回特定の i に対してこの現象が起きる。しかし、式 5 のようなサイクル型の場合、ターゲットパターンが一回ごとに $\xi_i^{R,k}$ から $\xi_i^{R,k+1}$ へと入れ替わるので、特定の i に対して、$\xi_i^{R,k+1}\xi_i^{\mu,\nu+1}$ と $(\xi^{\mu,\nu} \cdot \mathrm{S}(t))$ が、毎回異符号を持ち、第二項が第一項をひっくり返すとは限らない。こういったしくみで、第二項に自己平均化効果が働き、擬似アトラクターに陥ることが避けられるわけである。この効果は、Meir Domany の提案した多層型フィードフォワード型神経回路 [25] におけるものと同様の効果である。本研究のようにサイクルを用いれば、この効果が、一層のみの神経回路でも実現できる。また、この第二項による自己平均化効果は、自己相関型モデルで温度を上げた場合にも似ており、自己アニール効果とも呼ぶことができる。

5 各時間ステップにおける重なり値の変化

式 4 の第 2 項、すなわちターゲットパターン以外のパターンとの重なりの効果を、近似して、重なり値 $q(t)$ の時間的変化を議論した研究例は多い。Amari[26] と Kinzel[23] は、式 4 の第二項を Gaussian 分布に従うランダムな項と仮定して、次の式を導いた。

$$q(t+1) = erf(q(t)/\sigma), \qquad \sigma = \sqrt{2M/N} \tag{6}$$

ここで erf は error function である。このレベルの近似では、自己相関型とサイクル型の違いを見ることはできない。Amari ら [24] は自己相関型の場合にさらに近似を上げて、また Meir と Domany は多層型神経回路の場合について厳密に解き [25]、上記 6 式において分散 $\sigma(t)$ も時間ステップに依存して変化するといった漸化式を導き出した。我々は式 4 および 5 の第二項の効果を時間を追って見るため、$N = 400,\ M = 30$ の場合に

ついて、時間tのときの重なりを時間$t-1$のときに対して、Lorentz plot にしたものを図5に示す。

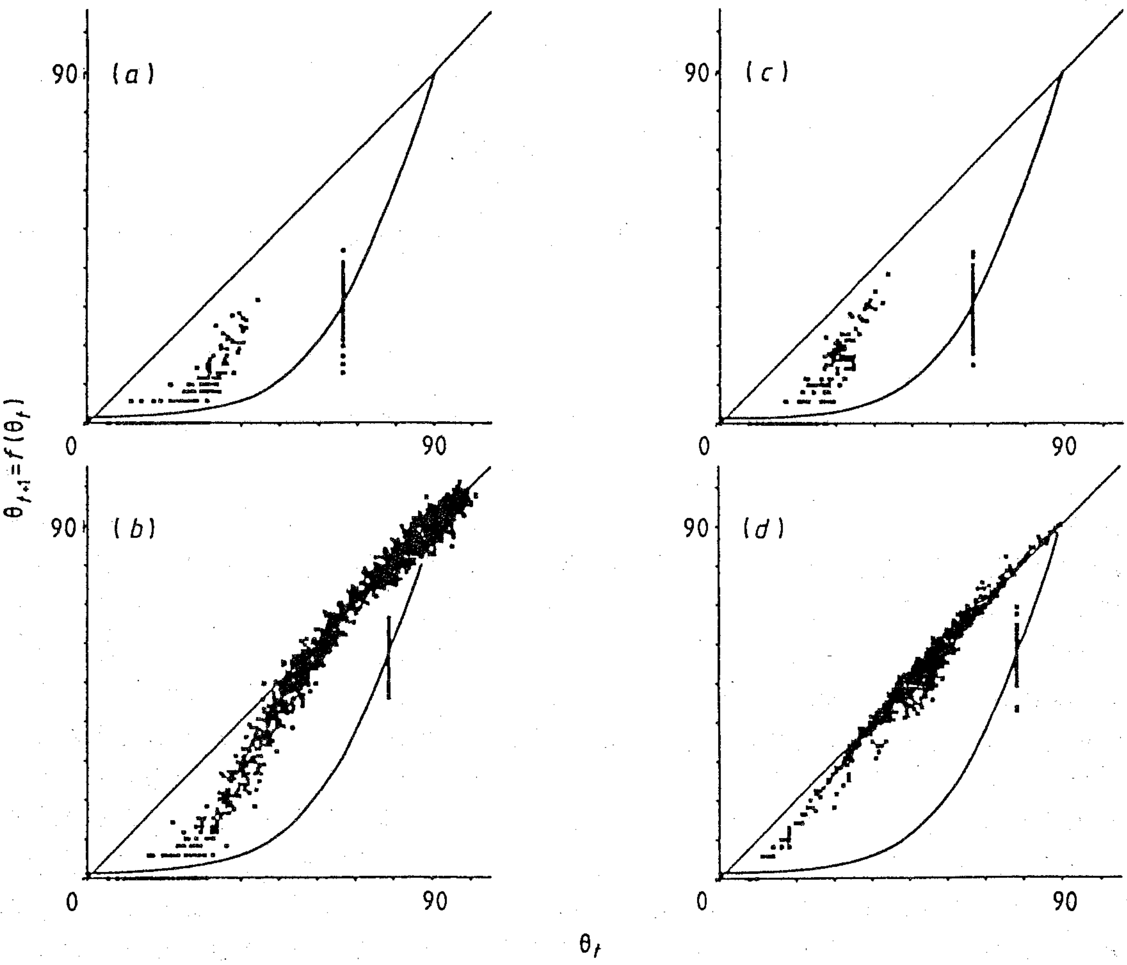

図5. 時間tのときの重なりを時間$t-1$のときの重なりに対してプロットした Lorentz plot。重なりは$\theta = cos^{-1}(q(t))$で示してある。ここで a,b は$L = 10$のサイクルの場合に対応し、c,d は自己相関の場合に対応する。b,d は初期重なりが 0.2 の場合であり、a,c は初期重なりが 0.4 の場合である。

この図で a と b は$L = 10$のサイクルの場合であり、c と d は自己相関の場合である。初期重なりがq_0 =0.2 と 0.4 の場合についてプロットしてある。それぞれの図で、100 回ずつ走らせてプロットした。初期重なり 0.4 のときは、ほとんどの軌道はターゲットパターン ($\theta = 0$) へ収束した。しかし初期重なり 0.2 で自己相関型の場合、かなりの割合が擬似アトラクターにひっかかっている。この図で実線は6式をプロットしたものである。この計算で初期パターンはランダムに選んだので、始めの一ステップのみは6式に従っている。しかし、二ステップめから収束はゆっくりになっている。つまり、$\theta(t+1) = \theta(t)$の直線に近くなる。これは、ターゲットパターン以外のパターンとの重なりがあり、式4、5の第二項の値が大きくなるためである。ターゲットパターンの近く ($\theta \approx 0$) では、ターゲットパターンとの重なりが大きくなるため、Lorentz plot は再び実線に

近づいてくる。図5でサイクルの場合と自己相関の場合とを較べてみよう。初期重なりが大きい(0.4)ときは、aとcとでプロットは似ている。しかし、初期重なりが小さい(0.2)ときは、式4、5の第二項の影響は大きく、bとdとでプロットは異なっている。自己相関の場合が収束はゆっくりであり、しかもほとんどの場合、$\theta = 50$付近の擬似アトラクターへトラップされる。サイクル型の場合は90%以上がターゲットパターン($\theta = 0$)へ収束し、残りは、$\theta = 90$付近の擬似アトラクターまたは目的以外のパターンへ行く。

6 他のパターン遷移型モデルとの比較

次に、時間送れを伴って一連の遷移を引き起こすモデル[15, 16, 17]と比較したい。このタイプのモデルでは、自己相関型の部分に加えて、異なるパターン間に直積を足し合わせた部分が、重み係数λで、また時間遅れτで、組み合わさっている。これらの系では、ある一定時間では一つのパターンに収束するが、さらにその後他のパターンへ遷移し、そこで収束し、また遷移をくりかえす。このモデルの長時間における、パターン間の遷移の振るまいは、それぞれ一時的に収束した時の重なり値で見ることができる。Gutfreundらによると[17]、この重なり値の各遷移での変化の様子が、上記6式と同じ形で表される。この対応は、係数λが充分大きくなった場合か、あるいは分散σをλを含む式で置き換えた場合に可能である。これらのモデルでは、係数λがある特定の値の時に、自己相関型よりも記憶容量が大きくなることがわかった[17]。この記憶容量の拡大は、式6をGutfreundらの方法と対応させて眺めると理解することができる。一方、我々のリミットサイクル型のモデルにおける記憶想起特性の改良は、式6のレベルでは理解することができないため、Gutfreundの記憶容量の拡大とは別の原理にもとづくものである。また、Gutfreundのモデルにおいても、熱的なノイズを加えることは、擬似アトラクターへ引き込まれることを避けるのに効果的である[17]。こういった、時間送れを伴って一連の遷移を引き起こすモデルと、我々のリミットサイクル型モデルを、今回の我々の結果を使いながら比較することは、今後の興味深い課題である。

7 まとめ

まとめると、リミットサイクルを埋め込んだモデルの方が、よく知られた自己相関を埋め込んだモデルよりも、記憶想起特性が良くなることを示した。この理由は擬似アトラクターに陥ることが避けられるためである。このことは、重なり値を求める相手の参照パターンが次々に変化するためであることから説明でき、自己平均化効果と呼ぶことができた。これらを定量的に解析することが残された課題である。このリミットサイクル型モデルは、パラメーターを調整することにより、収束すべきサイクルを不安定化させたり、カオス的な挙動を引き起こしたりすることができる[27]。将来的には、こういった系が示す動力学的現象を、情報検索などへ並列処理的な情報処理機能として応用することが期待できる。

参考文献

[1] Nakano K 1972 *IEEE Trans. Systems, Man, Cybern.* **SMC-2** 380

[2] Amari S 1972 *IEEE Trans. Systems, Man, Cybern.* **SMC-2** 643

[3] Anderson J A 1972 *Math. Biosci.* **14** 197

[4] Kohonen T 1972 *IEEE Trans. Comput. C-21* 353

[5] Hopfield J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554; Hopfield J 1985 *Biol. Cybern.* **52** 141

[6] Amit D J 1987 *Proc. Int. Symp. on the physics of Structure Formation, Tubingen,* 1986 (Berlin: Springer)

[7] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. A* **32** 1007

[8] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30

[9] Feigelman M V and Ioffe L B 1986 *europhys. Lett.* **1** 197

[10] Ackley D H, Hinton G E and Sejnowski T J 1985 *Cog. Sci.* **9** 147

[11] Feigelman M V and Ioffe L B 1987 *Int. J. Mod. Phys. B* **1** 51

[12] Shinomoto S 1987 *Biol. Cybern.* **57** 197

[13] Parisi G 1986 *J. Phys. A: Math. Gen.* **19** L675

[14] Peretto P and Niez J J 1986 *Disordered Systems and Biological Organization* ed E Bienenstock, F Fogelman-Soulie and G Weisbuch (Berlin: Springer)

[15] Kleinfeld D 1986 *Proc. Natl Acad. Sci. USA* **83** 9469

[16] Sompolinsky H and Kanter I 1986 *Phys. Rev. Lett.* **57** 2861

[17] Gutfreund H and Mezard M 1988 *Phys. Rev. Lett.* **61** 235

[18] Buhmann J and Schulten K 1987 *Europhys. Lett.* **4** 1205

[19] Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev. A* **34** 4217

[20] Dehaene S, Changeux J P and Nadal J P 1987 *Proc. Natl Acal. Sci. USA* **84** 2727

[21] Guyon I, Personnaz L, Nadal J P and Dreyfus G 1988 *Phys. Rev. A* **38** 6365

[22] Goles-Chacc E, Fogelman-Soulie F and Pellegrin D 1985 *Discrete Appl. Math.* **12** 261.

[23] Kinzel W 1985 *Z. Phys. B* **60** 205

[24] Amari S and Maginu K 1988 *Neural Networks* **1** 63

[25] Meir R and Domany E 1988 *Phys. Rev. A* **37** 608

[26] Amari S 1977 *Biol. Cybern.* **26** 175

[27] Nara S, Mori Y and Davis P *ATR Technical Report TR-O-0015*