



HOKKAIDO UNIVERSITY

Title	2004年度 混沌系工学特論講義ノート
Author(s)	井上, 純一; Inoue, Jun-ichi
Description	当講義資料は著者のホームページ http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/ からダウンロードできます。 http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Issue Date	2004
Doc URL	https://hdl.handle.net/2115/370
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	learning object
File Information	konton2004_4.pdf, 第4回講義ノート



混沌系工学特論 配布資料 #4

担当：井上 純一 (情報科学研究科棟 8-13)

URL : http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

平成 16 年 11 月 15 日

目次

1.12 想起過程のダイナミクス	41
1.12.1 シグナル-ノイズ解析	42
1.12.2 ノイズ項を固定分散ガウスで近似した発展方程式	43
1.12.3 計算機シミュレーションによるチェックと問題点	44
1.12.4 Amari-Maginu 理論	45
1.12.5 付録： m_t, σ_t に関する発展方程式の導出	47
1.13 学習の結果としての Hebb 則：もう一つの時間スケール	54
1.14 対称結合回路網とそのエネルギー関数	56

1.12 想起過程のダイナミクス

今まで述べてきた連想記憶に関する解析は各素子が同期的に

$$S_i(t+1) = \operatorname{sgn} \left(\frac{1}{N} \sum_{j \neq i} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} S_j(t) \right) \quad (135)$$

で記述される状態変更式に従って自らの状態を更新して行った後の収束点 ($t \rightarrow \infty$) に関するものであった。つまり、十分に時間が経った後に全ての素子 i に対して

$$\lim_{t \rightarrow \infty} S_i(t) = \xi_i^1 \quad (136)$$

となるための条件：

$$\xi_i^1 = \operatorname{sgn} \left(\frac{1}{N} \sum_{j \neq i} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \xi_j^1 \right) \quad (137)$$

をクロストークノイズを評価することにより解析したり (ξ^1 : ターゲットパターン), あるいは確率的動作をする神経素子が非同期的に状態更新している場合, 十分時間が経過した後の素子状態に関する時間平均：

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T S_i(t) = \langle S_i \rangle \quad (138)$$

に対して平均場方程式を作り, それを解析することにより, 回路網に記憶させることのできるパターン数を評価した。しかし, 初期状態 $S(0)$ から固定点 $S(\infty)$ へ至る過程を調べることで, 平衡系 (十分時間が経過

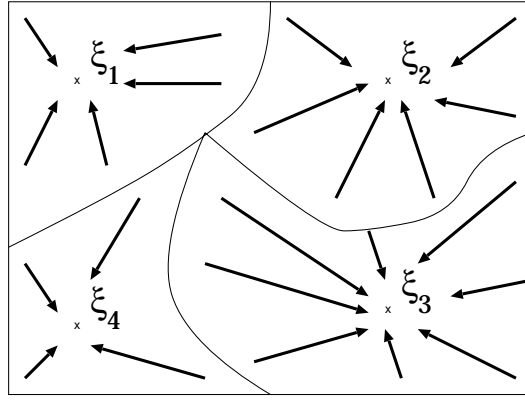


図 25: 連想記憶の想起過程の概念図. 初期状態の選び方によっては, 想起したいパターン以外の固定点に引き込まれる可能性がある. それぞれのパターンに引き込まれる領域のことを引き込み領域 (basin) と呼ぶ.

した後の系)の解析では得られない様々な連想記憶に関する性質や概念が明確になる (図 25 参照). そこで, ここでは記憶の「想起過程」のダイナミクスを調べる代表的な理論である Amari-Maginu 理論を紹介することにしよう.

ただし, ここでは各素子が (135) 式に従って状態変更して行くものを採用する. このときに問題となるのは, 素子群が一斉に状態を変更するのか (同期式状態変更), あるいは, 個々に状態変更するのか (非同期式状態変更) という点であるが, ここでは神経素子群が前者の更新様式をとるものとする. すなわち

$$\begin{pmatrix} S_1(0) \\ S_2(0) \\ \vdots \\ S_N(0) \end{pmatrix} \rightarrow \begin{pmatrix} S_1(1) \\ S_2(1) \\ \vdots \\ S_N(1) \end{pmatrix} \rightarrow \begin{pmatrix} S_1(2) \\ S_2(2) \\ \vdots \\ S_N(2) \end{pmatrix} \rightarrow \begin{pmatrix} S_1(3) \\ S_2(3) \\ \vdots \\ S_N(3) \end{pmatrix} \rightarrow \dots \quad (139)$$

で状態変更するものとして議論を進める.

1.12.1 シグナル - ノイズ解析

ダイナミクスを扱う場合でも, 既に学んだ平衡状態の解析方法 — シグナル - ノイズ解析 (SN 解析) が役に立つ. そこで, まず時刻 t における素子 i への入力:

$$h_i(t) = \frac{1}{N} \sum_{j=1}^N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu S_j(t) \quad (140)$$

をとり, ターゲットパターン ξ^1 を想起するプロセスを考える. このとき, ターゲットパターン ξ^1 の安定性解析で行なったように $h_i(t)$ を $\mu = 1$ の項と, それ以外の項に分離して

$$\begin{aligned} h_i(t) &= \frac{1}{N} \sum_j \xi_i^1 \xi_j^1 S_j(t) + \frac{1}{N} \sum_j \sum_{\mu \neq 1} \xi_i^\mu \xi_j^\mu S_j(t) \\ &= \xi_i^1 \left(\frac{1}{N} \sum_j \xi_j^1 S_j(t) \right) + \frac{1}{N} \sum_j \sum_{\mu \neq 1} \xi_i^\mu \xi_j^\mu S_j(t) \equiv \xi_i^1 m_t + N_i^t \end{aligned} \quad (141)$$

としておく. ここで

$$m_t \equiv \frac{1}{N} \sum_j \xi_j^1 S_j(t) \quad (142)$$

$$N_i^t \equiv \frac{1}{N} \sum_j \sum_{\mu \neq 1} \xi_i^\mu \xi_j^\mu S_j(t) \quad (143)$$

とおいた. この N_i^t は 2 番目以降のパターンからのノイズである.

1.12.2 ノイズ項を固定分散ガウスで近似した発展方程式

既に学んだ SN 解析に従えば, 我々が次にすべきことは, 確率変数 N_i^t の統計的な性質を調べることである. そこで, N_i^t は多数の ± 1 の和であり, 各項が独立であると仮定する. すると中心極限定理より N_i^t は平均がゼロ, 分散 σ^2 が

$$\sigma^2 = \frac{1}{N^2} \sum_j \sum_{\mu \neq 1} (\xi_i^\mu \xi_j^\mu S_j(t))^2 = \frac{1}{N^2} \times N \times (p-1) \simeq \alpha \left(\equiv \frac{p}{N} \right) \quad (144)$$

のガウス分布に従う. よって, 時刻 $t+1$ での重なり m_{t+1} は

$$\begin{aligned} m_{t+1} &= \frac{1}{N} \sum_j \xi_j^1 S_j(t+1) \\ &= \frac{1}{N} \sum_j \xi_j^1 \text{sgn}(h_j(t)) \\ &= \frac{1}{N} \sum_j \xi_j^1 \text{sgn}(\xi_j^1 m_t + N_j(t)) \\ &= \frac{1}{N} \sum_j (\xi_j^1 \xi_j^1) \text{sgn}(\xi_j^1 \xi_j^1 m_t + \xi_j^1 N_j(t)) \\ &= \frac{1}{N} \sum_j \text{sgn}(m_t + \xi_j^1 N_j(t)) \end{aligned} \quad (145)$$

となる.

さて, N_i^t が平均ゼロ, 分散 $\sigma^2 = \alpha$ のガウス分布に従うのであるから, $x \equiv \xi_j^1 N_j^t$ もやはり, 平均ゼロ, 分散 σ^2 のガウス分布に従う. よって $\text{sgn}(\dots)$ の自己平均性から

$$\begin{aligned} m_{t+1} &= \frac{1}{N} \sum_j \text{sgn}(m_t + \xi_j^1 N_j(t)) \\ &= \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}} \text{sgn}(m_t + x) = \sqrt{\frac{2}{\pi}} \int_0^{m_t/\sigma} dz e^{-\frac{z^2}{2}} \end{aligned} \quad (146)$$

が得られ, 回路網の状態 $S(t)$ と想起するパターン ξ^1 の重なり m_t の時間発展は

$$m_{t+1} = \sqrt{\frac{2}{\pi}} \int_0^{m_t/\sigma} dz e^{-\frac{z^2}{2}} = 1 - 2H\left(\frac{m_t}{\sqrt{\alpha}}\right) \quad (147)$$

と書ける. ここで, $\sigma = \sqrt{\alpha}$ であり, $H(x)$ を $H(x) = \int_x^\infty Dz, Dz = dz e^{-z^2/2}/\sqrt{2\pi}$ で定義した. この (147) 式から直ちにわかることは, 十分時間が経過した後に重なりが $m_{t+1} = m_t = m$ という一定値に収束したとすると, $\alpha = 0$, 及び, $\alpha = \infty$ でそれぞれ

$$m(\alpha = 0) = 1 - 2H(\infty) = 1 \quad (148)$$

$$m(\alpha = \infty) = 1 - 2H(0) = 0 \quad (149)$$

であるから, $0 < \alpha < \infty$ の間に必ず, 重なり m が有限値からゼロに変化する点が存在する. この α の増加に伴う想起相から非想起相への変化が 2 次転移だとすれば, (147) 式を m で展開することにより, $\alpha_c = 2/\pi$ が得られる. これは, 先に見た非同期的状態更新の場合の値 $\alpha_c = 0.138$ と比べてかなり大きい. そこで, 簡単にこのダイナミクスの式 (147) の妥当性を手っ取り早くチェックするために, 計算機シミュレーションを用いて (147) 式からの結果と比較してみることにしよう.

1.12.3 計算機シミュレーションによるチェックと問題点

図 26 にノイズ項を固定分散ガウスで近似した場合の想起過程のダイナミクス: $m_{t+1} = 1 - 2H(m_t/\sqrt{\alpha})$ をいくつかの α の値に対してプロットしたものを載せる. これらの図から, いずれの初期状態から出発して

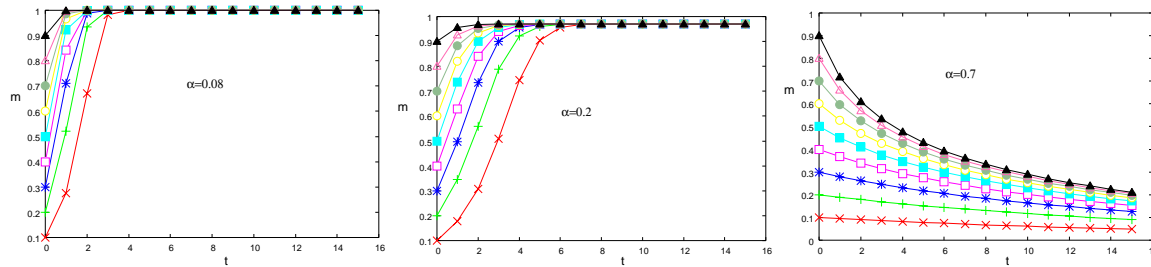


図 26: ノイズを単純に分散 $\sigma^2 = p/N$ とした場合の想起過程のダイナミクス (147). 左から, $\alpha = 0.08, 0.2, 0.7$ に対する結果.

も α の値によって決まる一定値に単調に収束することがわかる. 次いで, 図 27 に計算機シミュレーションによる結果を載せる. このシミュレーションでは神経素子数を $N = 3600$ に選んである. $\alpha = 0.08$ の場合の (147) 式からの結果と比較するために, 回路網に記憶させるパターン数を $p = \alpha N = 0.08 \times 3600 = 288$ と選んだ. この図 27 からわかるように, 初期状態の選び方により, 重なりの時間発展に「閾値現象」が見られ, (147) 式からの結果 (図 26 の左側) とは明らかに異なる. 従って, この結果を見る限りにおいて, (147) 式は間違っており, 近似として考えてみても回路網の定性的振る舞いを全くとらえていないことがわかる.

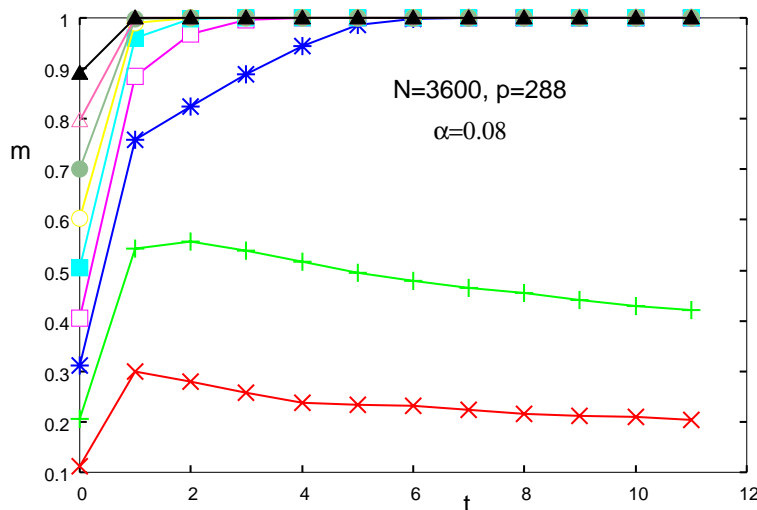


図 27: $N = 3600, p = 288, \alpha = p/N = 0.08$ での計算機シミュレーションの結果. 初期状態の選び方による閾値現象が見られる.

どこに問題があったのであろうか？そこで、確率変数 N_i^t に立ち帰って考えて直してみると、この N_i^t は時刻 t 以前の各素子の「履歴」に非常に複雑な形で依存しているわけであり、我々が方程式 (147) を導出する際に用いた「 N_i^t は平均ゼロ、分散 σ^2 のガウス分布に従う」という仮定が果たして妥当なものであるかどうかは一見しただけではわからない。この状況で具体的に我々が調べるべき点は次の2つであろう。

- N_i^t が従う分布は何か？
- N_i^t がガウス分布に従うことは認めるとすると、その平均と分散の時間発展はどうなるのか？

これらはいずれも、もっともな疑問である。1番目の疑問に関しては、中心極限定理が適用できるのは、あくまでも N_i^t を構成する各項が独立である場合であるから、上の場合、ガウス分布から大きくずれる可能性は大いにあり得ることである。また、2番目の問題に関して言えば、上の (147) では平均ゼロで分散 σ^2 が時間に依らず一定であるとおいたが、分散に時間依存性を持たせた場合、ダイナミクスはどう変わるのか、というのは調べるべき重要事項である。

現段階で言うと、1番目の問題は完全には解っておらず¹、従って、この講義で1番目の答えを述べることはできない。一方、2番目の問題に関しては、 σ^2 を時間に依存する分散 σ_t^2 としてこの σ_t^2 に関する閉じた方程式を書き下して、 (m_t, σ_t) でもって想起過程のダイナミクスを記述するというアイデアが Amari と Maginu によって提唱され (1988)、今では代表的な近似理論の一つとなっている²。そこで次節ではこの理論の詳細を見て行くことにしよう。

1.12.4 Amari-Maginu 理論

まずは N_i^t がガウス分布に従うものとしよう。このとき、Amari と Maginu はこのガウス分布が平均ゼロ、分散 σ_t^2 を持つとした。このとき、重なり m の時間発展は

$$m_{t+1} = \sqrt{\frac{2}{\pi}} \int_0^{m_t/\sigma_t} dz e^{-\frac{z^2}{2}} \tag{150}$$

となる。

ここで重要なのは m_t, σ_t で閉じた方程式が得られるということであるから、 σ_t の時間発展が σ_t, m_t のみで書けるかどうか解析上の焦点となる。これ以降の計算はやや複雑なので、その導出は各自の自習に任せ、計算結果をはじめに示そう。

解析の結果、重なり m_t 、分散 σ_t^2 は次の方程式：

$$m_t = \operatorname{erf}\left(\frac{m_{t-1}}{\sqrt{2}\sigma_{t-1}}\right) \tag{151}$$

$$\sigma_t^2 = \alpha + 4P\left(\frac{m_{t-1}}{\sigma_{t-1}}\right)^2 + 4\alpha\left(\frac{m_{t-1}m_t}{\sigma_{t-1}}\right)P\left(\frac{m_{t-1}}{\sigma_{t-1}}\right) \tag{152}$$

となる。ただし、 $P(x) \equiv (1/\sqrt{2\pi})e^{-\frac{x^2}{2}}$ であり、誤差関数を $\operatorname{erf}(x) = (2/\sqrt{\pi})\int_0^x dz e^{-z^2}$ で定義している。以下にこれらの方程式を記憶容量 $\alpha = 0.08, 0.2$ の2つの場合に関して解いた結果を載せておこう。ところで、ガウス分布の正当性であるが、想起が成功する場合にはガウス近似が正しく、想起が失敗する場合にはガウス近似から外れてくることがわかっている。しかし、ガウス近似から外れた場合、クロストーク項がどのような分布になるのかに関してはわかっていない。これらの図から、 m_t の時間発展では計算機シミュ

¹ ある条件下 (後述) ではガウス分布の近似が正しく、ある条件では正しくないことはわかっている。正しくない場合、それがどのような分布になるのかに関してはわかっていない。

² その後、この理論は Okada(1995) によって近似精度の改善がなされた

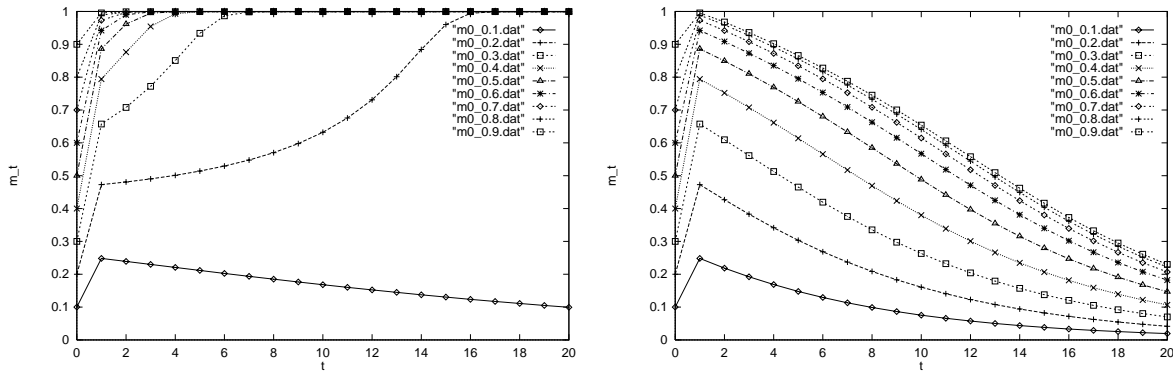


図 28: 重なり m の時間発展. 記憶させたパターン数が $p = 0.08N$ の場合 (左) と $p = 0.2N$ の場合 (右). $p = 0.08N$ では標準偏差の初期値は $\sigma_{t=0} = 0.1$ に選んである. 重なりの初期値が小さいと想起に失敗することがわかる. $p = 0.2N$ の場合は重なりの初期値をどんなに正解 ($m = 1$) に近い値に選んでも想起に失敗することがわかる. 最大記憶パターン以上のパターンをネットワークに記憶させてしまったためである.

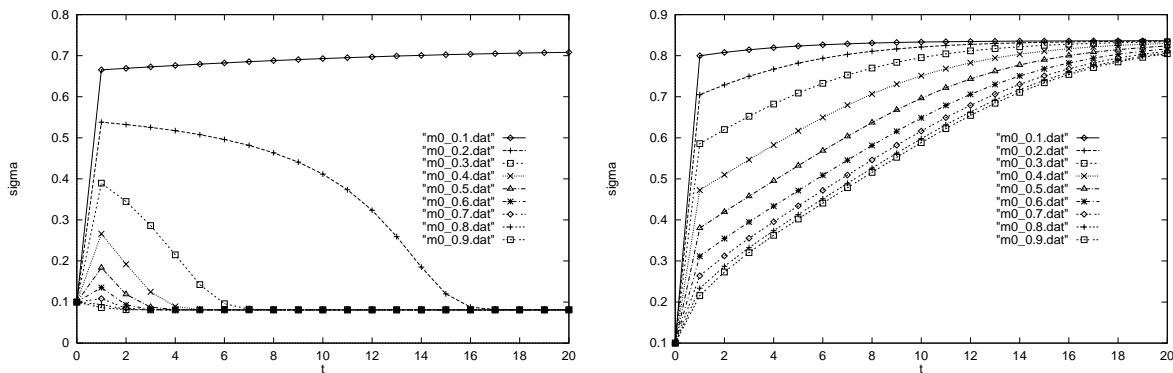


図 29: 標準偏差 σ の時間発展. 記憶させたパターン数が $p = 0.08N$ の場合 (左) と $p = 0.20$ の場合 (右). 標準偏差の初期値は $\sigma_{t=0} = 0.1$ に選んである. 重なりの初期値は $m_{t=0} = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ の 9 通りに選んだ. $p = 0.08N$ の場合には重なりの初期値が小さいと, クロストーク項の分散が成長し, 注目するパターンの想起を妨害することがわかる. $p = 0.2N$ の場合には全ての重なりの初期値に対しクロストーク項の分散が成長し, 注目するパターンの想起を妨害することがわかる.

レーションで確認できたような初期状態の選び方による閾値現象が見られ, また定性的な振る舞いも計算機シミュレーションの結果と良く合っていることがわかる.

問 5: 得られた m_t, σ_t に関する方程式: (151)(152) で十分時間が経過した後に m_t, σ_t が一定値: m, σ 収束したとすれば (151)(152) 両式からはそれぞれ

$$m = \operatorname{erf}\left(\frac{m}{\sqrt{2}\sigma}\right)$$

$$\sigma^2 = \alpha + 4P\left(\frac{m}{\sigma}\right)^2 + 4\alpha\left(\frac{m^2}{\sigma}\right)P\left(\frac{m}{\sigma}\right)$$

なる方程式が得られる. この連立方程式をニュートン法などを用いて数値的に解くことにより, 記憶容量 α_c を求めよ (記憶容量 α_c はこれまでと同様に α を大きくしていったときに $m = 0$ のみが解として得られる α として定義される). この値は先に求めた $\alpha_c = 0.138$ と同じになるか, あるいは異なるか? 異なる場合にはその理由として考えられる点を述べよ.

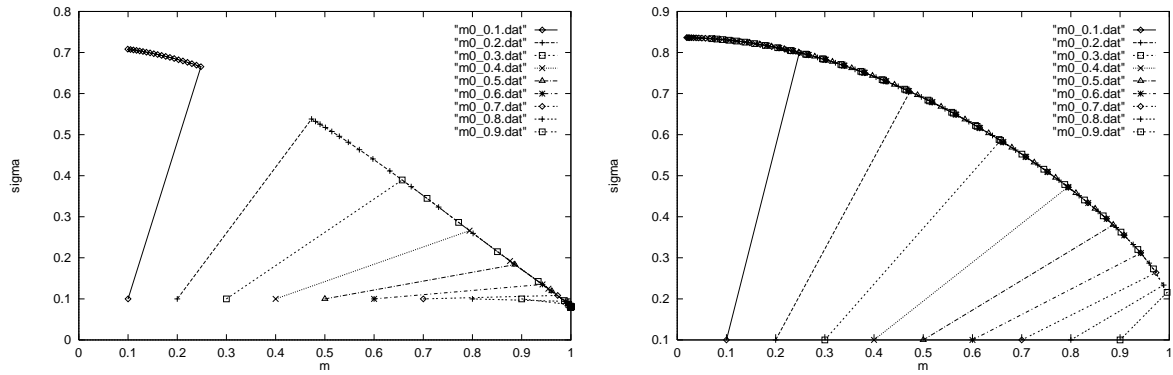


図 30: 記憶させたパターン数が $p = 0.08N$ の場合 (左) と $p = 0.2N$ の場合の重なり m_t と標準偏差 σ_t の流れ図.

問 6: ここでの議論で仮定した時間依存するクロストークノイズ:

$$N_i^t = \frac{1}{N} \sum_j \sum_{\mu \neq 1} \xi_i^\mu \xi_j^\mu S_j(t)$$

のガウス性を各更新ステップ t で N_i^t の 3 次のキュムラントを計算することによりチェックせよ. 計算には計算機シミュレーションを用いる. 具体的な手続きとしては, 同じ ξ^μ ($\mu = 1, \dots, p$) から作られる結合を持つ回路網の動作を S の初期状態を変えて複数回実行させ, そこから得られる各更新ステップ t でのアンサンブルで $C_3(N_i(t)) = \langle N_i(t)^3 \rangle - 3\langle N_i(t) \rangle \langle N_i(t)^2 \rangle + 2\langle N_i(t) \rangle^3$ を計算する.

1.12.5 付録: m_t, σ_t に関する発展方程式の導出

付録 (興味のある者は自習せよ)

まず σ_t^2 を具体的に書き出してみると

$$\sigma_t^2 + \{E[N_i^t]\}^2 = E[(N_i^t)^2] = \frac{1}{N^2} \sum_{jj'} \sum_{\mu\mu'} E \left[\xi_i^\mu \xi_j^\mu S_j(t) \xi_i^{\mu'} \xi_{j'}^{\mu'} S_{j'}(t) \right] \quad (153)$$

となる. ここで $E(x)$ は確率変数 x の期待値を意味する. 以下で μ, μ', j, j' の添え字間の関係に応じていくつかの場合にわけて考えてみる.

(1) $\mu = \mu', j = j'$ の場合

項の値は 1, 項の数は $(p-1)(N-1)$ である.

(2) $\mu \neq \mu', j = j'$ の場合

$$E[\xi_i^\mu \xi_j^\mu \xi_i^{\mu'} \xi_j^{\mu'}] = 0 \quad (\text{独立なパターンだから})$$

(3) $\mu = \mu', j \neq j'$ の場合

$$v = E[\xi_j^\mu \xi_{j'}^\mu S_j(t) S_{j'}(t)] \quad (154)$$

項の数 $(p-1)(N-1)(N-2)$.

(4) $\mu \neq \mu', j \neq j'$ の場合

$$v' = E[\xi_j^\mu \xi_{j'}^\mu S_j(t) \xi_i^{\mu'} \xi_{j'}^{\mu'} S_{j'}(t)] \quad (155)$$

項の数 $(p-1)(p-2)(N-1)(N-2)$.

従って、以下では v, v' を評価して行くことにする.

(A) v の評価

$$S_j(t) = \text{sgn}(h_j(t)) = \text{sgn} \left(\xi_j^1 m_{t-1} + \frac{1}{N} \sum_{k \neq j} \sum_{\nu=2}^p \xi_j^\nu \xi_k^\nu S_k(t-1) \right) \quad (156)$$

であるが、(153) 式右辺 $E[\dots]$ に現れる $\xi_j^\mu, \xi_{j'}^\mu, S_j(t), S_{j'}(t)$ の各項の相関を見たいのだから、 $h_j(t)$ の中の ν についての和をの中で μ の項と、それ以外の項に分離すると

$$h_j(t) = \xi_j^1 m_{t-1} + \frac{1}{N} \sum_{k \neq j} \sum_{\nu \neq \mu} \xi_j^\nu \xi_k^\nu S_k(t-1) + \frac{1}{N} \sum_{k \neq j} \xi_j^\mu \xi_k^\mu S_k(t-1) \quad (157)$$

となる。さらに、この $h_j(t)$ の最後の項を $k = j'$ とそれ以外に分離すれば

$$\begin{aligned} h_j(t) &= \xi_j^1 m_{t-1} + \frac{1}{N} \sum_{k \neq j} \sum_{\nu \neq \mu} \xi_j^\nu \xi_k^\nu S_k(t-1) + \frac{1}{N} \xi_j^\mu \xi_{j'}^\mu S_{j'}(t-1) + \frac{1}{N} \sum_{k \neq j, j'} \xi_j^\mu \xi_k^\mu S_k(t-1) \\ &\equiv \xi_j^1 m_{t-1} + Q + \frac{1}{N} \xi_j^\mu \xi_{j'}^\mu S_{j'}(t-1) + \xi_j^\mu R \end{aligned} \quad (158)$$

と書ける。ここで

$$Q \equiv \frac{1}{N} \sum_{k \neq j} \sum_{\nu \neq \mu} \xi_j^\nu \xi_k^\nu S_k(t-1) \quad (159)$$

$$R \equiv \frac{1}{N} \sum_{k \neq j, j'} \xi_k^\mu S_k(t-1) \quad (160)$$

とおいた。

さて、 Q, R は平均がゼロのガウス分布に従うと仮定すると、その分散は

$$Q: \quad \frac{1}{N^2} \sum_{k \neq j} \sum_{\nu \neq \mu} (\xi_j^\nu \xi_k^\nu S_k(t-1))^2 = \frac{1}{N^2} (N-1)(p-1) \simeq \alpha \equiv \sigma_{t-1}^2 \quad (161)$$

$$R: \quad \frac{1}{N^2} \sum_{k \neq j, j'} (\xi_k^\mu S_k(t-1))^2 = \frac{1}{N^2} \times (N-2) \simeq \frac{1}{p} \frac{p}{N} = \frac{\sigma_{t-1}^2}{p} \quad (162)$$

となる.

全く同様にして $S_{j'}(t)$ も v の表式の $E[\dots]$ 内の別の因子に依存する部分と, 他の部分に分離すると

$$S_{j'}(t) = \text{sgn} \left(\xi_{j'}^1 m_{t-1} + Q' + \xi_{j'}^\mu R + \frac{1}{N} \xi_{j'}^\mu \xi_{j'}^\mu S_j(t-1) \right) \quad (163)$$

$$Q' \equiv \frac{1}{N} \sum_{k \neq j'} \sum_{\nu \neq \mu} \xi_{j'}^\nu \xi_k^\mu S_k(t-1) \quad (164)$$

となる. ここで Q' は平均ゼロ, 分散 σ_{t-1}^2 のガウス分布である.

次に, $\xi_j^\mu, \xi_{j'}^\mu$ の値に応じて期待値 E を求める. そこで

$$Y_{pq} \equiv E \left[\xi_j^\mu \xi_{j'}^\mu S_j(t) S_{j'}(t) \mid \xi_j^\mu = p, \xi_{j'}^\mu = q \right] \quad (165)$$

で Y_{pq} を定義すると, v は

$$v = \frac{1}{4} (Y_{11} + Y_{1-1} + Y_{-11} + Y_{-1-1}) \quad (166)$$

と書けるので $Y_{11}, Y_{1-1}, Y_{-11}, Y_{-1-1}$ をそれぞれ評価すればよい.

また, 1 番目のパターン ξ^1 を $\xi^1 = (1, 1, \dots, 1)$ としても一般性を失わないので, 以下このようにおくことにする. つまり, $\xi_{j'} = \xi_j^1 = 1$ である.

このとき Y_{11}, Y_{1-1} は

$$\begin{aligned} Y_{11} &= E[S_j(t) S_{j'}(t)] \\ &= E \left[\text{sgn} \left(m_{t-1} + Q + R + \frac{1}{N} S_{j'}(t-1) \right) \text{sgn} \left(m_{t-1} + Q' + R + \frac{1}{N} S_j(t-1) \right) \right] \end{aligned} \quad (167)$$

$$\begin{aligned} Y_{1-1} &= -E[S_j(t) S_{j'}(t)] \\ &= -E \left[\text{sgn} \left(m_{t-1} + Q + R - \frac{1}{N} S_{j'}(t-1) \right) \text{sgn} \left(m_{t-1} + Q' - R - \frac{1}{N} S_j(t-1) \right) \right] \end{aligned} \quad (168)$$

と書ける. ここで, 確率変数 $Q \pm R, Q' \pm R$ は平均ゼロ, 分散 1 のガウス分布に従う確率変数 u, v を用いて次のように書き直せることに注意しよう.

$$Q \pm R = \sigma_{t-1} \left(u \pm \frac{1}{2p} v \right) \quad (169)$$

$$Q' \pm R = \sigma_{t-1} \left(v \pm \frac{1}{2p} u \right) \quad (170)$$

[念のための証明]

上の左辺, 右辺での分散, 共分散が等しいことを示せばよい.

まず, 分散に関して (169) 式の左辺では

$$E[(Q \pm R)^2] = E[Q^2] \pm 2E[Q]E[R] + E[R^2] = \sigma_{t-1}^2 + \frac{\sigma_{t-1}^2}{p} \simeq \sigma_{t-1}^2 \quad (171)$$

一方, (169) 式の右辺では

$$E \left[\sigma_{t-1}^2 \left(u \pm \frac{v}{2p} \right)^2 \right] = \sigma_{t-1}^2 E[u^2] \pm \sigma_{t-1}^2 E[u] E \left[\frac{v}{2p} \right] \pm \frac{\sigma_{t-1}^2}{(2p)^2} E[v^2] \simeq \sigma_{t-1}^2 \quad (172)$$

となり, 両者は一致する. (170) 式に関しても全く同様にして両辺の分散が一致することが示せる.

次に共分散に関して (169)(170) 両式の左辺どうしの積の期待値は

$$E[(Q \pm R)(Q' \pm R)] = \pm E[R^2] = \pm \frac{\sigma_{t-1}^2}{p} \quad (173)$$

となる. 一方 (169)(170) 両式の右辺どうしの積の期待値は

$$E \left[\sigma_{t-1}^2 \left(u \pm \frac{v}{2p} \right) \left(v \pm \frac{u}{2p} \right) \right] = \pm \frac{\sigma_{t-1}^2}{2p} (E[u^2] + E[v^2]) = \pm \frac{\sigma_{t-1}^2}{p} \quad (174)$$

となり共分散も一致する. □

従って, 以下では使いやすい u, v を用いて議論を進めていく. まず, この変換の下で Y_{11}, Y_{1-1} は

$$\begin{aligned} Y_{11} &= E \left[\operatorname{sgn} \left(m_{t-1} + \sigma_{t-1}u + \frac{\sigma_{t-1}v}{2p} + \frac{S_{j'}(t-1)}{N} \right) \operatorname{sgn} \left(m_{t-1} + \sigma_{t-1}v + \frac{\sigma_{t-1}u}{2p} + \frac{S_j(t-1)}{N} \right) \right] \\ &= E \left[\operatorname{sgn} \left(\frac{m_{t-1}}{\sigma_{t-1}} + u + \frac{v}{2p} + \frac{S_{j'}(t)}{N\sigma_{t-1}} \right) \operatorname{sgn} \left(\frac{m_{t-1}}{\sigma_{t-1}} + v + \frac{u}{2p} + \frac{S_j(t-1)}{N\sigma_{t-1}} \right) \right] \\ &= E \left[\operatorname{sgn}(\bar{m} + u + \tilde{v} + \tilde{x}') \operatorname{sgn}(\bar{m} + v + \tilde{u} + \tilde{x}) \right] \end{aligned} \quad (175)$$

となる. ここで

$$\bar{m} \equiv \frac{m_{t-1}}{\sigma_{t-1}} \quad (176)$$

$$\tilde{v} \equiv \frac{v}{2p} \quad (177)$$

$$\tilde{u} \equiv \frac{u}{2p} \quad (178)$$

$$\tilde{x}' \equiv \frac{S_{j'}(t-1)}{N\sigma_{t-1}} \quad (179)$$

$$\tilde{x} \equiv \frac{S_j(t-1)}{N\sigma_{t-1}} \quad (180)$$

で新しい変数を導入した. Y_{1-1} に関してもこれらの変数で書き直せば $Y_{11} + Y_{1-1}$ は

$$\begin{aligned} Y_{11} + Y_{1-1} &= E[\operatorname{sgn}\{(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v + \tilde{u} + \tilde{x})\}] - E[\operatorname{sgn}\{(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v + \tilde{u} - \tilde{x})\}] \\ &= \operatorname{Prob}[(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v + \tilde{u} + \tilde{x}) > 0] \\ &\quad - \operatorname{Prob}[(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v + \tilde{u} + \tilde{x}) < 0] \\ &\quad - \operatorname{Prob}[(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v - \tilde{u} - \tilde{x}) > 0] \\ &\quad + \operatorname{Prob}[(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v - \tilde{u} - \tilde{x}) < 0] \\ &= 2\operatorname{Prob}[(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v + \tilde{u} + \tilde{x})] \\ &\quad - 2\operatorname{Prob}[(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v - \tilde{u} - \tilde{x})] \end{aligned} \quad (181)$$

となる. ここで $\operatorname{Prob}[A > 0]$ は確率変数 A が $A > 0$ を満たす確率を表す. また, 最後の式変形では

$$\operatorname{Prob}[A > 0] + \operatorname{Prob}[A < 0] = 1 \quad (182)$$

を用いた.

さて, $p = N\alpha$ であったので p は N のオーダーであり, 従って $\tilde{u}, \tilde{v}, \tilde{x}, \tilde{x}'$ は \bar{m}, u, v に比べて非常に小さい. よって, これら小さな量の 2 次以上の積 ($\tilde{x}\tilde{v}$ 等) は無視できる. このとき $Y_{11} + Y_{1-1}$ はさらに簡単な形

にすることができる. つまり

$$\begin{aligned} Y_{11} + Y_{1-1} &= 2\text{Prob}[(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v) > 0] - 2\text{Prob}[(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v) > 0] \\ &+ 2\text{Prob}[(\bar{m} + u)(\bar{m} + v + \tilde{u} + \tilde{x}) > 0] - 2\text{Prob}[(\bar{m} + u)(\bar{m} + v - \tilde{u} - \tilde{x}) > 0] \\ &\equiv F_1 + F_2 \end{aligned} \quad (183)$$

となる. ここで

$$\frac{F_1}{2} \equiv \text{Prob}[(\bar{m} + u + \tilde{v} + \tilde{x}')(\bar{m} + v) > 0] - \text{Prob}[(\bar{m} + u - \tilde{v} - \tilde{x}')(\bar{m} + v) > 0] \quad (184)$$

$$\frac{F_2}{2} \equiv \text{Prob}[(\bar{m} + u)(\bar{m} + v + \tilde{u} + \tilde{x}) > 0] - \text{Prob}[(\bar{m} + u)(\bar{m} + v - \tilde{u} - \tilde{x}) > 0] \quad (185)$$

で F_1, F_2 を定義した.

上の F_1 を評価しよう.

$$f(v) \equiv -\tilde{v} - \bar{m} - \tilde{x}' \quad (186)$$

$$g(v) \equiv \tilde{v} - \bar{m} + \tilde{x}' \quad (187)$$

とおき, u, v の分布が

$$P(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad P(v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \quad (188)$$

であったことを思い出すと

$$\begin{aligned} \frac{F_1}{2} &= \int_{-\bar{m}}^{\infty} P(v)dv \int_{f(v)}^{\infty} P(u)du + \int_{-\infty}^{-\bar{m}} P(v)dv \int_{-\infty}^{f(v)} P(u)du \\ &- \int_{-\bar{m}}^{\infty} P(v)dv \int_{g(v)}^{\infty} P(u)du - \int_{-\infty}^{-\bar{m}} P(v)dv \int_{-\infty}^{g(v)} P(u)du \\ &= \int_{-\infty}^{-\bar{m}} P(v)dv \int_{g(v)}^{f(v)} P(u)du + \int_{-\bar{m}}^{\infty} P(v)dv \int_{f(v)}^{g(v)} P(u)du \end{aligned} \quad (189)$$

となる. ここで

$$f(v) - g(v) = -2(\tilde{v} + \tilde{x}') \sim \mathcal{O}\left(\frac{1}{N}\right) \quad (190)$$

であるから, 積分区間 $[f(v), g(v)], [g(v), f(v)]$ の積分を次のように書き直すことができる.

$$\begin{aligned} \int_{g(v)}^{f(v)} P(u)du &\simeq (f - g)P\left(\frac{f + g}{2}\right) \\ &= (f - g)P(-\bar{m}) \\ &= (f - g)P(\bar{m}) \end{aligned} \quad (191)$$

$$\int_{f(v)}^{g(v)} P(u)du \simeq (g - f)P(\bar{m}) \quad (192)$$

従って $F_1/2$ は

$$\frac{F_1}{2} = P(\bar{m}) \left\{ \int_{-\infty}^{-\bar{m}} (f - g)P(v)dv + \int_{-\bar{m}}^{\infty} (g - f)P(v)dv \right\}$$

$$\begin{aligned}
&= P(\bar{m}) \left\{ \int_{-\infty}^{-\bar{m}} \left(-\frac{v}{p} - \frac{2S_{j'}(t-1)}{N\sigma_{t-1}} \right) P(v) dv \right. \\
&+ \left. \int_{-\bar{m}}^{\infty} \left(\frac{v}{p} + \frac{2S_{j'}(t-1)}{N\sigma_{t-1}} \right) P(v) dv \right\} \\
&= \frac{P(\bar{m})}{N} \left\{ \frac{1}{\alpha} \int_{-\bar{m}}^{\infty} v P(v) dv - \frac{1}{\alpha} \int_{-\infty}^{\bar{m}} v P(v) dv \right. \\
&+ \left. \frac{2S_{j'}(t-1)}{\sigma_{t-1}} \left[\int_{-\bar{m}}^{\infty} P(v) dv + \int_{-\infty}^{\bar{m}} P(v) dv \right] \right\} \\
&= \frac{P(\bar{m})}{N} \left[\frac{2}{\alpha} P(\bar{m}) + \frac{2S_{j'}(t-1)}{\sigma_{t-1}} \int_{-\bar{m}}^{\bar{m}} P(v) dv \right] \\
&= \frac{P(\bar{m})}{N} \left[\frac{2}{\alpha} P(\bar{m}) + \frac{2S_{j'}(t-1)}{\sigma_{t-1}} \operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) \right] \tag{193}
\end{aligned}$$

と計算される.

全く同様の手続きにより, $F_2/2$ も計算することができて

$$\frac{F_2}{2} = \frac{P(\bar{m})}{N} \left[\frac{2}{\alpha} P(\bar{m}) + \frac{2S_j(t-1)}{\sigma_{t-1}} \operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) \right] \tag{194}$$

となるので, $Y_{11} + Y_{1-1}$ は

$$\begin{aligned}
Y_{11} + Y_{1-1} &= F_1 + F_2 \\
&= \frac{4}{N} \left[\frac{2}{\alpha} P(\bar{m})^2 + \frac{1}{\sigma_{t-1}} (S_j(t-1) + S_{j'}(t-1)) P(\bar{m}) \operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) \right] \\
&= \frac{4}{N} \left[\frac{2}{\alpha} P(\bar{m})^2 + \frac{2m_{t-1}}{\sigma_{t-1}} P(\bar{m}) \operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) \right] \tag{195}
\end{aligned}$$

となる. 上式の最後の式変形では

$$E[S_j(t-1)] = E[S_{j'}(t-1)] = m_{t-1} \tag{196}$$

を用いた.

上の $Y_{11} + Y_{1-1}$ を求めた方法は, そのまま $Y_{-11} + Y_{-1-1}$ の計算にも適用できて, ほとんど同じ手続きを
 ぶむことにより

$$Y_{-11} + Y_{-1-1} = Y_{11} + Y_{1-1} \tag{197}$$

を得る. 従って最終的に v は

$$\begin{aligned}
v &= \frac{1}{4} (Y_{11} + Y_{1-1} + Y_{-11} + Y_{-1-1}) \\
&= \frac{1}{N} \left[\frac{4}{\alpha} P(\bar{m})^2 + \frac{4m_{t-1}}{\sigma_{t-1}} P(\bar{m}) \operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) \right] \\
&= \frac{1}{N} \left[\frac{4}{\alpha} P(\bar{m})^2 + \frac{4m_{t-1}}{\sigma_{t-1}} P(\bar{m}) m_t \right] \tag{198}
\end{aligned}$$

となる. ここで, 最後の変形では (147) より

$$\operatorname{erf} \left(\frac{\bar{m}}{\sqrt{2}} \right) = \operatorname{erf} \left(\frac{m_{t-1}}{\sqrt{2}\sigma_{t-1}} \right) = m_t \tag{199}$$

であることを用いている.

(B) v' の評価

次に v' の評価に移るが、これは今の仮定のもとでは

$$v' = E[\xi_i^\mu \xi_j^\mu S_j(t) \xi_i^{\mu'} \xi_j^{\mu'} S_j'(t)] = E[\xi_i^\mu \xi_j^\mu S_j(t)] E[\xi_i^{\mu'} \xi_j^{\mu'} S_j'(t)] \quad (200)$$

となり、また

$$E[N_i^t] = E\left[\frac{1}{N} \sum_{j \neq i} \sum_{\mu=2}^p \xi_i^\mu \xi_j^\mu S_j(t)\right] = \frac{1}{N} (N-1)(p-1) E[\xi_i^\mu \xi_j^\mu S_j(t)] = p E[\xi_i^\mu \xi_j^\mu S_j(t)] \quad (201)$$

であるから

$$\{E[N_i^t]\}^2 = p^2 v' = N \alpha^2 v' \quad (202)$$

が成り立つ。よって、最終的に (153) より

$$\begin{aligned} \sigma_t^2 &= E[(N_i^t)^2] - \{E[N_i^t]\}^2 \\ &= \alpha + \frac{1}{N^2} \times p N^2 \frac{1}{N} \left[\frac{4}{\alpha} P(\bar{m})^2 + 4 \frac{m_{t-1}}{\sigma_{t-1}} P(\bar{m}) m_t \right] \\ &\quad + \frac{1}{N^2} \times p^2 N^2 v' - N^2 \alpha^2 v' \\ &= \alpha + 4P \left(\frac{m_{t-1}}{\sigma_{t-1}} \right)^2 + 4\alpha \left(\frac{m_{t-1} m_t}{\sigma_{t-1}} \right) P \left(\frac{m_{t-1}}{\sigma_{t-1}} \right) \end{aligned} \quad (203)$$

が得られる。

以上まとめると、 m_t, σ_t の時間発展は

$$m_t = \operatorname{erf} \left(\frac{m_{t-1}}{\sqrt{2} \sigma_{t-1}} \right) \quad (204)$$

$$\sigma_t^2 = \alpha + 4P \left(\frac{m_{t-1}}{\sigma_{t-1}} \right)^2 + 4\alpha \left(\frac{m_{t-1} m_t}{\sigma_{t-1}} \right) P \left(\frac{m_{t-1}}{\sigma_{t-1}} \right) \quad (205)$$

となる。ただし、

$$P(x) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (206)$$

であり、誤差関数を

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz \quad (207)$$

で定義している。

1.13 学習の結果としての Hebb 則 : もう一つの時間スケール

我々はこの節を通じて神経素子間の結合が Hebb 則で与えられ, 具体的には

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} \quad (208)$$

で任意の素子間の結合を決めたのであった. しかし, 「神経回路網はどのようにして上記のような結合を獲得したのであるか?」という疑問が残る.

実は, この Hebb 則は「記憶」と並ぶ脳の重要な高次機能である学習により得られたのであった. この「学習」においても, 回路網は「動的」にその結合を変化させる. μ でその更新ステップを表すとすれば, 脳は任意の素子間の結合を

$$w_{ij}(\mu+1) = w_{ij}(\mu) + \frac{1}{N} \xi_i^{\mu} \xi_j^{\mu} \quad (209)$$

の規則によって更新 (学習) させる. すると (208) 式はこの更新式の p ステップ後の結果であるとして理解できる.

ところで, 脳神経系では各神経素子 S の動きが始まる前にその学習 (209) を終えている必要がある. つまり, [学習フェーズ] と [想起フェーズ] は別個のものであり, 同じ時間スケール内では起こらないと考える. 我々が物事を記憶するというプロセスとそれを呼び起こすというプロセスを同時に行うことは実際の日常生活を考えても無いであろう. 物事を記憶から呼び起こすためには, それに先んじて記憶するというプロセスが無ければならない. 従って, 計算機シミュレーションをする際にも予め結合 w_{ij} を与えて神経素子 S の動きを計算機上でシミュレートしたわけである (w_{ij} が (208) に確定した後の w_{ij} の緩和時間は無限大である. 従って, 回路網の状態ベクトル S は「速い変数」, 結合ベクトルは「遅い変数 (事実上, 時間変化無し)」とみなすことができる). この学習については後に [学習の数理] で詳しくみることになる. ここではこれだけに留めておこう.

問 7: 神経素子の状態更新 ($t \geq 1$):

$$S_i(t+1) = \text{sgn} \left(\sum_{j \neq i} w_{ij} S_j(t) \right)$$

と結合の学習:

$$w_{ij}(t+1) = w_{ij}(t) + \frac{1}{N} \sum_{\nu=1}^q \xi_i^{(t+1)\nu} \xi_j^{(t+1)\nu}$$

$$w_{ij}(0) = 0$$

(ξ^1 をターゲットパターンとする) を同時に (同じ時間スケール内で) 行った場合の想起過程を計算機上でシミュレートせよ. このとき, 既に行った w_{ij} を初めに固定した場合のような連想記憶が実現できるであろうか?

[考察へのヒント]

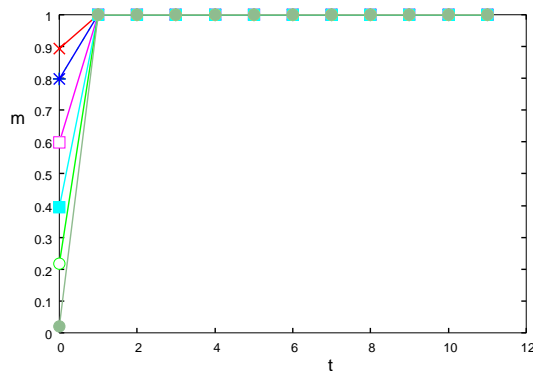
上記の w_{ij} の更新が $t = T$ ステップ経過した後の結合は直ちに

$$w_{ij}(T) = \frac{1}{N} \sum_{t=1}^T \sum_{\nu=1}^p \xi_i^{t\nu} \xi_j^{t\nu} = \frac{1}{N} \sum_{\mu=1}^{Tq} \xi_i^\mu \xi_j^\mu$$

となる. ここで q の選び方により, 次のように記憶のスピードが異なってくる.

- $q = 1$ の場合には各ステップで 1 個ずつのパターンを記憶して行く.
- $q = \alpha N$ の場合には各ステップで $\mathcal{O}(N)$ 個ずつのパターンを記憶して行く.

$q = 1$ の場合にはターゲットパターンを初めに記憶し, それ以降はステップごとに 1 個ずつのパターンを記憶に追加して行く.



前に見た $\mathcal{O}(1)$ のパターンの埋め込みの結果からわかるように, 1 ステップ後には既にターゲットパターンのアトラクタに落ち込んでいる (ターゲットパターンを想起している) はずであるから (図参照), その後に追加して行くパターンがターゲットパターンを不安定化しない限り問題は生じないことが予想される (特に, $\mathcal{O}(N)$ ステップ経過後の安定性を調べて見ると良い). 一方, $q = \alpha N$ の場合には, α の大きさに依存し, これとターゲットパターンへ落ち込むステップ数との兼ね合いから想起の可能性が決まることとが予想される.

1.14 対称結合回路網とそのエネルギー関数

この [連想記憶の数理] の節の最後では後に見る形式的統計力学の処方箋への導入のために、神経細胞のノイズゼロでの非同期態更新式：

$$S_i(t+1) = \operatorname{sgn} \left(\sum_{j \neq i} w_{ij} S_j(t) \right) \quad (210)$$

(ここで、 w_{ij} は任意の素子 i, j 間の結合。以下の議論ではこの結合が特に Hebb 則に従うとは仮定しない) が次で定義される神経系のエネルギー関数：

$$E = - \sum_{i,j} w_{ij} S_i S_j \quad (211)$$

を単調に減少させることを見ておくことにする。

まず、非同期状態更新であるから、任意の時刻に状態更新する神経素子を S_1 と選ぼう。そして、エネルギー関数を次のように書き換えておくことにする。

$$E = -S_1 \sum_{k=2}^N (w_{1k} + w_{k1}) S_k - \sum_{i,j \neq 1} w_{ij} S_j \quad (212)$$

すると、素子 S_1 は任意の時刻で

$$S_1 = \operatorname{sgn} \left(\sum_{k \neq 1} w_{k1} S_k \right) \quad (213)$$

に従ってその状態を更新するわけであるから、(211) 式で定義されるエネルギー関数は任意の隣接する時刻間 $t+1$ と t で

$$\begin{aligned} \Delta E &= -\operatorname{sgn} \left(\sum_{k=2}^N w_{1k} S_k \right) \sum_{k=2}^N (w_{1k} + w_{k1}) S_k - \sum_{i,j \neq 1} w_{ij} S_j \\ &\quad - \left\{ -S_1 \sum_{k=2}^N (w_{1k} + w_{k1}) S_k - \sum_{i,j \neq 1} w_{ij} S_i S_j \right\} \\ &= S_1 \sum_{k=2}^N (w_{1k} + w_{k1}) S_k - \left\{ \sum_{k=2}^N (w_{1k} + w_{k1}) S_k \right\} \operatorname{sgn} \left(\sum_{k=2}^N w_{1k} S_k \right) \end{aligned} \quad (214)$$

だけの差を持つことになる。ところで、神経素子間の全ての結合 w_{ij} が対称：

$$w_{ij} = w_{ji} \quad (215)$$

であるのであれば、エネルギー関数の差 ΔE は次のように変形することができる。

$$\begin{aligned} \Delta E &= 2S_1 \sum_{k=2}^N w_{1k} S_k - 2 \left(\sum_{k=2}^N w_{1k} S_k \right) \operatorname{sgn} \left(\sum_{k=2}^N w_{1k} S_k \right) \\ &= 2S_1 \sum_{k=2}^N w_{1k} S_k - 2 \left| \sum_{k=2}^N w_{1k} S_k \right| \leq 0 \end{aligned} \quad (216)$$

従って、ノイズゼロでの非同期状態更新式 (210) はエネルギー関数 (211) を単調に減少させることがわかった。このことから、我々が今まで調べてきた対称 Hebb 則：

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} \quad (217)$$

ではエネルギー関数 (リャプノフ関数)：

$$E = -\frac{1}{N} \sum_{ij} \sum_{\mu=1}^p \xi_i^{\mu} \xi_j^{\mu} S_i S_j \quad (218)$$

が存在する。「平均場アニーリング」の導入の際に述べた神経系のエネルギー関数とはこのことであったわけである。一方、回路網の想起の過程がリミットサイクル解を持つような非対称 Hebb 則：

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^{\mu+1} \xi_j^{\mu} \quad (219)$$

の場合には上のエネルギーに関する単調減少性の証明からわかるように、この関数が単調に減少するということは起こらない。

エネルギー関数が存在すれば、形式的な統計力学の処方箋に従って、この節 [連想記憶の数理] で得られた対称結合を持つ回路網に関する全ての結果を再び導出することができる。実際、この節 [連想記憶の数理] では部分的に統計力学/熱力学の考え方を使っているとはいうものの、本格的に統計力学を用いたわけではなかった。次の節 [最適化問題の数理] ではまず最初に統計力学の方法を簡単に説明し、個別の問題にあたって行くことにする。そして、統計力学の考え方にある程度慣れてから、もう一度このトピックス [連想記憶の数理] に立ち返り、問題を再考することにしよう。