

## 基本パーセプトロンの等比学習とその収束条件

著者	三好 誠司, 池田 和司, 中山 謙二
雑誌名	電子情報通信学会論文誌. A, 基礎・境界
巻	J81-A
号	5
ページ	844-853
発行年	1998-05-25
URL	<a href="http://hdl.handle.net/2297/5629">http://hdl.handle.net/2297/5629</a>

基本パーセプトロンの等比学習とその収束条件

三好 誠司<sup>†</sup>      池田 和司<sup>††</sup>      中山 謙二<sup>†††</sup>

Geometric Learning Algorithm for Elementary Perceptron and Its Convergence Condition

Seiji MIYOSHI<sup>†</sup>, Kazushi IKEDA<sup>††</sup>, and Kenji NAKAYAMA<sup>†††</sup>

あらまし 基本パーセプトロンの学習アルゴリズムとして等比学習アルゴリズム (GLA) を提案する。GLA は適応フィルタにおけるアフィン射影アルゴリズム (APA) を基本パーセプトロンに適用したものである。APA における係数更新の目標が“ある 1 点”であるのに対して、GLA における結合荷重更新の目標は全パターンを正しく分類する“ある領域”である。このため両者の収束条件は異なる。本論文では、まず、2 個のパターンに対して 1 次 GLA が収束するために 2 パターンの角度  $\theta$  と学習係数  $\lambda$  が満たすべき関係を理論的に導出する。次に、パターンが 3 個以上の場合に“解領域の角度  $\psi_{\min}$ ”を導入し、1 次 GLA が収束するための  $\psi_{\min} - \lambda$  の関係が、2 パターンの場合の  $\theta - \lambda$  の関係で近似できることを計算機シミュレーションで示す。更に、 $\theta - \lambda$  の関係において  $\theta$  によらずに収束する条件である  $\lambda = 2$  に対して、1 次 GLA がパターン数にかかわらずに常に収束することを証明する。これらにより、1 次 GLA の収束性が保証され、学習アルゴリズムとしての有効性が確認される。また、1 次 GLA が収束するための  $\psi_{\min}$  と  $\lambda$  が分布する範囲が近似的に明らかになる。

キーワード パーセプトロン、パターン分類、ニューラルネットワーク、アフィン射影アルゴリズム、等比学習アルゴリズム

1. ま え が き

基本パーセプトロンによるパターン分類のための学習アルゴリズムとしては、パーセプトロン学習がよく知られている。パーセプトロン学習の最大の長所は、与えられた問題が線形分離可能であるならば、有限回の更新で結合荷重が必ず解領域に到達することである。この性質はパーセプトロンの学習収束定理としてよく知られている [1]。しかしながらパーセプトロン学習にはいくつかの短所もある。すなわち、学習速度が遅いことや、見つかった解が雑音を含むパターンの分類に対して必ずしも良好な特性を示すとは限らないことなどである。双方向連想記憶のために提案された PRLAB [2] は 1 個のユニットの 1 方向の学習に着目

すれば基本パーセプトロンの学習アルゴリズムと等価であり、これはパーセプトロン学習の学習速度を改善したものであるとすることができる。

一方、適応フィルタの分野においては、正規化 LMS アルゴリズムをブロック信号処理に拡張したアルゴリズムとしてアフィン射影アルゴリズム (Affine Projection Algorithm, APA) [3] がよく知られている。服部らは 2 次 APA を双方向連想記憶の学習に適用し、パーセプトロン学習よりも高速であることを計算機シミュレーションにより確認した [4]。しかしながら、文献 [4] は双方向連想記憶というやや特殊なモデルに限定された内容であり、また、学習収束条件については詳細な検討がなされていない。

本論文では、ニューラルネットワーク研究の原点であり、また現在でも基本的な要素として重要である基本パーセプトロンを対象とし、この基本パーセプトロンへの APA の適用として、等比学習アルゴリズム (Geometric Learning Algorithm, GLA) を提案する [5], [6]。次に、学習の発振状態が生じうる条件についての解析を行うことにより、1 次 GLA で 2 個のパターンを分類する場合に有限回の更新で結合荷重重

<sup>†</sup> 神戸高専電子工学科, 神戸市  
Department of Electronic Engineering, Kobe City College of Technology, Kobe-shi, 651-2194 Japan

<sup>††</sup> 京都大学大学院情報学研究所, 京都市  
Graduate School of Informatics, Kyoto University, Kyoto-shi, 606-8501 Japan

<sup>†††</sup> 金沢大学工学部電気・情報工学科, 金沢市  
Faculty of Engineering, Kanazawa University, Kanazawa-shi, 920-8667 Japan

クトルが解領域に到達するために、2パターンの角度  $\theta$  と学習係数  $\lambda$  が満たすべき関係を理論的に導出する。次に、パターンが3個以上の場合に“解領域の角度  $\psi_{\min}$ ”を導入し、1次GLAが収束するための  $\psi_{\min} - \lambda$  の関係および  $\theta - \lambda$  との関係を計算機シミュレーションで調べる。更に、 $\theta - \lambda$  の関係において  $\theta$  によらずに収束する条件である  $\lambda = 2$  に対して、1次GLAがパターン数にかかわらずに常に収束することを証明する。これらをふまえてGLAの有効性と解領域の角度の意義について述べる。

なお、本論文においては、結合荷重ベクトルが解領域に到達して更新が終了することを、学習が収束する、と表現するものとする。更に、結合荷重ベクトル初期値によらずに学習が有限回の更新で収束することを、学習が大域収束する、と表現することにする。

## 2. 基本パーセプトロン

Rosenblatt によって提案された単純パーセプトロン[7]において、ユニットの出力関数がしきい値関数である場合を基本パーセプトロンとよぶ。その構造を図1に、入出力関係を式(1)、(2)に示す。

$$u = \sum_{i=0}^{N-1} w_i x_i \quad (1)$$

$$y = \begin{cases} +1, & u \geq 0 \\ -1, & u < 0 \end{cases} \quad (2)$$

式(1)の  $u$  に0を代入すると、傾きが  $w_i$  で決定され原点を通る超平面を表す式になる。よって、基本パーセプトロンはパターン空間において超平面を境界とする二つのクラスを分類する能力を有すると言える。

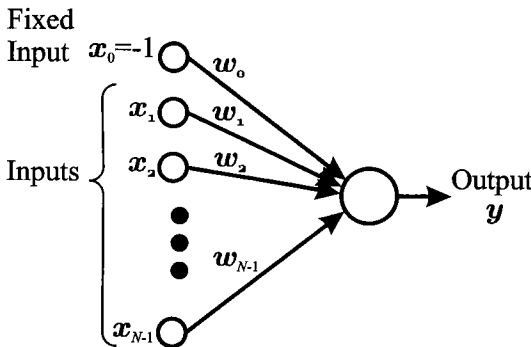


図1 基本パーセプトロン  
Fig.1 Elementary perceptron.

なお図1は、値として常に  $-1$  をとり、対応する結合荷重が  $w_0$  であるような固定入力  $x_0$  を追加することにより、しきい値を0とした基本パーセプトロンである。この基本パーセプトロンを、固定入力  $x_0$  をもたずにしきい値が0に限定されない場合と比較すると、入力数が1だけ増えるかわりに境界超平面が原点を通ることになり都合がよい。よって本論文においては図1の基本パーセプトロンを扱うことにする。また、この固定入力を含む入力の総数  $N$  をパターンの次元とよぶことにする。

## 3. 等比学習アルゴリズム

### 3.1 アフィン射影アルゴリズム

アフィン射影アルゴリズム (Affine Projection Algorithm, APA) は適応フィルタのアルゴリズムとして提案された[3]。図2はAPAによるフィルタ係数更新を示している。図2において、 $h^*$  は適応の目標、すなわち未知システムに対する最適な同定である。また二つの超平面は、それぞれ異なる時間ステップでの二つのタップ入力ベクトル  $u^1, u^2$  を法線とし、それぞれのタップ入力ベクトルに対して望みの値を出力するようなフィルタ係数の集合である。フィルタ係数ベクトルは超平面の交わりに向かって垂直に更新される。k次APAにおいては、1回の更新にk個の超平面を用いる。すなわち図2は、次数kが2でタップ入力ベクトル  $u^1, u^2$  が3次元である場合を表していることになる。この場合、超平面は2次元になるので、その交わりは1次元の直線  $l$  である。APAにおいては  $PR/PQ$  を一定として各更新を行う。1次APA

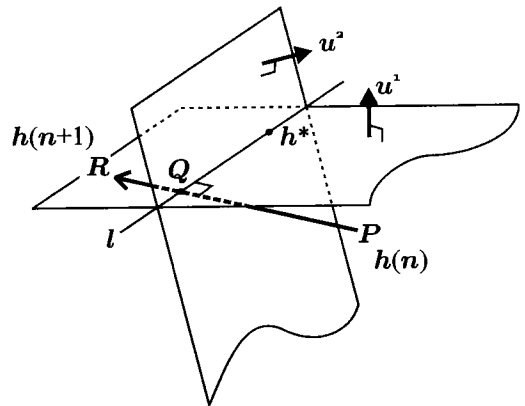


図2 APAにおけるフィルタ係数更新  
Fig.2 Update of filter coefficients in APA.

は NLMS アルゴリズム [8],[9] と等価であるので [3], APA は NLMS アルゴリズムをブロック適応アルゴリズムに拡張したアルゴリズムであると言える。  $\overline{PR}/\overline{PQ}$  を  $\mu$  とおくと, APA が収束するための必要十分条件は以下のようなものである [3].

$$0 < \mu < 2 \tag{3}$$

### 3.2 等比学習アルゴリズム

前節で述べたように, 図 1 に示す基本パーセプトロンの分離超平面は原点を通る. よってその交わりも原点を通る. すなわち, この交わりは入力ベクトル空間の“部分空間”であり, “アフィン部分空間”とよぶ必要はない.

この理由により, パターン分類を行うための基本パーセプトロンの学習に APA を適用するアルゴリズムを新たに“等比学習アルゴリズム (Geometric Learning Algorithm, GLA)”とよぶことにする. この名称は, 更新ベクトルの長さ ( $\overline{PR}$ ) と, 結合荷重ベクトルから交わりに下ろした垂線の長さ ( $\overline{PQ}$ ) の比が一定であることに由来している. そしてこの比  $\overline{PR}/\overline{PQ}$  を学習係数とよび,  $\lambda$  で表すことにする. また,  $k$  次 APA に対応するアルゴリズムを  $k$  次の等比学習アルゴリズム ( $k$  次 GLA) とよぶ. すなわち,  $k$  次 GLA においては, 各時点での結合荷重ベクトルの更新のために, その時点で式 (1), (2) とクラスが矛盾するパターンの中の  $k$  個を用いる.  $k$  次 GLA を以下に示す.

- [step1]  $w(0)$  を乱数で初期設定する.
- [step2] 式 (1), (2) とクラスが矛盾するパターンの個数を  $k_0$  とする.
- [step3]  $k_0 = 0$  なら収束と判定し終了する.
- [step4]
  - もし  $k_0 \geq k$  ならば
  - $$X = (x^1, x^2, \dots, x^k)^T \tag{4}$$
  - そうでないなら
  - $$X = (x^1, x^2, \dots, x^{k_0})^T \tag{5}$$
- [step5] 結合荷重を次式で更新する.
- $$w(i+1) = w(i) - \lambda X^+ X w(i) \tag{6}$$
- [step6] step2 に戻る.

ここで  $X^+$  は  $X$  の Moore-Penrose 一般逆行列である. また,  $T$  は行列の転置を表す. 式 (4), (5) におい

て  $x^1 \sim x^k$  または  $x^1 \sim x^{k_0}$  は式 (1), (2) とクラスが矛盾するパターンから選択する. この選択をどのような規則で行うかについてはここでは規定しない.

APA は適応フィルタのアルゴリズムであるから, フィルタ係数更新の目標は未知システムに対する最適な同定であり, これは“ある 1 点”である. これに対して, GLA は基本パーセプトロンでパターン分類を行う際の学習アルゴリズムであるから, 結合荷重更新の目標は与えられたすべてのパターンを正しく分類するような“ある領域”である. それゆえ, GLA の収束条件は式 (3) に示された APA の収束条件とは異なる. GLA の収束条件の解析は本論文の主題であり次章以降で詳しく述べる.

図 3 はパーセプトロン学習を, 図 4 は 1 次 GLA を示している. これらの図は 2 次元 2 パターンの場合の学習の進み方を表している. 図 4 内の円は結合荷重ベクトルのノルムが毎回の更新において正規化されることを表している. この正規化を行うと, 結合荷重ベク

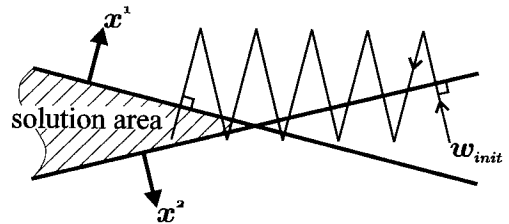


図 3 パーセプトロン学習における結合荷重の更新  
Fig.3 Weights update process in Perceptron learning.

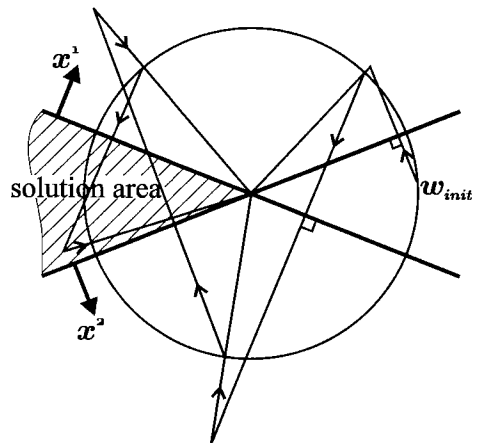


図 4 1 次 GLA における結合荷重の更新  
Fig.4 Weights update process in 1-GLA.

トルのノルムの変化がなくなるので好都合である。また、しきい値が0である基本パーセプトロンを考慮しており、式(6)が  $w$  について線形であるので、この正規化が収束条件に影響を与えることはない。よって本論文ではこの正規化を行う場合について解析を行うことにする。

#### 4. 2パターンに対する1次GLAの理論収束条件

ここでは2個のパターンに対して1次GLAが大域収束するための条件について理論的な解析を行う。1次GLAは1次APAの基本パーセプトロンへの適用であり、その特性解析は  $k$  次GLAの特性解析の出発点でもある。

前にも述べたように、与えられた問題が線形分離可能であればパーセプトロン学習は有限回の結合荷重更新で収束する[1]。これに対して、GLAは常に収束するというわけではない。図5は1次GLAにおける発振状態 ( $\rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow$ ) を示している。このような状態に陥ると結合荷重ベクトルは解領域にこれ以上近づいていかないので学習は収束しない。本章では、もっとも基本的な場合として、1次GLAで2個のパターンの分類を学習する際に発振状態が生じうる条件についての解析を行い、学習が大域収束するための必要十分条件を導出する。ここで、図5が、 $N (\geq 2)$  次元空間において原点と二つのパターンベクトルを含むような2次元の平面を表していると考え

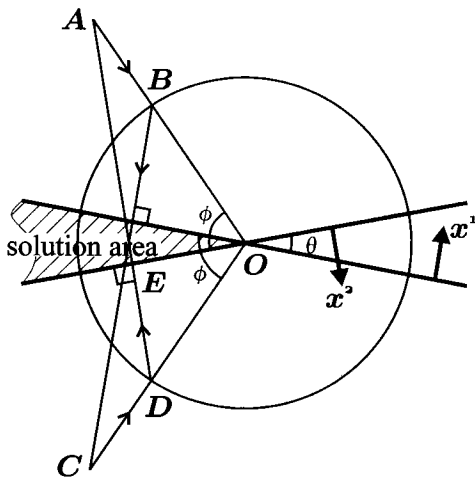


図5 1次GLAにおける発振状態  
Fig.5 Oscillation phenomenon in 1-GLA.

ことにより、本章の議論は  $N (\geq 2)$  次元に関して成り立つ。また、 $\lambda \leq 1$  の場合には有限回の更新で超平面を越えることができないので、解領域に到達することはない。よって以下の解析においては  $\lambda > 1$  の場合について考えることにする。

いま、図5のような発振状態にあるとする。なお、図5においては、 $\overline{DA/DE}$  が  $\lambda$  である。二つの直角三角形  $\triangle OED$ ,  $\triangle OEA$  は辺  $OE$  を共有するので、以下の式が成り立つ。

$$(\lambda - 1) \sin(\phi - \theta) \cot \phi = \sin(\phi - \theta) \cot(\phi - \theta) \tag{7}$$

この方程式を  $\lambda$  に関して解くと次式を得る。

$$\lambda = \frac{\tan \phi}{\tan(\phi - \theta)} + 1 \tag{8}$$

図5から明らかのように、 $\theta$  は分類しようとする二つのパターンおよびそれぞれが属するクラスによって一意に決定される。式(8)を満足する  $\lambda, \phi (\geq \theta)$  が存在する場合には、発振状態が生じうる可能性があることになる。

ここで式(8)の右辺を  $\phi$  の関数  $f$  とおく。すなわち、

$$f(\phi) = \frac{\tan \phi}{\tan(\phi - \theta)} + 1 \tag{9}$$

図6、図7はそれぞれ  $0 < \theta \leq \frac{\pi}{2}$ ,  $\frac{\pi}{2} < \theta \leq \pi$  の場合の関数  $f(\phi)$  の概形を示している。

- $0 < \theta \leq \frac{\pi}{2}$  の場合 図6より、 $\lambda \geq \frac{\tan(\frac{\pi}{4} + \frac{\theta}{2})}{\tan(\frac{\pi}{4} - \frac{\theta}{2})} + 1$   
 または  $\lambda \leq \frac{\tan(\frac{\pi}{4} - \frac{\theta}{2})}{\tan(\frac{\pi}{4} + \frac{\theta}{2})} + 1$  の場合に、 $f(\phi) = \lambda$  なる  $\phi (\geq \theta)$  が二つ (等号の場合には一つ) 存在するこ

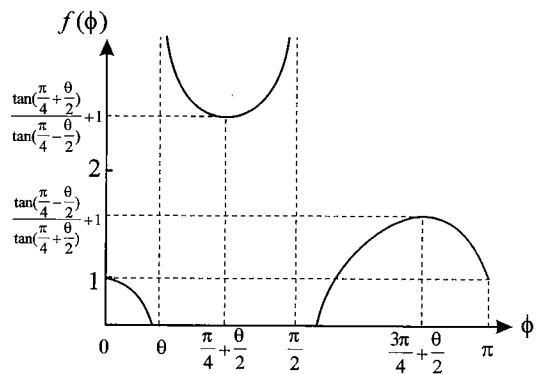


図6  $\phi$  の関数  $f (0 < \theta \leq \frac{\pi}{2})$   
Fig.6 Function  $f$  of  $\phi (0 < \theta \leq \frac{\pi}{2})$ .

とがわかる。これは二つ（等号の場合には一つ）の発振状態が存在することを意味している。付録 1. に示すように、この二つの発振状態のうち一つは不安定であるが、もう一つは安定である（等号の場合の一つの発振状態は片側からのみ安定である）。それゆえ、 $\lambda \geq \frac{\tan(\frac{\pi+\theta}{4})}{\tan(\frac{\pi-\theta}{4})} + 1$  あるいは  $\lambda \leq \frac{\tan(\frac{\pi-\theta}{4})}{\tan(\frac{\pi+\theta}{4})} + 1$  が成り立つ場合には、学習が解領域に到達する前に、発振状態に陥る可能性がある。 $\lambda \geq \frac{\tan(\frac{\pi+\theta}{4})}{\tan(\frac{\pi-\theta}{4})} + 1$  あるいは  $\lambda \leq \frac{\tan(\frac{\pi-\theta}{4})}{\tan(\frac{\pi+\theta}{4})} + 1$  が成り立つ場合に学習が発振状態に陥るか、あるいは解領域に到達するかは結合荷重ベクトルの初期値に依存する。

図 6 より、 $\frac{\tan(\frac{\pi+\theta}{4})}{\tan(\frac{\pi-\theta}{4})} + 1 > \lambda > \frac{\tan(\frac{\pi-\theta}{4})}{\tan(\frac{\pi+\theta}{4})} + 1$  の場合には、 $f(\phi) = \lambda (> 1)$  を満足する  $\phi (\geq \theta)$  は存在しないことがわかる。付録 2. に示すように、このような場合には学習は有限回の更新で必ず収束する。

- $\frac{\pi}{2} < \theta \leq \pi$  の場合 図 7 より、 $f(\phi) = \lambda$  なる  $\phi (\geq \theta)$  は存在しないことがわかる。そしてこの場合には、付録 2. に示すように、学習は有限回の更新で必ず収束する。

上記の考察より、2個のパターンに対して1次 GLA が大域収束するための必要十分条件は以下の通りである。

- $0 < \theta \leq \frac{\pi}{2}$  の場合

$$\frac{\tan(\frac{\pi+\theta}{4})}{\tan(\frac{\pi-\theta}{4})} + 1 > \lambda > \frac{\tan(\frac{\pi-\theta}{4})}{\tan(\frac{\pi+\theta}{4})} + 1 \quad (10)$$

- $\frac{\pi}{2} < \theta \leq \pi$  の場合

$$\lambda > 1 \quad (11)$$

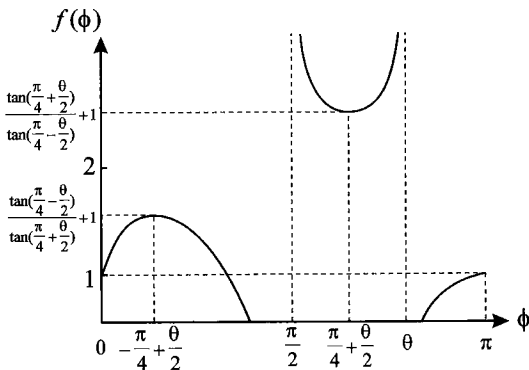


図 7  $\phi$  の関数  $f$  ( $\frac{\pi}{2} < \theta \leq \pi$ )  
Fig.7 Function  $f$  of  $\phi$  ( $\frac{\pi}{2} < \theta \leq \pi$ ).

## 5. 多パターンに対する 1 次 GLA の収束特性

### 5.1 解領域の角度

前章において、パターンが 2 個の場合の 1 次 GLA の学習収束性は、図 5 における  $\theta$  によって決定されることが明らかになった。すなわち式 (10), (11) より、 $\theta$  が大きいほど大域収束する  $\lambda$  の範囲が広い。

パターンが 3 個以上の場合（以後、パターンが多い場合、と表現する）を 2 個の場合と比較すると、パターン数の増加のため解領域の形状が複雑になること、および、パターン提示の順序という自由度が加わることにより、収束あるいは発振状態に至るまでの学習の過程は一般には複雑になる。

しかし、パターンが多い場合でも、そのパターン集合に固有な何らかの角度が式 (10) の  $\theta$  に相当し学習収束性を決定しているのではないかと考えられる。ここではパターンが与えられると比較的容易に一意に求められ、しかもパターンが 2 個の場合の  $\theta$  を包含する概念として、次に述べる“解領域の角度”を導入することにする。

図 8 は解領域が 5 個のパターンにより決定されている場合を表す。この図においては、各パターンを法線とし原点を通る超平面と超球面の交わりが描かれているので、斜線で示した部分が解領域を表していることになる。図 8 において、 $S, T$  は解領域の周囲を動く点であるとする。いま、 $S$  を固定して  $T$  を一周させたときの  $\angle SOT$  の最大値を  $S$  の関数  $\psi(S)$  とする。そして、 $\psi(S)$  の  $S$  に関する最小値を、解領域の角度  $\psi_{\min}$  と定義する。 $\psi_{\min}$  は原点から解領域を見た角度

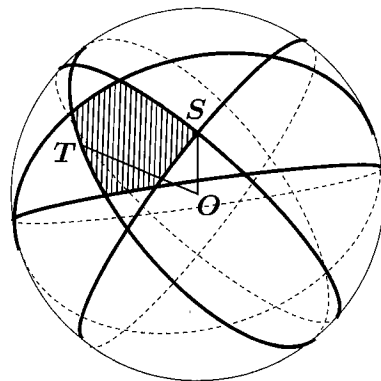


図 8 解領域の角度  
Fig.8 Angle of solution area.

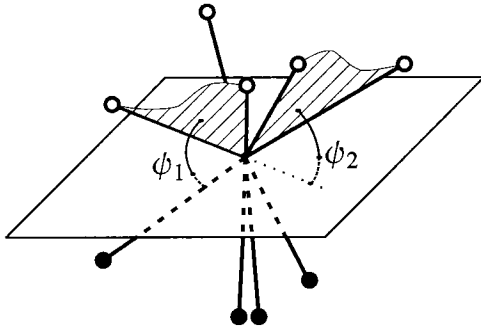


図9 パターン集合と分割超平面の関係  
Fig.9 Relation between patterns and separating hyperplane.

の最小値であると言える。

与えられたパターン集合について解領域の角度  $\psi_{\min}$  を求めるための考え方は以下の通りである。結合荷重ベクトルが解領域内に存在する場合には、そしてその場合に限り、結合荷重ベクトルを法線とし原点を通る超平面がパターンを正しく分割する。図9は多次元空間におけるパターン集合と分割超平面の関係を、3次元パターンを例に図示したものである。この図において、白丸で表されたパターンと黒丸で表されたパターンは異なるクラスに属する。すなわち結合荷重ベクトルが解領域内に存在する場合には、そしてその場合に限り、超平面はこれら二つのクラスを正しく分割している。 $\psi_1$  は二つの白丸パターンではさまれたくさび形領域と、一つの黒丸パターンのなす角度を表している。 $\psi_2$  は二つの白丸パターンではさまれたくさび形領域と、一つの白丸パターンの方向を180度変えたパターンのなす角度を表している。 $\psi_{\min}$  とは、原点から解領域を見た角度の最小値であるから、すべてのパターンの組合せについての  $\psi_1$  と  $\psi_2$  の最小値として求めることができる。ここで、すべてのパターンの組合せには、黒と白を入れ替えた場合も含む。また、 $N$  次元の場合のくさび形領域としては、最大  $N - 1$  個のパターンベクトルの非負係数線形和を考える必要がある。

### 5.2 計算機シミュレーション

1次GLAが大域収束するための“解領域の角度  $\psi_{\min}$ ”と学習係数  $\lambda$  の関係を計算機シミュレーションにより調べた。

まず、ランダムにクラス分けされた5個の7次元パターンからなるパターン集合を10組作成し、各パターン集合の  $\psi_{\min}$  を前節で述べた考え方により求めた。

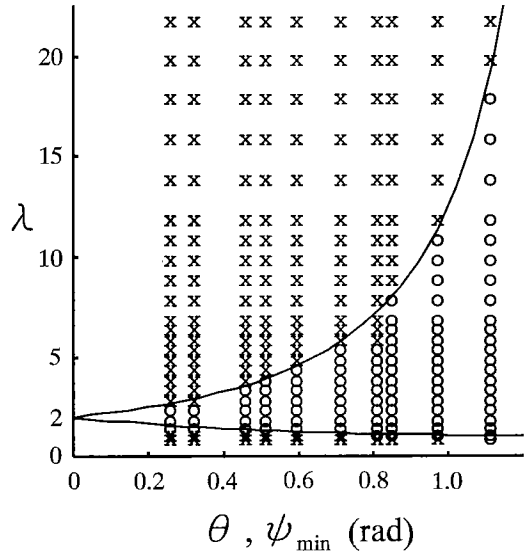


図10 1次GLAの収束条件  
Fig.10 Condition for 1-GLA convergence.

ここで、すべてのパターンは、その第一成分は  $-1.0$ 、それ以外の6個の成分は  $-2.0 \sim +2.0$  のランダムな実数である。次にこれらのパターン集合を用いて  $\lambda$  の値を変えて学習収束特性を調べた。1次GLAの理論収束条件である式(10)は学習が大域収束するための必要十分条件であるので、この式との比較のためには多数の結合荷重初期値に関して収束性を調べる必要がある。ここでは100個のランダム初期値を用いた。また、パターンの提示は、パターン数5を1周期とする周期的提示とランダム提示の両方で行った。すなわち、それぞれの結合荷重ベクトル初期値において  $5! = 120$  通りの周期的提示と20通りのランダム提示で学習を行った。このように、あるパターン集合に対してある  $\lambda$  を用いたとき学習収束性は14000回の試行により判定した。また、各試行では、更新を10000回行っても解領域内に到達しない場合を“収束しない”と判定した。

図10は理論収束条件と、10個のパターン集合についてのシミュレーション結果を示している。2本の曲線は式(10)で与えられる理論収束条件の上限と下限である。また、縦方向に並ぶ“o”と“x”の列10個のそれぞれは、一つのパターン集合に関してのいろいろな  $\lambda$  の値における学習収束性を表している。ここで、“o”は14000回の試行のすべてで学習が収束したことを表しており、“x”は14000回の試行において、学習が収束しない試行が1回以上あったことを表している。

図 10 より, 多パターンに対して 1 次 GLA が大域収束するための  $\psi_{\min} - \lambda$  の関係と, 2 パターンに対して 1 次 GLA が大域収束するための  $\theta - \lambda$  の関係がきわめてよく一致していることがわかる. すなわち, 解領域の角度  $\psi_{\min}$  を導入することによって, 多パターンの場合に 1 次 GLA が大域収束するための条件が, 2 パターンの場合の理論収束条件で近似できる. このことから, 解領域の角度  $\psi_{\min}$  が多パターンに対する 1 次 GLA の大域収束条件を決める重要なパラメータであると言える.

### 6. 1 次 GLA の収束定理

4. において, パターンが 2 個の場合に 1 次 GLA が大域収束するための  $\theta$  と  $\lambda$  の関係が明らかになった. また, 5. において, パターンが多い場合に 1 次 GLA が大域収束するための, 解領域の角度  $\psi_{\min}$  と  $\lambda$  の関係が, パターンが 2 個の場合において理論的に導き出された  $\theta - \lambda$  の関係で近似できることが明らかになった.

式 (10), (11) より, パターンが 2 個の場合には, 学習係数  $\lambda$  を 2 とすれば  $\theta$  によらずに 1 次 GLA は大域収束する. また図 10 は, パターンが多い場合でも, 学習係数  $\lambda$  を 2 とすることにより 1 次 GLA が大域収束することを示唆している. これらの事実より,  $\lambda = 2$  は 1 次 GLA において特別な条件であると言える.

5. の計算機シミュレーションにより, 1 次 GLA が大域収束するための,  $\psi_{\min} - \lambda$  の関係が,  $\theta - \lambda$  の関係で近似できることが明らかになったが, より一般的に GLA の学習アルゴリズムとしての有効性を示すためには, ある学習係数における大域収束性を示す必要がある. そのためにここでは, 学習係数  $\lambda$  を 2 とすれば, 線形分離可能なパターン分類問題に対して, パターン数にかかわらず 1 次 GLA が大域収束することを証明する. 図 11 に  $\lambda = 2$  の場合の 1 次 GLA による結合荷重ベクトルの更新を示す. 1 次 GLA においては, 更新の目標となる部分空間  $C$  は  $N - 1$  次元の超平面である.

(証明) いま, 結合荷重ベクトル初期値を  $w(0)$ ,  $i (\leq n)$  回更新後の結合荷重ベクトルを  $w(i)$  とし  $w(i)$  はいずれも解領域内にないとする. 但し,  $\|w(0)\| = 1$  であるとする.  $\lambda = 2$  の 1 次 GLA においては結合荷重ベクトルの更新は, その時点で式 (1), (2) とクラスが矛盾するパターンベクトル  $x$  の直交補空間  $C$  に対称に行われるから, 図 11 からも明らかなように以下の

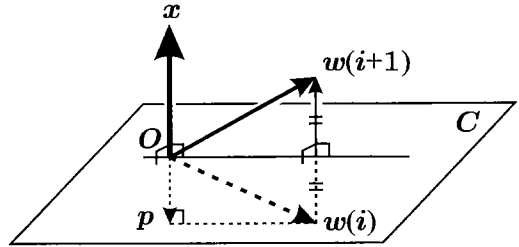


図 11 1 次 GLA ( $\lambda = 2$ ) における結合荷重ベクトルの更新

Fig. 11 Weights update in 1-GLA ( $\lambda = 2$ ).

式が成り立つ.

$$\|w(i+1)\| = \|w(i)\| \tag{12}$$

いま,  $\|w(0)\| = 1$  であるから,

$$\|w(i)\| = 1 \tag{13}$$

ここで,  $i$  回目の更新ベクトルを  $\Delta w(i)$ , そのとき用いられたパターンベクトルを  $x$ ,  $w(i)$  の  $x$  への射影を  $p$  とすると [10],

$$\begin{aligned} p &= x(x^T x)^{-1} x^T w(i) \\ &= x(\|x\|^2)^{-1} x^T w(i) \\ &= \frac{xx^T}{\|x\|^2} w(i) \end{aligned} \tag{14}$$

$$\begin{aligned} \therefore \Delta w(i) &= -2p \\ &= -2 \frac{xx^T}{\|x\|^2} w(i) \end{aligned} \tag{15}$$

ノルムが 1 であり, すべてのパターンを正しく分類するような結合荷重ベクトルのひとつを  $w^*$  とすると

$$\begin{aligned} w^{*T} \Delta w(i) &= w^{*T} \frac{-2xx^T}{\|x\|^2} w(i) \\ &= -\frac{2}{\|x\|^2} w^{*T} x x^T w(i) \\ &> 0 \end{aligned} \tag{16}$$

( $\because w^{*T} x$  と  $x^T w(i)$  は異符号)

また,

$$\begin{aligned} w(n)^T w^* &= \left( w(0)^T + \sum_{i=0}^{n-1} \Delta w(i)^T \right) w^* \\ &= w(0)^T w^* + \sum_{i=0}^{n-1} \Delta w(i)^T w^* \\ &\geq -1 + \sum_{i=0}^{n-1} \Delta w(i)^T w^* \end{aligned} \tag{17}$$



ここで、 $w(n)$ ,  $w^*$  はともにノルムが1であるから、

$$|w(n)^T w^*| \leq 1 \quad (18)$$

式(17), (18)より、

$$\begin{aligned} \sum_{i=0}^{n-1} \Delta w(i)^T w^* &= \sum_{i=0}^{n-1} \|\Delta w(i)\| \|w^*\| \cos \alpha_i \\ &\leq 2 \end{aligned} \quad (19)$$

ここで、 $\Delta w(i)$  と  $w^*$  のなす角を  $\alpha_i$  とおいた。式(16)より  $\alpha_i < \frac{\pi}{2}$  であるが、 $\Delta w(i)$  は  $i$  回目の更新に用いられるパターンと平行であり、パターン数は有限であるから、 $\alpha_i$  には最大値がある。これを  $\beta$  とすると次式が成り立つ。

$$0 \leq \alpha_i \leq \beta < \frac{\pi}{2} \quad (20)$$

$$\therefore \cos \alpha_i \geq \cos \beta > 0 \quad (21)$$

式(19), (21)より、次式が成り立つ。

$$\sum_{i=0}^{n-1} \|\Delta w(i)\| \|w^*\| \cos \beta \leq 2 \quad (22)$$

$$\therefore \sum_{i=0}^{n-1} \|\Delta w(i)\| \leq \frac{2}{\|w^*\| \cos \beta} \quad (23)$$

以下では、数学的帰納法を用いて  $n$  が有界でないと仮定すると矛盾が生じることを示す。

いま、 $n$  が有界でないと仮定する。この場合、Dirichlet-Riemann の定理 [11] より、式(23)は  $w(i)$  が  $w_\infty$  に収束することを意味する。このとき、 $\|w_\infty\| = 1$  である。また、更新が無限回行われるということであるから、分類の対象としている  $P$  個のパターンの中に、更新のために無限回用いられるパターンが存在することになる。これらを、 $x_j$  ( $j = 1, 2, \dots, P'$ ) とする。また、これらが張る空間を  $Z$  とする。 $x_j$  ( $j = P' + 1, P' + 2, \dots, P$ ) は有限回しか用いられないので、これらによる更新が終わった後の  $w(i)$  について考えることにする。なお、1個のパターンについては1回の更新で必ず式(1), (2)とクラスの矛盾が解消されるので、 $P' \geq 2$  である。

式(23)より  $\lim_{i \rightarrow \infty} \|\Delta w(i)\| = 0$  である。これは、 $w(i)$  の収束点が  $Z$  の直交補空間内にあることを示している。

まず、パターン次元が2の場合について考える。この場合、 $P' \geq 2$  であるから、 $Z$  の直交補空間は原点

であり、 $\|w_\infty\| = 1$  に矛盾する。よって、 $n$  は有界である。

次に、パターン次元が  $N + 1$  の場合について考える。但し、パターン次元が  $2 \sim N$  の場合には  $n$  が有界であるとする。 $P' \geq N + 1$  の場合には  $Z$  の直交補空間は原点となり、 $\|w_\infty\| = 1$  に矛盾するので、 $P' \leq N$  である。 $w(i)$  を  $Z$  の成分  $w(i)'$  とその直交補空間の成分  $w(i)^\perp$  に分けると、 $w(i)^\perp$  は更新により変化しない。 $P' \leq N$  より  $Z$  の次元は  $N$  以下なので、 $w(i)$  は  $Z$  において、有限回の更新で  $x_j$  ( $j = 1, 2, \dots, P'$ ) に関する解領域に到達する。すなわち、 $\text{sgn}(x_j^T w(i)') = y_j^*$  ( $j = 1, 2, \dots, P'$ ) が成り立つようになる。但し、 $y_j^*$  は  $x_j$  に対して基本パーセプトロンが出力すべき値である。ここで、 $x^T w(i) = x^T (w(i)' + w(i)^\perp) = x^T w(i)'$  であることに注意すると、 $\text{sgn}(x_j^T w(i)) = y_j^*$  ( $j = 1, 2, \dots, P'$ ) も成り立つことになる。これは、 $w(i)$  が  $x_j$  ( $j = 1, 2, \dots, P'$ ) に対して正しい出力をすることを示している。また、 $x_j$  ( $j = P' + 1, P' + 2, \dots, P$ ) に関しては既に更新が終了していることを考え合わせると、このパターン分類問題に関して結合荷重は解領域に到達していることになる。すなわち、 $n$  が有界でないという仮定は誤りである。

以上よりパターン次元にかかわらず  $n$  が有界であることが示された。すなわち、学習係数  $\lambda$  を2とすれば、線形分離可能なパターン分類問題に対して、パターン数にかかわらず1次GLAは大域収束する。□

本章の証明は、学習アルゴリズムとしてのGLAの理論的な有効性を示すものである。

## 7. むすび

本論文では、まず、基本パーセプトロンの学習アルゴリズムとして等比学習アルゴリズム (Geometric Learning Algorithm, GLA) を提案した。次に、2個のパターンに対して1次GLAが収束するために、2パターンの角度  $\theta$  と学習係数  $\lambda$  が満たすべき関係を理論的に導出した。また、パターンが3個以上の場合に“解領域の角度  $\psi_{\min}$ ”を導入し、1次GLAが収束するための  $\psi_{\min} - \lambda$  の関係が、2パターンの場合の  $\theta - \lambda$  の関係で近似できることを計算機シミュレーションで示した。更に、 $\theta - \lambda$  の関係において  $\theta$  によらずに収束する条件である  $\lambda = 2$  に対して、1次GLAがパターン数にかかわらずに収束することを証明した。こ

れらにより、1次 GLA の収束性が保証され、学習アルゴリズムとしての有効性が確認された。また、1次 GLA が収束するための  $\psi_{\min}$  と  $\lambda$  が分布する範囲が近似的に明らかになった。収束範囲における学習のふるまいを明らかにすることが今後の課題である。

文 献

[1] S. Haykin, Neural Networks—A Comprehensive Foundation—, Macmillan College Publishing Company, New York, 1994.  
 [2] H. Oh and S.C. Kothari, "Adaptation of the relaxation method for learning in bidirectional associative memory," IEEE Trans. Neural Networks, vol.5, no.4, pp.576-583, July 1994.  
 [3] 尾関和彦, 梅田哲夫, "アフィン部分空間への直交射影を用いた適応フィルタ・アルゴリズムとその諸性質," 信学論 (A), vol.J67-A, no.2, pp.126-132, Feb. 1984.  
 [4] 服部元信, 萩原将文, "双方向連想メモリの交点学習法," 電学論 (C), vol.116-C, no.7, pp.755-761, July 1996.  
 [5] S. Miyoshi and K. Nakayama, "A geometric learning algorithm for elementary perceptron and its convergence analysis," Proc. 1997 IEEE Int. Conf. on Neural Networks, Houston, Texas, USA, pp.1913-1918, June 1997.  
 [6] 三好誠司, 中山謙二, "基本パーセプトロンの等比学習とその収束条件," 1997 信学総大, no.D-2-35, March 1997.  
 [7] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," Psychological Review, vol.65, pp.386-408, 1958.  
 [8] J.I. Nagumo and A. Noda, "A learning method for system identification," IEEE Trans. Autom. Control, vol.AC-12, pp.282-287, 1967.  
 [9] A.E. Albert and L.S. Gardner, Jr, Stochastic Approximation and Nonlinear Regression, MIT Press, 1967.  
 [10] G. ストラング, 線形代数とその応用, p.126, 産業図書, 東京, 1978.  
 [11] 広中平祐ほか, 現代数理科学事典, p.1083, 大阪書籍, 大阪, 1991.

付 録

1. 発振状態の安定性

$0 < \theta \leq \frac{\pi}{2}$  で、 $\lambda \geq \frac{\tan(\frac{\pi}{4} + \frac{\theta}{2})}{\tan(\frac{\pi}{4} - \frac{\theta}{2})} + 1$  または  $\lambda \leq \frac{\tan(\frac{\pi}{4} - \frac{\theta}{2})}{\tan(\frac{\pi}{4} + \frac{\theta}{2})} + 1$  の場合に存在する二つ (等号の場合に一方は一つ) の発振状態のうち、一方は安定であり、他方は不安定である (等号の場合の発振状態は片側からのみ安定である) ことを証明する。ここで、結合荷重ベクトルの値は図 5 の  $\phi$  で表すことにする。

(証明) いま、 $f(\phi_0) = \lambda_0$  が成り立つとすると、式 (8) より、

$$\lambda_0 = \frac{\tan \phi_0}{\tan(\phi_0 - \theta)} + 1 \tag{A.1}$$

結合荷重ベクトルが  $\phi_0$  から微小だけずれた位置  $\phi_0 + \delta$  ( $0 < \delta \ll 1$ ) にあり、更新が 1 回行われて  $\phi_0 + \epsilon$  に移動したとすると、

$$\lambda_0 = \frac{\tan(\phi_0 + \epsilon)}{\tan(\phi_0 + \delta - \theta)} + 1 \tag{A.2}$$

•  $\frac{\pi}{4} + \frac{\theta}{2} \leq \phi_0 < \frac{\pi}{2}$  または  $\frac{\pi}{2} < \phi_0 \leq \frac{3\pi}{4} + \frac{\theta}{2}$  の場合  
 このとき、図 6 より  $\left. \frac{d}{d\phi} \left\{ \frac{\tan \phi}{\tan(\phi - \theta)} + 1 \right\} \right|_{\phi=\phi_0} \geq 0$  であるので、

$$\lambda_0 < \frac{\tan(\phi_0 + \delta)}{\tan(\phi_0 + \delta - \theta)} + 1 \tag{A.3}$$

が成り立つ。式 (A.2)、(A.3) より、

$$\frac{\tan(\phi_0 + \epsilon)}{\tan(\phi_0 + \delta - \theta)} < \frac{\tan(\phi_0 + \delta)}{\tan(\phi_0 + \delta - \theta)} \tag{A.4}$$

•  $\frac{\pi}{4} + \frac{\theta}{2} \leq \phi_0 < \frac{\pi}{2}$  の場合は式 (A.4) より  $\epsilon < \delta$  が成り立つ。よって発振状態は安定である。

•  $\frac{\pi}{2} < \phi_0 \leq \frac{3\pi}{4} + \frac{\theta}{2}$  の場合は式 (A.4) より  $\epsilon > \delta$  が成り立つ。よって発振状態は不安定である。

•  $\theta \leq \phi_0 < \frac{\pi}{4} + \frac{\theta}{2}$  または  $\frac{3\pi}{4} + \frac{\theta}{2} < \phi_0 \leq \pi$  の場合  
 このとき、図 7 より  $\left. \frac{d}{d\phi} \left\{ \frac{\tan \phi}{\tan(\phi - \theta)} + 1 \right\} \right|_{\phi=\phi_0} < 0$  であるので、

$$\lambda_0 > \frac{\tan(\phi_0 + \delta)}{\tan(\phi_0 + \delta - \theta)} + 1 \tag{A.5}$$

が成り立つ。式 (A.2)、(A.5) より、

$$\frac{\tan(\phi_0 + \epsilon)}{\tan(\phi_0 + \delta - \theta)} > \frac{\tan(\phi_0 + \delta)}{\tan(\phi_0 + \delta - \theta)} \tag{A.6}$$

•  $\theta \leq \phi_0 < \frac{3\pi}{4} + \frac{\theta}{2}$  の場合は式 (A.6) より  $\epsilon > \delta$  が成り立つ。よって発振状態は不安定である。

•  $\frac{3\pi}{4} + \frac{\theta}{2} < \phi_0 \leq \pi$  の場合は式 (A.6) より  $\epsilon < \delta$  が成り立つ。よって発振状態は安定である。

以上より、発振状態が二つ存在する場合には、一方は安定、他方は不安定であり、発振状態が一つだけ存在する場合には片側からのみ安定であることが証明された。 □

2. 式 (10), (11) の十分性

式 (10) または式 (11) が成り立つ場合に有限回の更新で学習が必ず収束することを証明する。ここで、結合荷重ベクトルの値は図 5 の  $\phi$  で表すことにする。

(証明) いま, 式 (10) または式 (11) が成り立っているとする. 結合荷重ベクトルの位置が角度  $\alpha (> \theta)$  にあり, 更新が 1 回行われて  $\beta$  に移動したとすると,

$$\lambda = \frac{\tan \beta}{\tan(\alpha - \theta)} + 1 \quad (\text{A}\cdot 7)$$

- $\theta < \alpha < \frac{\pi}{2}$  の場合

この場合, 式 (10) が成り立っていることになるから, 図 6 より

$$\lambda < \frac{\tan \alpha}{\tan(\alpha - \theta)} + 1 \quad (\text{A}\cdot 8)$$

$$\therefore \frac{\tan \beta}{\tan(\alpha - \theta)} + 1 < \frac{\tan \alpha}{\tan(\alpha - \theta)} + 1 \quad (\text{A}\cdot 9)$$

$$\therefore \tan \beta < \tan \alpha \quad (\text{A}\cdot 10)$$

$$\therefore \beta < \alpha \quad (\text{A}\cdot 11)$$

すなわち, 結合荷重ベクトルは単調に解領域に近づいてゆく. また, 式 (10) または式 (11) が成り立つ場合には発振状態が存在しないのであるから, ある値に漸近することはない. よって, 必ず有限回の更新で解領域に到達し学習は収束する.

- $\frac{\pi}{2} < \alpha < \pi$  の場合

いま式 (10) または式 (11) が成り立っているので, 図 6, 図 7 より

$$\lambda > \frac{\tan \alpha}{\tan(\alpha - \theta)} + 1 \quad (\text{A}\cdot 12)$$

$$\therefore \frac{\tan \beta}{\tan(\alpha - \theta)} + 1 > \frac{\tan \alpha}{\tan(\alpha - \theta)} + 1 \quad (\text{A}\cdot 13)$$

- $0 < \alpha - \theta < \frac{\pi}{2}$  の場合

式 (A-7) より

$$\tan \beta = (\lambda - 1) \tan(\alpha - \theta) > 0 \quad (\text{A}\cdot 14)$$

$$\therefore 0 < \beta < \frac{\pi}{2} \quad (\text{A}\cdot 15)$$

すなわち, 更新が 1 回行われることにより結合荷重ベクトルの位置は  $\frac{\pi}{2}$  より大きい値から  $\frac{\pi}{2}$  より小さい値に移動する.

- $\frac{\pi}{2} < \alpha - \theta < \pi$  の場合

式 (A-13) より

$$\tan \beta < \tan \alpha \quad (\text{A}\cdot 16)$$

$$\therefore \beta < \alpha \quad (\text{A}\cdot 17)$$

すなわちこの場合, 結合荷重ベクトルは単調に減少する. また, 式 (10) または式 (11) が成り立つ

場合には発振状態が存在しないのであるから, ある値に漸近することはない. よって, 必ず有限回の更新で  $0 < \alpha - \theta < \frac{\pi}{2}$  になる.

以上より, 式 (10) または式 (11) が成り立つ場合には有限回の更新で学習が必ず収束することが証明された. □

(平成 9 年 9 月 24 日受付, 11 月 13 日再受付)



三好 誠司 (正員)

昭 61 京大・工・電気卒, 昭 63 同大学院修士課程了. 同年 NEC 入社. 宇宙開発事業部にて人工衛星搭載用通信機器の研究開発に従事. 平 6 神戸高专・電子工学科・講師, 平 10 同助教授, 現在に至る. 平 10 工博 (金沢大). ニューラルネットワークによる連想記憶, パターン分類に関する研究に従事. 電気学会会員.



池田 和司 (正員)

平 1 東大・工・計数卒. 平 6 同大学院博士課程了. 工博. 同年金沢大・工・電気・情報工学科助手. 平 9 同講師. 平 10 京大学院情報学研究科講師, 現在に至る. 平 7 香港中文大・工・計算機科学科研究員. 計算論的学習理論, 適応システム, ニューロダイナミックスの研究に従事. IEEE, 神経回路学会, 応用数理学会各会員.



中山 謙二 (正員)

昭 46 東工大・工・電子卒. 昭 46~47 同大学院研究生. 昭 58 工博 (東工大). 昭 47 NEC 入社. 伝送通信事業部および C&C システム研究所にて, 通信用各種フィルタおよびデジタル信号処理の研究開発に従事. 昭 63 金沢大・工・電気・情報工学科助教授. 平 2 同教授となり現在に至る. 最近の研究テーマは主として適応信号処理およびニューラルネットワーク. 昭 62.9 IEEE Circuits & Devices Mag. 論文賞受賞, 著書「SC 回路網の設計と応用」(東海大学出版会) 他. IEEE シニア会員, INNS 会員.