パーセプトロンの直交射影学習とその収束性に関する研究

氏　　　名	三　好　誠　司
生 年 月 日	
本　　　籍	大阪府
学 位 の 種 類	博士（工学）
学 位 記 番 号	博甲第246号
学位授与の日付	平成10年3月25日
学位授与の要件	課程博士（学位規則第4条第1項）
学位授与の題目	パーセプトロンの直交射影学習とその収束性に関する研究
論 文 審 査 委 員	（主査）中山　謙二 （副査）船田　哲男，村本健一郎，橋本　秀雄，西川　清

学位論文要旨

This thesis proposes a fast learning algorithm for the perceptron applied to pattern classification and analyzes its convergence properties.

The affine projection algorithm for adaptive filters is modified to perceptron. In a new learning algorithm, the connection weight vector $\boldsymbol{w}(n)$ is updated vertically towards the orthogonal complement of k patterns. A rate of the distances among $\boldsymbol{w}(n)$, the complement and $\boldsymbol{w}(n+1)$ is constant. Thus, it is called "geometric learning algorithm (GLA)". When k patterns are simultaneously used for each update, it is called kth–order GLA. First, convergence property of the 1st–order GLA applied to two–patterns classification is analyzed. The relation between the angle θ of two patterns and the learning rate λ for convergence is theoretically derived. λ determines the previously described rate. When the number of the patterns is larger than 2, a new concept "an angle of the solution area $\psi_{\min}$" is introduced, in which the weight vector solutions exist. It is numerically confirmed that the $\theta - \lambda$ relation is good approximation of $\psi_{\min} - \lambda$. Furthermore, it is theoretically proved that the 1st–order GLA with $\lambda = 2$ always converges for any number of patterns.

The GLA demonstrates some typical behavior when $\lambda = 2$, which means that $\boldsymbol{w}(n)$ and $\boldsymbol{w}(n+1)$ are symmetric with respect to the complement. Thus, the GLA with $\lambda = 2$ is discriminated as "symmetric learning algorithm (SLA)". The followings are theoretically and numerically proved. First, for N–dimensional patterns, $k < N$ is necessary for convergence. Second, the convergence speed is the maximum when $k = N/2$.

パーセプトロンは，McCulloch と Pitts によって提案された神経細胞のモデルを基本に，Rosenblatt によって考案された線形識別回路であり，ニューラルネットワークの基本要素として重要である．パーセプトロンによるパターン分類の学習アルゴリズムとしては，いわゆる "パーセプトロン学習" がよく知られている．時刻 n における結合荷重ベクトルを $\boldsymbol{w}(n)$ ，学習係数を c ，時刻 n においてパーセプトロンの出力が望みの出力とは異なるようなパターンの1つを $\boldsymbol{x}$，そのパターンに対する望みの出力を y^* とするとき，パーセプトロン学習における結合荷重更新の式は， $\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + cy^*\boldsymbol{x}$ である．この学習アルゴリズムの特長は，線形分離可能なパターン分類問題に対して，有限回の更新で，すべてのパターンを正しく分類する結合荷重に必ず到達することである．しかし更新式からわかるように，毎回の更新に1個のパターンのみを用い，また，結合荷重の修正量が一定であるため，複雑なパターン分類問題においては収束が遅いという問題点がある．

一方，適応フィルタの分野において，NLMS アルゴリズムをブロック信号処理に拡張した適応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）が知られている．

本研究では，パーセプトロンのための高速な学習アルゴリズムとして，APA を適用した新しい直交射影学習アルゴリズムの提案と，その収束性に関する理論的，数値的解析を行った．

本研究で対象とするパーセプトロンの構造を図 1 に，動作方程式を式（1），(2）に示す．

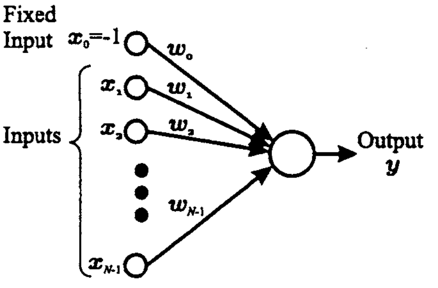

図 1: パーセプトロン

$$u = \sum_{i=0}^{N-1} w_i x_i \tag{1}$$

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

式（1）の u に 0 を代入すると，傾きが w_i $(i = 0, 1, \cdots, N-1)$ で決定され原点を通る超平面を表す式になる．よって，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 2 つのクラスを分類する能力を有すると言える．

適応フィルタの分野においては，正規化最小 2 乗平均（NLMS）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）がよく知られている．本研究では，まず，APA をパーセプトロンに適用することにより，等比学習アルゴリズム（geometric learning algorithm : GLA）を提案した．GLA を以下に示す．

[step1] $\boldsymbol{w}(0)$ を乱数で初期設定する．
[step2] パーセプトロンの出力が望みの出力と異な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

$$\boldsymbol{X} = (\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^k)^T \tag{3}$$

そうでないなら

$$\boldsymbol{X} = (\boldsymbol{x}^1, \boldsymbol{x}^2, \cdots, \boldsymbol{x}^{k_0})^T \tag{4}$$

[step5] 結合荷重を次式で更新する．

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) - \lambda \boldsymbol{X}^+ \boldsymbol{X} \boldsymbol{w}(n) \tag{5}$$

[step6] **step2** に戻る．

ここで $\boldsymbol{X}^+$ は $\boldsymbol{X}$ の Moore–Penrose 一般逆行列である．式（3），(4）において $\boldsymbol{x}^1 \sim \boldsymbol{x}^k$ または $\boldsymbol{x}^1 \sim \boldsymbol{x}^{k_0}$ は，パーセプトロンの出力が望みの出力と異なるパターンから選択する．この選択をどのように行うかについてはここでは規定しない．式（5）は，GLA では結合荷重ベクトルの毎回の更新を，k 個のパターンベクトルの直交補空間に向かって垂直に行うことを示している．このときの k を GLA の“次数”とよぶ．なお，この学習アルゴリズムにおいては，k 個のパターンに対する解領域までの距離と更新ベクトルの大きさの比 λ を一定に毎回の更新を行うので，これを等比学習アルゴリズムと呼ぶ．

パーセプトロン学習と異なり，GLA の場合には，線形分離可能なパターン分類問題に対しても，常に収束するとは限らない．図 2 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．なお，図 2 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が，この正規化が収束条件に影響を与えることはない．

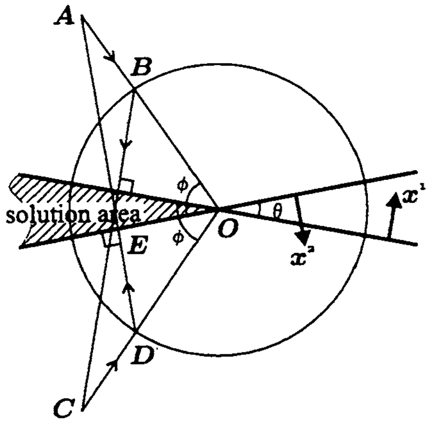

図 2: 1 次 GLA における発振状態.

まず，もっとも基本的な場合として，図 2 に示すような発振状態が生じうる条件の解析を行い，1 次 GLA で 2 パターンの分類を行う際に結合荷重の初期値に関わらず有限回の更新で収束するための必要十分条件を導出した．導出された理論収束条件を以下に示す．

- $0 < \theta \leq \frac{\pi}{2}$ の場合

$$\frac{\tan(\frac{\pi}{4}+\frac{\theta}{2})}{\tan(\frac{\pi}{4}-\frac{\theta}{2})}+1 > \lambda > \frac{\tan(\frac{\pi}{4}-\frac{\theta}{2})}{\tan(\frac{\pi}{4}+\frac{\theta}{2})}+1 \tag{6}$$

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

$$\lambda > 1 \tag{7}$$

次に，パターンが多い場合について考える．図 3 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．このように，パターンが 3 個以上になると，解領域の形状が複雑になり，また，パターン提示の順序という自由度が加わるので，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（6）の θ 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，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 θ を包含する新しい概念として，“解領域の角度 $\psi_{\min}$ ” を導入した．すなわち，解領域の角度 $\psi_{\min}$ とは，“原点から解領域を見た角度の最小値” である．さらに，解領域の角度の計算方法を考案した．

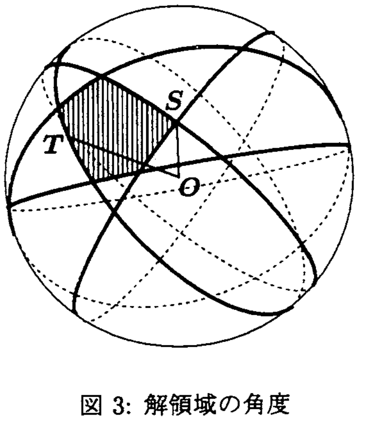

図 3: 解領域の角度

1 次 GLA が収束するための，解領域の角度 $\psi_{\min}$ と学習係数 λ の関係を計算機シミュレーションにより調べた．図 4 は，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 (6) 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，ひとつのパターン集合に関してのいろいろな λ の値における学習収束性を表している．ここで，"o" は多数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，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 4 より，多パターンに対して 1 次 GLA が収束するための $\psi_{\min}-\lambda$ 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 の収束条件を決める重要なパラメータであることがわかった．

式 (6)，(7) より，パターンが 2 個の場合には，学習係数 λ を 2 とすれば θ によらずに 1 次 GLA は収束する．また図 4 は，パターンが多い場合でも，学習係数 λ を 2 とすることにより 1 次 GLA が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．そこでさらに，$\lambda=2$ の場合に 1 次 GLA がパターン数に関わらずに必ず収束することを，背理法と数学的帰納法を用いて証明した．この証明は，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．

GLA においては，k 個のパターンを正しく分類する領域，すなわち，解領域までの距離が大きい場合には結合荷重の更新量が大きくなるので，パーセプトロン学習より高速である．パーセプトロン学習と 1 次 GLA

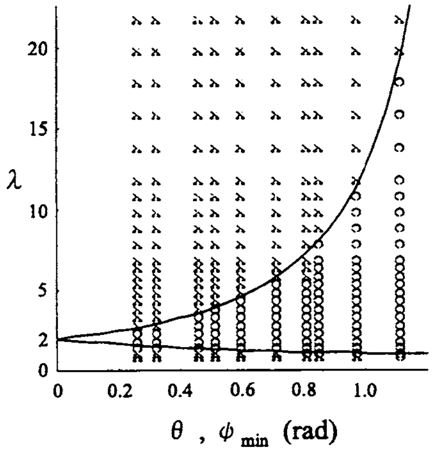

図 4: 1 次 GLA の収束条件

表 1: SLA が収束する次数 k の上限

P	$N=3$	$N=4$	$N=7$	$N=10$
2	P	P	P	P
3	P	P	P	P
4	1, P	P	P	P
5	2, P	2, 3, P	P	P
6	2	2, 3, P	P	P
7	2	3	P	P
8	2	3	P	P
9	2	3	P	P
10	2	3	5, 6, P	P
11	2	3	5, 6, P	P
12	2	3	5, 6, P	P
13	2	3	6	P
14	2	3	6	9, P
15	2	3	6	8, 9, P
16	2	3	6	7, 8, 9, P
17	2	3	6	8, 9, P
18	2	3	6	8, 9
19	2	3	6	9
20	2	3	6	9

の学習速度の比較に関する計算機シミュレーションを行い，特に解領域の角度が小さい問題に対しては 1 次 GLA が高速であることを明らかにした．

すでに述べたように，GLA において $\lambda = 2$ の場合は，学習が必ず収束し，また，収束速度も最大となるなど特徴的な特性を有し，学習アルゴリズムとして重要である．この場合には，結合荷重の更新をその目標となる直交補空間に対して対称なベクトルへと行うことになるので，第 4 章ではこの場合を特に区別して対称学習アルゴリズム（symmetric learning algorithm : SLA）と呼び，その収束条件および収束速度について解析した．まず，SLA の次数 k，パターン数 P，パターン次元 N と収束性の関係についての理論検討を行い，$P \geq 2N$ の場合に結合荷重初期値によらずに学習が収束するためには，$k < N$ でなければならないことを明らかにした．この条件を検証するために，SLA の次数 k と学習収束性に関する計算機シミュレーションを行った．学習が収束する次数 k の上限を表 1 に示す．

表 1 で，たとえば，$P = 18, N = 7$ において "6" となっているのは，それぞれが 18 個の 7 次元パターンからなる 20 組のパターン集合すべてが，1 ～ 6 次では収束し，7 次以上では収束しなかったことを表している．また，次数上限として複数の数字が書かれているのは，パターン集合により上限が異なることを意味している．たとえば，$P = 15, N = 10$ において "$8, 9, P$" となっているのは，それぞれが 15 個の 10 次元パターンからなる 20 組のパターン集合の中に，1 ～ 8 次でのみ収束したパターン集合，1 ～ 9 次でのみ収束したパターン集合，および 1 ～ 15 次すべてで収束したパターン集合の 3 種類のパターン集合が存在したことを表している．この計算機シミュレーションにより，$P \geq 2N$ の場合に結合荷重初期値によらずに学習が収束するためには，$k < N$ でなければならないことが検証された．

さらに，k と収束速度の関係についての理論検討を行い，次数がある程度小さい場合には次数が大きいほど収束が速く，また，k 次 SLA と（$N - k$）次 SLA が平均的な収束速度に関して等価であることを示すことにより，SLA においては $k = N/2$ で平均的な収束速度が最大になることを明らかにした．図 5 は，この条件を検証するために行った次数 k と収束速度に関する計算機シミュレーションの結果であり，理論検討の結果が正しいことを示している．APA の収束速度が次数と単調な関係にあることと比較すると，この特徴は SLA のきわめて興味深い特徴であると言える．

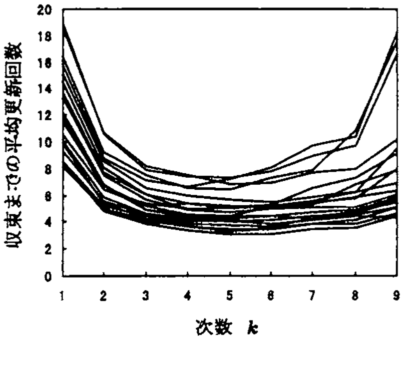

図 5: 学習収束までの平均更新回数（$N = 10, P = 20$）

本研究においては収束までの更新回数により学習速度の評価を行ったが，1 回の更新に要する計算量まで考慮した学習速度の解析や，GLA や SLA を多層のニューラルネットワークに適用する方法は今後の課題として残されている．

学位論文審査結果の要旨

平成10年1月27日に第1回学位論文審査委員会，平成10年2月5日に口頭発表会並びに第2回学位論文審査委員会を開催し，以下の通り判定した。

ニューラルネットワークの基本回路であるパーセプトロンに対して，結合荷重の学習法は従来から提案されている。しかし，パターン分類において1回の更新に1パターンを用いること，更新量が誤差に関わらず一定であることなどのために収束が遅いという問題がある。

本論文では，この問題を解決するためにアフィン射影アルゴリズムを適用した新しい直交射影学習アルゴリズムを提案している。1回の重み更新をk個のパターン（ベクトル）の直交補空間に向かって垂直に行い，解領域までの距離と更新後の荷重ベクトルの大きさの比が一定（λ）になるように更新を行う。線形識別ニューロンを用いたパターン分類に対する結合荷重の解はある領域にあるが，この領域を規定する”解領域の角度”を導入し，学習が収束するための更新比λとの関係を理論的に求めている。特に$\lambda=2$の場合にパターン数によらず収束することを証明した。$\lambda=2$の場合は，収束性が保証され，学習速度も速い特徴を有しているため，この場合の収束特性について更に解析し，パターン数（P）≧パターン次元（N）の場合に学習が収束するためには学習パターン数（k）＜Nが必要であること，及びk＝N／2において収束速度が最大になることを証明した。

以上の研究成果はニューラルネットワークの基本回路において新たな視点を開き，理論的な基礎を固めると同時に，実用的にも有用な方法論を確立しており，本論文は博士論文に値するものと判定する。