

研究論文

神経学習における正規化 LMS アルゴリズムの収束条件¹

池田和司², 三好誠司^{3,4*}, 中山謙二³²金沢大学工学部, ³金沢大学大学院自然科学研究科⁴神戸市立工業高等専門学校

Conditions for Convergence of the Normalized LMS Algorithm in Neural Learning

Kazushi Ikeda², Seiji Miyoshi^{3,4*}, Kenji Nakayama³²Faculty of Engineering, Kanazawa University³Graduate School of Natural Science & Technology, Kanazawa University⁴Kobe City College of Technology

The Perceptron Learning algorithm for linear dichotomy can be regarded as the LMS algorithm which is one of the most popular algorithms for transversal filters. The Normalized LMS (N-LMS) algorithm is one of the improved versions of the LMS algorithm for transversal filters and we apply it to linear dichotomies. In this paper, the proof of the convergence of the N-LMS algorithm for linear dichotomies in a finite number of iterations when the learning coefficient μ is unity, and a sufficient condition of μ for the convergence are given.

1. はじめに

適応フィルタの分野においては、入力ベクトル $x \in R^n$ に対して

$$y = x^T w \in R \quad (1)$$

を出力するトランスバーサルフィルタと呼ばれる線形フィルタがもっともよく用いられている。Least Mean Square (LMS) アルゴリズムは最も基本的なトランスバーサルフィルタの学習法であり、ある時刻の重みが w である時の重みの変化量 Δw を

$$\Delta w = \mu x(d - x^T w) \quad (2)$$

と定める。ここで μ は学習定数、 d は所望出力であり、 $(d - x^T w)$ は出力誤差を意味している。これは確率的最急降下法の一種であり、学習定数 μ が $0 < \mu < 2/\lambda_{\max}$ を満たす時、重みは平均二乗誤差を最小にする重みに確率収束する。ここで λ_{\max} は入力の自己相関行列 Σ の固有値の最大値である¹⁾。従って入力信号

の自己相関行列が不明である場合には、収束を保証する学習係数 μ を設定することができない。この点を改良し、収束性を高めたアルゴリズムとして、正規化 LMS アルゴリズム^{1,3,4)}が知られている。これは入力 x の大きさを考慮して重みの変化量 Δw を

$$\Delta w = \mu x(d - x^T w)/|x|^2 \quad (3)$$

とすることにより、 Δw が x の大きさに依らなくなっている。この時、確率収束する学習係数 μ の条件は $0 < \mu < 2$ である。

一方、パーセプトロンと呼ばれる線形二分機械

$$y = \text{sign}[x^T w] \quad (4)$$

はトランスバーサルフィルタに符号関数を組み合わせたものであり、離散値を出力するニューラルネットワークの最小構成要素としても用いられている。その学習法として提案されたパーセプトロン学習は、真のパラメータ w_0 を用いて

$$\Delta w = \frac{1}{2} x(\text{sign}[x^T w_0] - \text{sign}[x^T w]) \quad (5)$$

と表すことができ、これは LMS アルゴリズムを線形二分機械に応用したものだと言える。パーセプトロン学習は有限回の更新で学習が停止することが示されて

¹⁾ 1997年7月7日受付²⁾ 〒920 金沢市小立野 2-40-20³⁾ 〒920 金沢市小立野 2-40-20⁴⁾ 〒651-21 神戸市西区学園東町 8-3

いる^{5,6)}が、式(5)からもわかる通り、 x と w の大きさが学習アルゴリズムに考慮されていない。そのため、 w に比べ x が極端に小さい場合 (Fig. 1 a) や大きい場合 (Fig. 1 b) には収束が遅くなる。Fig. 1 では、斜線部分が例題に対して正しい出力を与える重み w の領域を表し、矢印が重みの変化を表している。線形二分機械についてもトランസバｰｻﾙフィルタと同様、入力と重みがわかっているので、その出力 $\text{sign}[x^t w]$ だけでなく $x^t w$ そのものの値がわかっている。従ってこれを利用すると、LMS アルゴリズムを正規化 LMS アルゴリズムに改良したのと同様に、パーセプトロン学習を改良することができる。入力 x に対して w が間違った符号を出力する時、 a を正定数とすると $-aw$ は正しい符号を出力するので、所望信号 d として $-ax^t w$ を利用することを考えよう。この時、LMS アルゴリズムを線形二分機械に応用したもののは

$$\begin{aligned}\Delta w &= \mu x(-ax^t w - x^t w) \\ &= -(1+a)\mu x x^t w\end{aligned}\quad (6)$$

と表される。ただし更新は $\text{sign}[x^t w]$ が正しく出力されない x についてだけ行う。また、 $x^t w = 0$ となる場合には重みの更新がされないので、本論文では等号になる場合は考えないことにする。 a による違いは μ で吸収できるので、以下では $a=1$ とする。式(6)では誤差が大きい時には重みは大きく更新され、誤差が小さい時には重みは小さく更新される。さらに x の大きさへの依存性をなくすため、正規化 LMS アルゴリズムで用いられている正規化の手法を応用すると、

$$\Delta w = -2\mu x x^t w / |x|^2 \quad (7)$$

とできる。このアルゴリズムを線形二分機械における正規化 LMS アルゴリズムと呼ぶことにする。これは幾何学的には、 $\mu=1/2$ の時には w を $x^t w=0$ の超平面上に正射影し、 $\mu=1$ の時には w を $x^t w=0$ の超平面について対称移動する (Fig. 2)。従って、線形二分機械における正規化 LMS アルゴリズムは直交射影アルゴリズムである。これを一度に複数の入力を用いたブロック信号処理に拡張した、ブロック直交射影アルゴリズム、アフィン射影アルゴリズムなどがトラン

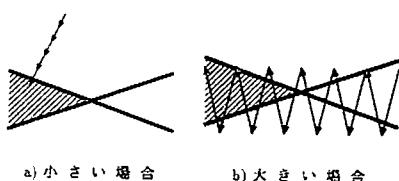


Fig. 1 パーセプトロン学習の概念図

スバｰｻﾙフィルタでは提案されており、それを線形二分機械に応用したものも等比学習アルゴリズム (GLA) として提案されている²⁾。これらの研究ではパーセプトロン学習における収束定理のような、収束性の厳密な解析は行われていない。しかし、基礎的な理論として、入力に対して正規化を行うアルゴリズムの収束性がそれを行わない場合のものとどのように異なるのかを明らかにすることは有意義なことである。また、トランｽバｰｻﾙフィルタでは正規化により収束条件が改善されたのに対し、符号関数という非線形要素が入ることでその性質の違いがどのように変わるものかも興味深いことである。

さらに、重みの正規化ということが収束性に与える影響も考察すべき問題である。例えば、正規化 LMS アルゴリズムは Δw が x の大きさに影響を受けず、また Δw の大きさが w の大きさに比例するから、 w の大きさを正規化することは学習の収束性に全く影響を与えない。重みの正規化を行うことは、その数値精度を十分に利用する上で重要なことである。しかし、通常のパーセプトロン学習に重みの正規化を併用すると有限回の更新で停止する保証が失われる。例えば Fig. 3 のような場合には、 x_1 を用いて更新すると A であった重みベクトルはベクトル x_1 が加えられて B となり、さらに正規化により C となる。次に x_2 を用いて更新すると C が D となり、正規化により A に戻る。従って、 x_1 と x_2 が交互に与えられる場合には学習は停止しない。

線形二分機械の正規化 LMS アルゴリズムの収束性の問題は、例題数が二つという特殊な場合については

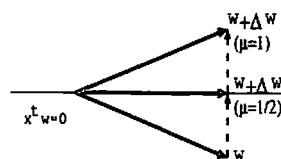


Fig. 2 正規化 LMS アルゴリズムの概念図

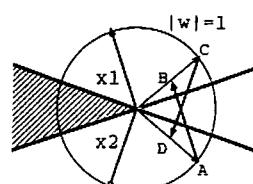


Fig. 3 正規化を伴うパーセプトロン学習の収束しない例 (入力が 2 次元で例題数が 2 の場合)

1次等比学習アルゴリズムの収束条件問題として三好ら²⁾によって解析されているが、一般の次元で一般的な例題数についての収束は示されていない。そこで本論文では、 $\mu=1$ ならば正規化LMSアルゴリズムが有限回の更新で学習が停止することを示す。さらに、与えられた例題に対して学習が有限回の更新で停止する μ の範囲を導出する。

2. 正規化LMSアルゴリズムの収束定理

トランスバーサルフィルタでは最適重み係数は出力誤差の期待値を最小にする1点であるのに対し、線形二分機械では与えられたすべての例題と同じ出力を出す重み（解と呼ぶ）が体積を持つ領域（正解領域と呼ぶ）になる。このことがトランスバーサルフィルタと線形二分機械の最大の違いであり、その学習特性に大きく影響を与える。すなわち、トランスバーサルフィルタでは、学習係数 μ が $0 < \mu < 2$ を満たす時には学習が進むにつれて重み係数は最適重み係数に収束するが、これは確率収束であるため、有限回の更新で最適重み係数が必ず得られるわけではない。しかし解が領域である線形二分機械の場合には、 $\mu=1$ ならば有限回の更新で重みは必ず解の一つになり学習が停止する。以下ではこのことを証明する。

与えられた*i*番目の例題は、入力ベクトル x_i と出力 y_i の組である。入力 x_i に対する正しい出力が -1 であるとき、入力 $-x_i$ に対する正しい出力は $+1$ であるので、与えられた例題の出力はすべて $+1$ であるとする。また、 x_i の大きさは学習に影響を与えないなので、 $|x_i|=1$ としてもよい。よって例題の入力 x_i の定義域は、 $m-1$ 次元超球面 S^{m-1} の半分、

$$S_i = \{x_i | x_i^T w > 0, x_i \in S^{m-1}\} \quad (8)$$

であると仮定する。ここで w_0 は真の重み係数である。例題の出力は常に $+1$ であるので、今後は x_i を例題と呼ぶ。

例題 $x_i, i=1, \dots, p$ が与えられた時、

$$D = \{w | x_i^T w > 0, i=1, \dots, p\} \quad (9)$$

で表される重み係数 w の集合 D が正解領域である。正解領域に含まれる任意の w は、入力 x_i に対して真の重み係数 w_0 と同じ符号を出力する。線形二分機械の重みの更新は

$$x_i^T w < 0$$

となる x_i について行われるので、重み係数 w が $w \in D$ となると学習は停止する。

本節では、 $\mu=1$ ならば正規化LMSアルゴリズムでの学習が有限回の更新で停止することを数学的帰納法を用いて証明する。 $\mu=1$ の時には例題 x_i が与えら

れると w は超平面 $x_i^T w = 0$ に対称に移動されるので、ベクトルの大きさは変わらない。従って、 w の大きさは常に1であるとしてよい。この時正解領域 D は超球面 S^{m-1} 上の多面体になる。正解領域 D に含まれる重みの一つを w^* とし、 w^* と x_i のなす角を $\pi/2 - \theta_i$ とすると $\theta_i > 0$ である。

まず重み係数ベクトル w の次元数 m が2である場合について示す。Fig. 4からわかる通り、 x_i を用いて更新を行うと、 $w + \Delta w$ と w^* のなす角は $2\theta_i$ だけ w と w^* のなす角より小さくなる。よって学習は有限回の更新で停止する。

次に、 $m \leq k$ ならば w の学習が有限回の更新で停止すると仮定し、 $m = k+1$ の時にも有限回の更新で学習が停止することを示す。

無限回の更新でも正解領域 D に入らないとすると、無限回適用される例題が存在する。これを $x_i, i=1, \dots, p'$ とする。 $x_i, i=p'+1, \dots, p$ は有限回しか用いられないもので、これらによる学習が終わった後の w を考えることにすると、学習に用いられる例題は $x_i, i=1, \dots, p'$ だけである。 w^* となす角が $\pi/2 - \theta_i$ であるような x_i で w を更新すると、 Δw の大きさが ε ならば w の w^* 方向の成分は $\varepsilon \sin \theta_i$ だけ増加する。 w の大きさは更新を行っても変化しないので、このことは w と w^* のなす角が減少することを意味する。よって無限回の更新でも正解領域 D に入らないためには、 ε は0に収束しなければならない。即ち $w(\infty)$ は $x_i^T w = 0$ 上に存在しなければならない。よって $x_i, i=1, \dots, p'$ が張る空間

$$X = \text{span}(x_i, i=1, \dots, p')$$

が R^{k+1} である場合には $w(\infty) = 0_{k+1}$ となり $|w| = 1$ に矛盾するので、 X は $R^k, m < k+1$ である。 w の更新は x_i の定数倍を加えることによって行われるから、 w の X に直交する成分は変化しない。そこで、 w を X とその直交補空間 X^\perp の成分に分解し、 $w' + w^\perp$ とする。 X 内において $x_i, i=1, \dots, p'$ によって定義される正解領域を D' とすると、 X は k 以下の次元の空

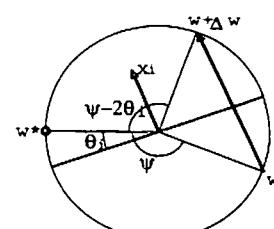


Fig. 4 正規化LMSによる学習（2次元の場合）

間なので、 w' は有限回の更新で D' に入り、この時任意の*i*について $x_i'w' > 0$ となる。ここで $x_i'w = x_i'w', i=1, \dots, p'$ が成り立つことに注意すると $x_i'w > 0, i=1, \dots, p'$ が成り立つので、 $x_i, i=1, \dots, p'$ は学習に用いることができず、また $x_i, i=p'+1, \dots, p$ による学習が終わっているという仮定によりこれらも学習に用いることができないため、学習は停止する。従って無限回の更新でも学習が停止しないという仮定は誤りであり、学習は有限回の更新で学習は停止する。

以上により、任意の次元の線形二分機械における正規化LMSアルゴリズムは $\mu=1$ ならば有限回の更新で学習が停止することが示された。

3. 収束のための学習係数の条件

前節では $\mu=1$ の時には有限回の更新で学習が停止することが示された。では、学習係数 μ は1でなければならぬのだろうか。1以外にも停止が保証され、なおかつ、収束性などの点で1よりすぐれた μ の値は存在しないのだろうか。そこで本節では、どんな初期値から学習を始めても有限回の更新で学習が停止するための μ の十分条件を導く。その結果得られる μ の範囲は与えられる例題に依存した1を含む区間となり、従って、実用上では $\mu=1$ としなければならないことが示される。

正解領域 D に含まれる重みの一つを w^* とした時、 x_i を用いた更新により w と w^* のなす角が有限の大きさだけ必ず減少するならば、前節と同様の方法で学習が有限回の更新で停止することを示すことができる。よって w が w^* となす角が必ず減少するような μ の範囲を求める。

w^*, x_i の張る2次元空間 W とその直交補空間 W^\perp にわけて考える。 w が W に含まれている時、Fig.5のような半径 $|w|$ の円を考える。 $-w^*$ と w が作る角 ψ と $-w^*$ と $w+\Delta w$ が作る角 ψ' をFig.5のようにとり、 $-\theta_i < \psi < \pi/2 - \theta_i$ について $\psi > \psi'$ となる μ の条件を考えると、 $-\theta_i < \psi \leq 0$ の時は $\mu > 0$ ならば $\psi' > 0$ となるので $\psi' > \psi$ となる。また、 $0 < \psi < \pi/2$

$-\theta_i$ の時には、Fig.5において重みベクトル $w + \Delta w$ の終点を表すDが、ACの延長線上にあれば $\psi' > \psi$ となる。ここで、ADの長さは μ に比例し、かつ $\mu = 1/2$ の時にDはBと一致することに注意すると、 $\mu / (1/2)$ がAC/ABよりも大きければ $\psi' > \psi$ となることがわかる。従って、

$$\begin{aligned} & [\sin(\psi + \theta_i) + \cos(\psi + \theta_i)\tan(\psi - \theta_i)]/\sin(\psi + \theta_i) \\ & < \mu / \frac{1}{2} \end{aligned} \quad (10)$$

を満たすための μ の条件

$$\mu > \frac{1}{2} \left[1 + \frac{\tan(\psi - \theta_i)}{\tan(\psi + \theta_i)} \right] \quad (11)$$

が、 $\psi < \psi'$ のための条件である。よって $\theta_i < \pi/4$ の時は

$$\mu > \frac{1}{2} \left[1 + \frac{\tan(\pi/4 - \theta_i)}{\tan(\pi/4 + \theta_i)} \right] \quad (12)$$

$$= \frac{1 + \tan^2 \theta_i}{(1 + \tan \theta_i)^2} \quad (13)$$

であり、 $\pi/4 \leq \theta_i < \pi/2$ の時は $\mu > \frac{1}{2}$ である($\mu = \frac{1}{2}$ の時には原点に射影されることがあるので除いてある)。

同様にして、 w^* と w のなす角 ψ と w^* と $w + \Delta w$ のなす角 ψ' をFig.6のようにとり、 $\theta_i < \psi < \pi/2 + \theta_i$ について $\psi > \psi'$ となる μ の条件を考えると、

$$\begin{aligned} & [\sin(\psi - \theta_i) + \cos(\psi - \theta_i)\tan(\psi + \theta_i)]/\sin(\psi - \theta_i) \\ & > \mu / \frac{1}{2} \end{aligned} \quad (14)$$

が $\psi > \psi'$ のための条件であり、従って

$$\mu < \frac{1}{2} \left[1 + \frac{\tan(\psi + \theta_i)}{\tan(\psi - \theta_i)} \right] \quad (15)$$

である。よって $\theta_i < \pi/4$ の時は

$$\mu < \frac{1}{2} \left[1 + \frac{\tan(\pi/4 + \theta_i)}{\tan(\pi/4 - \theta_i)} \right] \quad (16)$$

$$= \frac{1 + \tan^2 \theta_i}{(1 - \tan \theta_i)^2} \quad (17)$$

であり、 $\pi/4 \leq \theta_i < \pi/2$ の時は $\mu > 0$ である。

以上をまとめると、 $\theta_i < \pi/4$ の時は

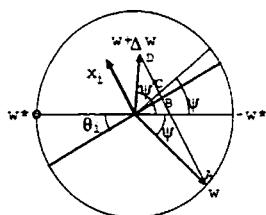


Fig. 5 $w \in \text{span}(w^*, x_i)$ の時 I

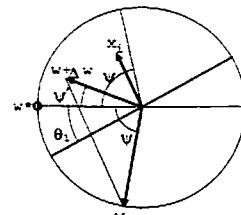


Fig. 6 $w \in \text{span}(w^*, x_i)$ の時 II

$$\frac{1+\tan^2\theta_i}{(1+\tan\theta_i)^2} < \mu < \frac{1+\tan^2\theta_i}{(1-\tan\theta_i)^2} \quad (18)$$

であり、 $\pi/4 \leq \theta_i < \pi/2$ の時は $\mu > \frac{1}{2}$ である。すべての例題 x_i について上記が成り立たなければならないので、 $\min_i \theta_i$ を最大にするように w^* をとりその時の θ_i の最小値を θ^* として $\theta^* < \pi/4$ の時

$$\frac{1+\tan^2\theta^*}{(1+\tan\theta^*)^2} < \mu < \frac{1+\tan^2\theta^*}{(1-\tan\theta^*)^2}, \quad (19)$$

$\pi/4 \leq \theta^* < \pi/2$ の時 $\mu > \frac{1}{2}$ が有限回の更新で学習が停止するための学習係数 μ の十分条件である。この条件を条件 I と呼ぶことにする。

次に、 w が W に含まれていない時を考える。あるベクトルと w^* のなす角を比較するには、ベクトルを正規化してから w^* 方向成分を比較すればよい。すなわち、その成分が増加することは、なす角が減少することを意味する。 w を $w = w' + w^\perp$, $w' \in W$, $w^\perp \in W^\perp$ と分解すると、 x_i を用いた更新において w^\perp は変化しない。そこで、 w , $w + \Delta w$ を W に射影すると Fig. 7 のようになる。ただし図中の円は原点を中心とした半径 $|w'|$ の円である。 $\mu < 1$ の時には $w' + \Delta w$ が円の内部 (B あるいは B') にあり、 $|w'| > |w' + \Delta w|$ であるので、 $w + \Delta w$ を w と同じ大きさに正規化した $\frac{|w|}{|w + \Delta w|}(w + \Delta w)$ は W^\perp 成分が正規化前より増加する。よってそれを W 上に射影した点はもとの点 (B または B') と原点を結ぶ直線 (l または l') 上でもとの点より外側かつ半径 $|w'|$ の円より内側 (BC または $B'C'$ 上) にある。 μ が条件 I を満たしている時には l または l' が w^* となす角が必ず減少するので、 BC または $B'C'$ 上の点の w^* 方向成分は増加する。すなわち $w + \Delta w$ と w^* のなす角は w と w^* のなす角より小さい。

$\mu > 1$ の時には $w' + \Delta w$ が円の外部にあり (D ある

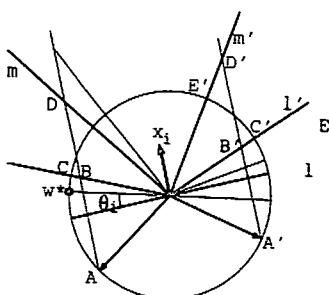


Fig. 7 $w \in \text{span}(w^*, x_i)$ の時

いは D'), $|w'| < |w' + \Delta w|$ であるので、 $w + \Delta w$ を w と同じ大きさに正規化した $\frac{|w|}{|w + \Delta w|}(w + \Delta w)$ は W^\perp 成分が正規化前より減少する。よってそれを W 上に射影した点はもとの点 (D または D') と原点を結ぶ直線 (m または m') 上でもとの点より内側かつ半径 $|w'|$ の円より外側 (DE または $D'E'$ 上) にある。 μ が条件 I を満たしている時には m または m' が w^* となす角が必ず減少するので、 DE または $D'E'$ 上の点の w^* 方向成分は増加する。すなわち $w + \Delta w$ と w^* のなす角は w と w^* のなす角より小さい。

以上により、条件 I が満たされる時には w^* となす角は減少することが示された。よって条件 I は線形二分機械における正規化 LMS アルゴリズムが有限回の更新で学習が停止するための十分条件である。例題数が 2 の場合について三好ら²⁾が導いた、有限回の更新で収束するための μ の必要十分条件は、条件 I と一致している。条件 I は与えられた例題と正解領域中のパラメータ w^* から得られる θ^* に依存するために実際には求めることができず、 $\theta^* > 0$ ということしかわからない。従って、有限回の更新での学習の停止を実用において保証するためには、学習係数 μ は 1 としなければならないことがわかる。

4. まとめ

いわゆるパーセプトロン学習はトランスバーサルフィルタにおける LMS アルゴリズムであることに着目すると、トランスバーサルフィルタの学習法である正規化 LMS アルゴリズムを線形二分機械 (パーセプトロン) の学習に導入することができる。この学習方法自体は新しいものではなく、それをブロック信号処理に拡張したアルゴリズムも提案されている²⁾が、パーセプトロン学習における収束定理のような厳密な収束性の議論はなされていなかった。そこで本論文では、学習係数 μ を 1 に設定すると正規化 LMS アルゴリズムが必ず有限回の更新で停止することを厳密に証明した。さらに、学習が有限回の更新で停止するための学習係数の十分条件を導出した。その条件は与えられた例題によって定まる 1 を含む区間であるため、実用上は $\mu = 1$ としなければならない。この結果は、パーセプトロン学習では学習係数は任意の正数でよいこと、あるいはトランスバーサルフィルタにおける正規化 LMS アルゴリズムの確率収束のための条件が $0 < \mu < 2$ であることと対比的である。

参考文献

- 1) Haykin, S. (1991) : Adaptive Filter Theory, 2nd ed., Eaglewood Cliffs, NJ.: Prentice-Hall
- 2) 三好誠司, 中山謙二, 池田和司 (1997) : 基本パーセプトロンの等比学習とその収束条件および雑音特性, 電子情報通信学会技報, NC 97-5, pp. 33-40
- 3) Nagumo, J. and Noda, A. (1967) : A learning method for system identification, IEEE Trans. AC, Vol. 12, pp. 282-287
- 4) 野田淳彦, 南雲仁一 (1968) : システムの学習的同定法, 計測と制御, Vol. 7, pp. 597-605
- 5) Rosenblatt, F. (1958) : The perceptron: A probabilistic model for information storage and organization in the brain, Psychological Review, Vol. 65, pp. 386-408
- 6) Rosenblatt, F. (1961) : Principle of Neurodynamics, Washington, D. C.: Spartan