

ニューラルネットにおける学習率の新しい更新則

—Delta-Bar-Delta-Bar 則—

落合 慶 広[†] 戸田 尚 宏[†] 臼井 支 朗[†]

ニューラルネット学習の加速化法として知られる Jacobs 法の収束の速さは、学習率の更新則：Delta-Bar-Delta 則 (DBD 則) に含まれる増減係数に依存しており、これらを適切に調整しなければ十分な加速効果が得られない。しかしながら、この問題の原因等については一切明らかにされていない。本論文では収束が遅くなる原因を考察し、DBD 則と慣性項を併用すると、評価関数曲面の谷において DBD 則による学習率の更新が適切に行われず、重みの振動を抑制できなくなることを明らかにする。次に、この問題点を解決した新しい学習率の更新則：Delta-Bar-Delta-Bar 則 (DBDB 則) を提案する。数値実験より、DBDB 則を用いると、学習率の増減率を調整しなくても、最も速く収束する場合と同程度の速さで収束することを示す。

A New Renewal Rule of Learning Rate in Neural Networks

—Delta-Bar-Delta-Bar Rule—

KEIHIRO OCHIAI,[†] NAOHIRO TODA^{††} and SHIRO USUI[†]

One disadvantage of the Jacobs algorithm is a convergence rate depends on the increment and decrement factors in the renewal rule of learning rates (Delta-Bar-Delta rule: DBD rule). It is hard to adjust these factors in order to get a faster convergence. However, the cause of this problem has not been cleared yet. In this paper, we point out that the DBD rule is not suitable for the renewal of learning rates in ravine regions of the error surface. Thus we propose a new renewal rule of learning rates, named the Delta-Bar-Delta-Bar rule (DBDB rule), to improve the robustness of setting the increment and decrement factors of the learning rate. We show that this rule is effective in extracting the best performance of the learning algorithm. Using the DBDB rule, the rate of convergence is substantially improved even though the learning rates are not optimally set.

1. はじめに

神経系をモデル化した階層型ニューラルネットワーク (以下 NN と略記) の学習には、誤差逆伝搬法¹⁾を利用した学習法が広く用いられている。しかし、この学習法には収束が遅いという欠点があり、これは、評価関数曲面が急峻な傾きをもつ“谷”，または、傾きをほとんどもたない“丘”のような形状となる部分における重みの振舞いに起因すると考えられている^{2),3)}。丘は2値から成る学習パターンを用いた学習課題で主に生じるが、谷はさまざまな学習課題において存在するものである。こうした観点から、われわれは、NN の学習において収束を遅くする主な原因は、評価関数曲面の谷で発生する重みの振動であると考え、Jacobs 法⁴⁾に基づいて振動を素早く抑制する加速化法

(Kick Out 法) を提案した^{5),6)}。

Jacobs 法は、勾配項のみをもつ重みの更新則と Delta-Bar-Delta 則 (DBD 則) と呼ばれる学習率の更新則から構成される加速化法である。この Jacobs 法の重みの更新則に慣性項を付加した改良法が Jacobs hybrid 法⁴⁾、さらに補正項を付加したものが Kick Out 法^{5),6)}である。

DBD 則は、収束を加速化するために、評価関数曲面の形状に基づいて学習率を自動的に調整する更新則であり、あらかじめ、学習率を増減させるための係数 (以下では、おのおのの係数を増加率、減少率と記し、まとめて増減率と略記する) を設定する必要がある。Jacobs 法の収束の速さは、こうした増減率の値に依存しており、増減率の値によって十分な加速効果が得られないことが知られている。特に、慣性項を含む Jacobs hybrid 法では、加速効果が著しく低下する場合がある。このため、速く収束させるためには増減率を調整しなければならず、この調整に手間を要すると

[†] 豊橋技術科学大学情報工学系
Department of Information and Computer
Sciences, Toyohashi University of Technology

いう問題が指摘されている³⁾。この問題は、Kick Out 法を用いた場合にも生じる。しかしながら、この問題がどのような場合に生じ、いかなる原因によって加速効果が低下するか等については一切明らかにされていない。

本論文では、初めに、上記の問題の原因について考察し、慣性項と DBD 則を併用すると、谷で生じる重みの振動を抑制できないため加速効果が低下することを明らかにする。次に、この問題は谷における学習率の更新方法が不適切であるために起こり、これは DBD 則で用いられている学習率の更新基準に起因することを指摘する。以上の点を考慮し、新たな更新基準を用いた学習率の更新則：Delta-Bar-Delta-Bar 則 (DBDB 則) を提案する。最後に、数値実験により、DBDB 則を用いると、増減率を調整しなくても、最も速く収束する場合と同程度の速さで収束することを示す。

2. 重みと学習率の更新則

本論文で用いる Jacobs 法、Jacobs hybrid 法、および Kick Out 法について簡単に説明する。NN の重みの個数を n 、重みを $w=(w_1, \dots, w_n)$ 、学習率を $\eta=(\eta_1, \dots, \eta_n)$ 、およびその増減率を κ (増加率)、 ϕ (減少率)、勾配 (評価関数の一次偏微分) を g 、平滑化した勾配 (以下、平滑化微分と略記) を $\bar{\delta}$ とすると、Jacobs 法、および DBD 則は次式で表される。

Jacobs 法

重みの更新則：

$$w_{k+1} = w_k + \Delta w_k, \quad (1)$$

$$\Delta w_k = -\text{diag}(\eta_k) g_k. \quad (2)$$

学習率の更新則：Delta-Bar-Delta 則 (DBD 則)

$$\begin{cases} \eta_{k,i} = \eta_{k-1,i} + \kappa & \text{if } \bar{\delta}_{k-1,i} \cdot g_{k,i} > 0, \\ \eta_{k,i} = \eta_{k-1,i} \cdot \phi & \text{if } \bar{\delta}_{k-1,i} \cdot g_{k,i} < 0, \\ \eta_{k,i} = \eta_{k-1,i} & \text{otherwise.} \end{cases} \quad (3)$$

$$\text{ただし, } \bar{\delta}_k = (1-\theta) \cdot g_k + \theta \cdot \bar{\delta}_{k-1}. \quad (4)$$

ただし、添字 k, i は、学習アルゴリズムの反復回数、NN の重みに対応した番号を表し、学習には、全学習パターンを提示した後に重みを更新する一括学習方式を用いる。また、(4)式の平滑化微分 $\bar{\delta}_k$ は、 $\bar{\delta}_k = (1-\theta)(g_k + \theta g_{k-1} + \theta^2 g_{k-2} + \dots)$ と書き表せる。これは、過去の反復点における勾配 (g_{k-1}, g_{k-2}, \dots) を係数 $\theta \in (0, 1)$ のべき乗で重み付けした値であり、重みの軌跡上において平滑化された大域的な勾配を表している。

さて、Jacobs 法では、NN の各重みに対応した個別の学習率が用いられており、各学習率を独立に更新する DBD 則((3), (4)式)が収束の加速化において重要な役割を果たしている。ある反復点における勾配の1成分 (i 成分とする)と、1反復前の点における平滑化微分の i 成分が同符号の場合、重みの更新に伴う評価関数曲面の形状には大きな変化がなかったと考えられる。この場合、重みの更新量は小さく、1反復前の点からほとんど移動していないと推察される。そこで、DBD 則は学習率の値を増加させ、より大きな更新量で重みを更新することにより収束の速さを改善するように働く。逆に、勾配と平滑化微分が異符号となる場合、1反復する前後での評価関数曲面の形状は大きく変化しており、これは、学習率の値が大きすぎたために重みが振動していると推察できる。この場合、DBD 則は学習率の値を減少させ、重みの更新量を小さくすることにより収束の速さを改善するように働く。

次に、重みの更新則について考える。勾配項 ((2)式)のみを用いた Jacobs 法では、評価関数曲面の谷 (図1(a))において発生する重みの振動を抑制するために非常に多くの反復を要する。これは、勾配項のみの更新則は、重みの更新量 Δw_k から重みの振動の原因となる谷の直角方向成分を除去する項を備えていないことに起因する。そこで、谷底を横切った直後の反復点において、重みの更新量 Δw_k に含まれる谷の直角方向成分を除去するために、次式のように慣性項を導入した Jacobs hybrid 法が提案された⁴⁾。ただし、学習率の更新則には(3), (4)式に示した DBD 則を用いる。

Jacobs hybrid 法

$$\Delta w_k = -\text{diag}(\eta_k) g_k + \alpha \Delta w_{k-1}. \quad (5)$$

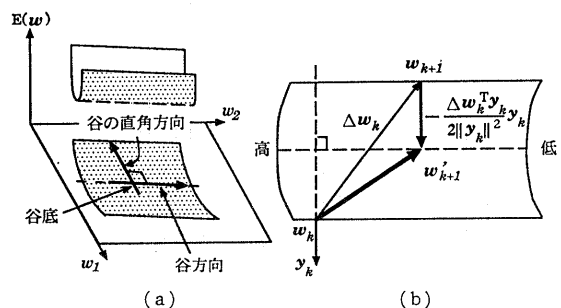


図1 (a)谷の形状, (b)重みの位置の補正方法
Fig. 1 (a) Shape of ravine, (b) Correction method of the weights position.

ここで、 α は慣性率を表し、以下では、慣性項による加速効果が最大となる定数 ($\alpha \approx 1$) として扱う。

しかしながら、慣性項を用いても、谷底を横切ってから2反復後の点において重みの更新量に含まれる谷の直角方向成分を除去することはできない。このため、重みは再び谷底を横切り、重みの振動は再発する⁵⁾。重みの更新量から谷の直角方向成分を根本的に除去するためには、その成分の発生源である勾配の谷の直角方向成分を生じさせないようにする必要がある。このためには、負の勾配方向である最急降下方向が谷方向と一致する谷底上を探索する学習法が有効であると考えられる。そこで、われわれは、重みの位置を谷底付近に補正する方法を Jacobs hybrid 法に導入した Kick Out 法を提案した^{5),6)}。ただし、学習率の更新則には(3)、(4)式に示した DBD 則を用いる。

Kick Out 法

$$w_{k+1} = w_k + \Delta w_k, \quad \begin{cases} \Delta w_k = -\text{diag}(\gamma_k)g_k + \alpha \Delta w_{k-1} & \text{if } y_k^T y_{k-1} > 0, \\ \Delta w_k = -\text{diag}(\gamma_k)g_k + \alpha \Delta w_{k-1} - \frac{y_k^T \Delta w_{k-1}}{2\|y_k\|^2} y_k & \text{if } y_k^T y_{k-1} < 0. \end{cases} \quad (6)$$

Kick Out 法の特徴は、谷において重みが振動する場合、補正項 ((6)式下段の更新則の右辺第3項)の働きにより、重みの位置 (w_{k+1}) が谷底付近 (w_{k+1}) に補正されることであり、これにより、重みの振動は素早く抑制される (図1(b))。この補正法は、振動時に重みが捕捉される特定の部分空間 (二次元鳥かご⁷⁾) から重みを蹴り出すことに相当するため、上記の学習法を Kick Out 法と呼んでいる。

数値実験から、Kick Out 法による収束が最も速く、Jacobs hybrid 法、Jacobs 法の順に収束が遅くなる結果が得られている^{5),6)}。しかしながら、これらの数値実験で示したように、各学習法による最高の収束速度を得るためには、DBD 則に含まれる増減率の値を調整しなければならないという問題がある。次章では、この問題について考察する。以下、Jacobs 法、Jacobs hybrid 法、Kick Out 法と記す場合、おのおのの重みの更新則を指すものとする。

3. Delta-Bar-Delta 則の問題点

DBD 則を用いた場合、重みの更新量は増減率により規定される。このため、増減率の値によっては、

最も速く収束する場合と比較して収束が遅くなるという問題、すなわち、加速効果の低下が生じる。特に、Minai らは、慣性項を含む重みの更新則と DBD 則を併用した場合、加速効果の低下のみならず発散する場合も生じ、勾配項のみの更新則を用いた場合と比較して、加速効果の低下に程度差があることを指摘している⁸⁾。しかしながら、こうした問題の原因等については一切明らかにされていない。本章では、この原因を明らかにするため、DBD 則の加速効果について考察する。この際、DBD 則による加速化は、重みの軌跡上における勾配に基づいて行われるため、以下においても同様に、重みの軌跡に基づいて DBD 則による加速効果を考察する。

DBD 則による加速効果としては、丘における重みの更新量の増大、および谷において発生する重みの振動の抑制により得られる2つの効果が考えられる。そこで、利用する重みの更新則が勾配項のみの場合と慣性項を含む場合では、加速効果の低下に程度差が生じることに着目し、これが丘または谷のどちらにおいて発生するかを考察する。

丘、および谷において2つの重みの更新則による重みの軌跡を比較すると、丘における両者の軌跡は大差はないが、谷においては全く異なる。すなわち、谷において勾配項のみの更新則を用いた場合、重みは1反復ごとに谷底を横切るが、慣性項を含む更新則を用いた場合、重みは2反復ごとに谷底を横切る。このことから、加速効果の低下の程度差は、更新則により重みの軌跡が異なる谷において生じていると推察できる。

さて、DBD 則による谷での加速効果は、重みの振動を抑制することにより得られる。これは、DBD 則が学習率の値を増減することにより、重みの更新量から振動の原因となる谷の直角方向成分が除去されるためである。このことから、DBD 則による振動抑制効果は、学習率の更新方法に大きく依存していることがわかる。そこで、重みの振動方向に関係した学習率の谷の直角方向成分の更新方法に着目し、谷での DBD 則による重みの振動抑制効果に絞って詳細に考察する。

初めに、勾配項のみをもつ Jacobs 法と DBD 則を併用した場合について考察する。谷における重みは1反復ごとに谷底を横切ること、および平滑化微分の定義より、各反復点における平滑化微分、および勾配の谷の直角方向成分は図2に示すようになる。説明の便宜上、谷方向、および谷の直角方向を w_1 , w_2 軸とし

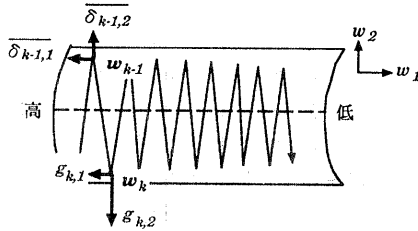


図 2 勾配項のみを用いた場合の重みの軌跡
Fig. 2 Locus of weights by using the algorithm with only a gradient term.

て、その方向を図 2 に示すようにとる。この図より、谷底を横切る前後の反復点における平滑化微分、および勾配の谷の直角方向成分は常に逆符号であり、学習率の更新基準は $\overline{\delta_{k-1,2}} \cdot g_{k,2} < 0$ となることがわかる。したがって、学習率の谷の直角方向成分は減少し続け、重みの振動は次第に抑制される。

しかしながら、重みの更新則に慣性項を含む Jacobs hybrid 法、あるいは、Kick Out 法と DBD 則を併用した場合、谷における重みの振動を抑制できない場合が生じる。慣性項を含む重みの更新則を用いた場合、谷における重みの軌跡は複雑であり解析が困難である。そこで、重みの動きを簡単化するために、Jacobs hybrid 法と DBD 則を併用し、各学習率を同一値に固定した場合を考える。この場合、学習法は Momentum 法と等価であるため、重みの軌跡は図 3 に示すようになる。以下では、この軌跡に基づいて学習率を可変にした場合の DBD 則の振動抑制効果を考察する。

各反復点における勾配の谷の直角方向成分は、図 4 (a) に示す方向を指し、これと平滑化微分の定義 ((4) 式) より、平滑化微分の谷の直角方向成分は図 4 (b) に示す方向を指す。この図より、各反復点における谷の直角方向成分に関する学習率の更新基準は表 1 に示すようになる。

この場合、谷底を横切った直後の反復点 (k 反復目とする) では、学習率の更新基準は $\overline{\delta_{k-1,2}} \cdot g_{k,2} < 0$ で

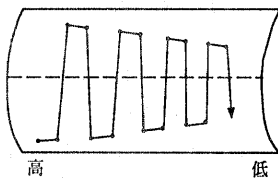


図 3 Momentum 法による重みの軌跡
Fig. 3 Locus of weights with the momentum algorithm.

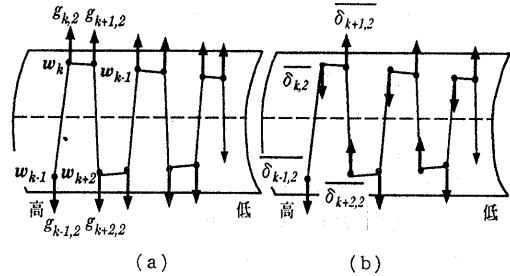


図 4 谷における勾配 (a)、および平滑化微分 (b) の谷の直角方向成分の方向
Fig. 4 Direction of gradients (a) and smoothed derivatives (b) concerning the perpendicular direction in ravine regions.

表 1 DBD 則における学習率の更新基準
Table 1 Criterion for renewal of learning rates in the DBD rule.

反復点	学習率の更新基準
k	$\overline{\delta_{k-1,2}} \cdot g_{k,2} < 0$
$k+1$	$\overline{\delta_{k,2}} \cdot g_{k+1,2} < 0$
$k+2$	$\overline{\delta_{k+1,2}} \cdot g_{k+2,2} < 0$
$k+3$	$\overline{\delta_{k+2,2}} \cdot g_{k+3,2} < 0$

あるため、学習率の谷の直角方向成分は減少する。このため、勾配項 $-\text{diag}(\eta_k)g_k$ と慣性項 $\alpha \Delta w_{k-1}$ は釣り合わず、重みは谷の側面を登る方向に更新される (図 5 (a))。また、 $k+1$ 反復目においても、谷の直角方向成分に関する学習率の更新基準は $\overline{\delta_{k,2}} \cdot g_{k+1,2} < 0$ となり、学習率の谷の直角方向成分はさらに減少する。このため、勾配項の大きさは非常に小さくなり、慣性項と釣り合わなくなる。その結果、重みの振動はさらに大きくなり、抑制されなくなる (図 5 (b))。これを繰り返すと、重みの軌跡は図 6 に示すようにな

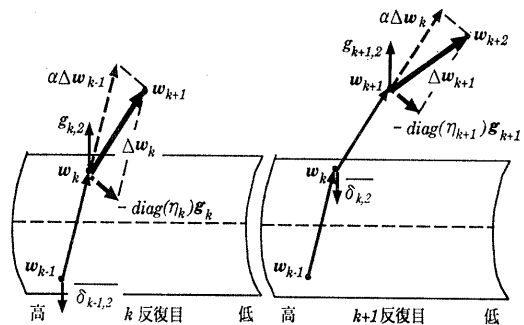


図 5 慣性項、および DBD 則を併用した場合の重みの軌跡
Fig. 5 Locus of weights with the momentum term and the DBD rule.

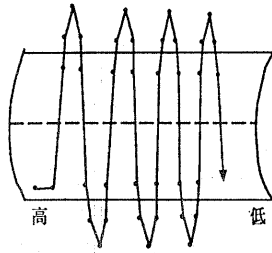


図 6 慣性項, および DBD 則を併用した場合の重みの軌跡
Fig. 6 Locus of weights with the momentum term and the DBD rule.

る。両者が釣り合う場合と同量の距離を谷方向に進むためには、より多くの反復回数を必要とするため、収束の速さは低下する (図 3, 図 6)。

以上の議論から、重みの振動の増大は、谷における DBD 則による学習率の更新方法が不適切であるために生じることが明らかにされた。次章では、こうした問題を解決する方法について検討する。

4. Delta-Bar-Delta-Bar 則の提案

重みの振動を抑制するためには、谷底を横切った直後で勾配項と慣性項を釣り合わせる必要がある。このためには、谷底を横切る際、学習率の谷の直角方向成分を減少させないことが望ましい。しかし、これとは反対に、谷における重みの振動を抑制するためには、学習率の谷の直角方向成分を減少させる必要がある。このような相反する考えを両立させるためには、学習率の谷の直角方向成分を、大域的に見ると減少するが、谷底を横切った直後では減少しないように更新すれば良いと考えられる。

ここで、慣性項を含む重みの更新則を用いた場合の重みの軌跡 (図 3, 図 6) に注目すると、重みは 2 反復以上経た後に谷底を横切る。このことから、重みが谷を横切ってから 2 反復目以降の反復点において学習率の谷の直角方向成分を減少させても、重みの振動を十分に抑制できると考えられる。以下では、こうした考えを具体化した新しい更新則を導出する。

さて、DBD 則では、学習率の更新基準として勾配と平滑化微分の積 ((3)式) を用いている。このため、谷底を横切った直後の反復点において、谷の直角方向成分に関する更新基準は負となり (表 1), 学習率の谷の直角方向成分の減少を招いていた。そこで、学習率の更新基準として、平滑化微分同士の積 ($\overline{\delta_{k-1,i}} \cdot \overline{\delta_{k,i}}$) を用いると、谷底を横切った直後の反復点 (表 2 の *

表 2 DBDB 則における学習率の更新基準
Table 2 Criterion for renewal of learning rates in the DBDB rule.

反復点	学習率の更新基準
$k(*)$	$\overline{\delta_{k-1,2}} \cdot \overline{\delta_{k,2}} > 0$
$k+1$	$\overline{\delta_{k,2}} \cdot \overline{\delta_{k+1,2}} < 0$
$k+2(*)$	$\overline{\delta_{k+1,2}} \cdot \overline{\delta_{k+2,2}} > 0$
$k+3$	$\overline{\delta_{k+2,2}} \cdot \overline{\delta_{k+3,2}} < 0$

印) における更新基準は正となり、学習率の谷の直角方向成分は減少しなくなる (表 2)。

以上のことから、DBD 則 ((3)式) の更新基準として、1 反復する前後での平滑化微分の積 $\overline{\delta_{k-1,i}} \cdot \overline{\delta_{k,i}}$ を用いた新しい更新則: Delta-Bar-Delta-Bar 則 (DBDB 則) を提案する。

Delta-Bar-Delta-Bar 則 (DBDB 則)

$$\begin{cases} \eta_{k,i} = \eta_{k-1,i} + \kappa & \text{if } \overline{\delta_{k-1,i}} \cdot \overline{\delta_{k,i}} > 0, \\ \eta_{k,i} = \eta_{k-1,i} \cdot \phi & \text{if } \overline{\delta_{k-1,i}} \cdot \overline{\delta_{k,i}} < 0, \\ \eta_{k,i} = \eta_{k-1,i} & \text{otherwise.} \end{cases} \quad (7)$$

DBDB 則を用いた場合、谷底を横切る前後での平滑化微分の符号は変化しない。したがって、学習率の谷の直角方向成分は谷を横切った直後には減少せず、1 反復遅れて減少するようになる。このため、谷を横切った直後における勾配項は慣性項と同程度の大きさとなり、両者の釣り合いがとれて重みの振動は抑制される (図 7)。

こうした DBDB 則による加速効果は、平滑化微分に用いる重み付け係数 $\theta \in (0, 1)$ の値に依存する。 $\theta = 0$ とした極端な場合、学習率の更新基準は勾配のみを用いた $g_{k-1,i} \cdot g_{k,i}$ と等価になり、学習率の谷の直角方向成分が適切に更新されず、振動は抑制されなくなる。したがって、DBDB 則による加速効果を得るためには、少なくとも $\theta \geq 0.5$ の範囲で設定することが

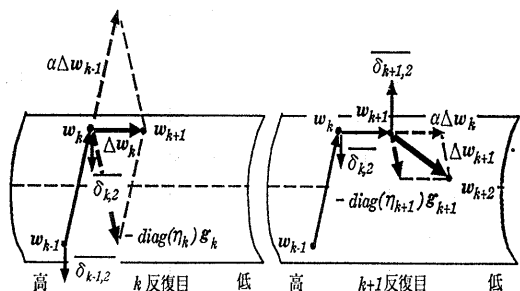


図 7 慣性項, および DBDB 則を併用した場合の重みの軌跡
Fig. 7 Locus of weights with the momentum term and the DBDB rule.

望ましい。

5. 数値実験

提案した DBDB 則の性能を評価するために、Sin 関数近似⁹⁾、およびパリティ問題¹⁾を用いて数値実験を行った。この結果、DBD 則を用いた場合には収束しなかった試行が、DBDB 則を用いると、さまざまな増減率に対して収束し、著しい収束の改善効果が確認された^{10), 11)}。

5.1 Sin 関数近似⁹⁾

まず、DBD 則と慣性項を併用すると加速効果が低下する問題に対し、DBDB 則を用いることにより収束速度が著しく改善されることを示す。加速効果の低下は谷において発生するため、評価関数曲面に谷が存在することで知られる Sin 関数近似を用いて数値実験を行った。

Sin 関数近似は、入力-中間-出力層に1 (線形関数, バイアスなし)-10 (シグモイド関数, バイアスあり)-1 (シグモイド関数, バイアスあり) 個のユニットをもつ NN に、 $0 \leq x < 2\pi$ の範囲を等分した 20 点の入力値に対して Sin 関数: $y = 0.5 + 0.5 \sin(x)$ の値を出力するように学習させる問題である。

前述したように、Jacobs 法に基づく学習法は、重みと学習率に関する2つの更新則から構成される。このため、重みの更新則として、勾配項のみの Jacobs 法⁴⁾、慣性項を含む Jacobs hybrid 法⁴⁾、さらに補正項を付加した Kick Out 法^{9), 6)}を、学習率の更新則として、DBD 則、および DBDB 則を用い、これらを組み合わせた6つの学習法に対して数値実験を行い、各学習法の収束性能を比較した。

数値実験は、 $(-1, 1)$ の範囲の1様乱数を用いて100組の初期値セットを作成し、同じ初期値セットを用いて、おのおのの学習法に対して100回の初期値試行を行い、全パターンに対する残差二乗和が0.01に減少するまで学習させた。また、37,500回反復しても収束しない場合は学習を打ち切った。収束性能を比較するための評価基準として、100回の初期値試行の中から収束した試行に関して収束するまでに要した計算時間の平均値 (CPU time) を用いた。

本実験では、Jacobs 法の収束の速さが増減率 κ , ϕ に依存するという問題に対して、DBDB 則を用いることにより、収束性能が改善される程度を調べることが目的である。そこで、増減率をおのおの $\kappa = (0.005, 0.01, 0.05, 0.1, 0.5)$, $\phi = (0.1, 0.2, \dots, 0.9)$ の範囲で

変化させ、 (κ, ϕ) の45通りの組合せについて数値実験を行った。この際、各学習法で用いた学習率の初期値は $\eta_1 = \dots = \eta_n = 0.5$ 、平滑化微分 δ に用いる重み付け係数は $\theta = 0.7$ と設定した。この結果を図8に示す。図8の各格子点は、100試行中の収束した試行に関する平均計算時間を表しており、Total Ave. は全格子点の平均値、すなわち、さまざまな κ , ϕ に対して収束に要した平均計算時間 (全平均計算時間) を、Min. Time は全格子点の内の最小値、すなわち、最も速く収束した場合の平均計算時間 (最小平均計算時間) を表す。

Jacobs 法は慣性項を用いていないため、DBD 則と慣性項の併用により重みの振動が増大するという問題は起こらない。このため、大部分の κ , ϕ の組合せに対して速く収束している。また、DBDB 則と DBD 則を比較すると、DBDB 則を用いることにより収束が著しく遅くなることはない。次に、慣性項を含む Jacobs hybrid 法、および Kick Out 法に DBD 則を適用した場合、増減率の値によっては収束しなくなる場合 (平均計算時間が最大値を示している部分) が現れる。これらは、谷において勾配項と慣性項の釣合いがとれず、振動が抑制できなくなるため、収束が加速化されなかったと考えられる。このことは、減少率の値が小さい場合 ($\phi = 0.1 \sim 0.5$)、特に顕著に現れている。これに対して、DBDB 則を用いた場合は、ほぼすべての κ , ϕ の組合せにおいて、最も速く収束する場合に近い収束速度で収束していることがわかる。また、全平均計算時間から、DBDB 則により収束の頑健性が改善される様子がわかり、最小平均計算時間からは、Kick Out 法により収束の速さが改善されていることがわかる。

ここで、DBD 則、および DBDB 則を用いた場合、重みの振動が抑制される様子を詳細に比較するため、学習曲線、および重みの軌跡について調べた。初めに、Jacobs hybrid 法、および Kick Out 法に DBD 則、および DBDB 則を適用した場合、初期値から200反復分の学習曲線を図9に示す。この際、重み、学習率の初期値、増減率などは同じ値に設定し、学習法のみを変えて実験した。ただし、増減率は $\kappa = 0.05$, $\phi = 0.4$ に設定した。

図9より、DBD 則では、評価関数値が激しく変化し全く収束していないが、DBDB 則では評価関数値の変化が滑らかになり、速く収束することがわかる。

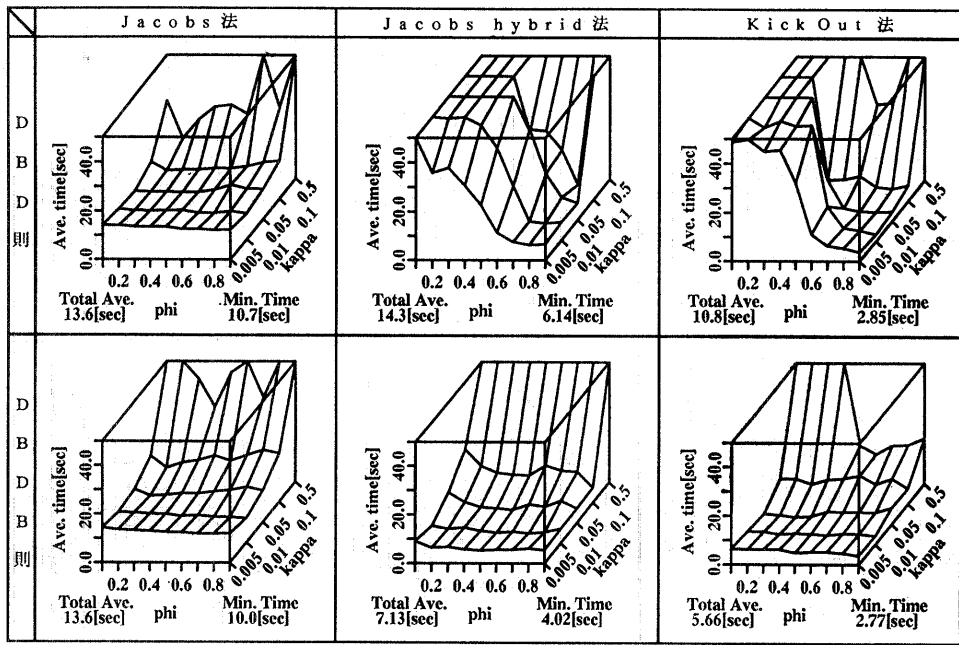


図 8 収束に要した平均計算時間 (Sin 関数近似)

Fig. 8 Average computational time for convergence (Sine function approximation).

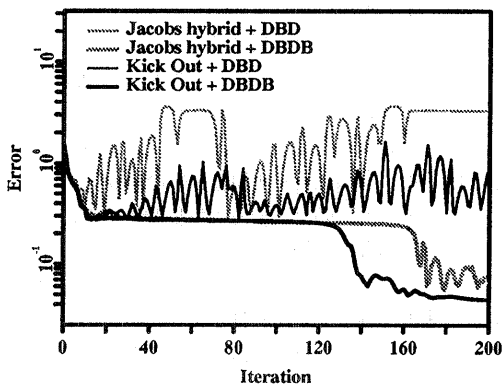


図 9 各学習法による学習曲線 (Sin 関数近似).

Fig. 9 Learning curves for each learning algorithm (Sine function approximation).

次に、重みの振舞いを観察するために、重みの軌跡を初期値の重みの位置でのヘッセ行列の最大固有ベクトル（谷の直角方向）および最小固有ベクトル（谷方向）で張られる超平面上へ射影した結果を図 10 に示す。ただし、最大および最小固有ベクトルは正規化したものを用いた。図中の×印は、初期値の位置であり、反復するに従い、谷方向である横軸に沿って左から右へ、重みの位置が変化する様子を示してい

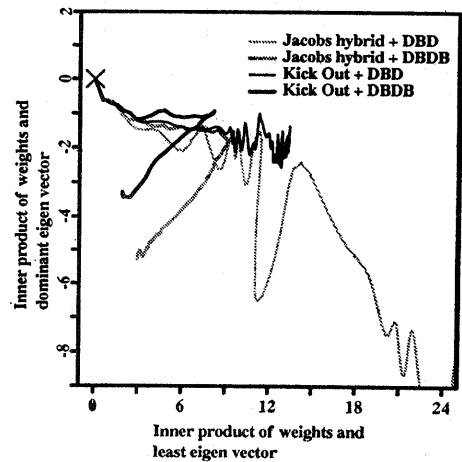


図 10 各学習法による重みの軌跡 (Sin 関数近似).

Fig. 10 Loci for each learning algorithm (Sine function approximation).

る。

図 10 から、DBD 則では、重みが激しく振動している様子がわかる。特に、Jacobs hybrid 法と DBD 則を併用した場合、反復ごとに激しくなり、最終的には重みが発散していると考えられる。これらの軌跡と図 9 の学習曲線から、DBD 則を用いた場合、谷の部

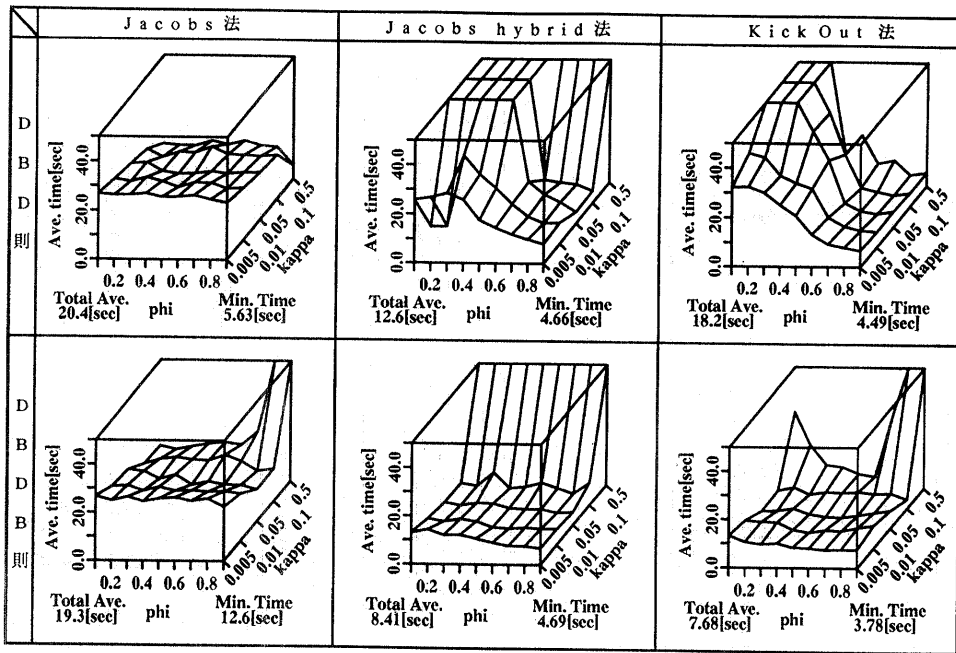


図 11 収束に要した平均計算時間 (5ビットパリティ問題)
 Fig. 11 Average computational time for convergence (5-bit parity problem).

分で発生する重みの振動を抑制できていないと推察できる。これに対して、DBDB 則を用いた場合、重みの軌跡に振動の跡が見られないことから、振動は十分に抑制されていると推察できる。このことは、DBDB 則による学習曲線が滑らかであり、速く収束していることから明らかである。

5.2 5ビットパリティ問題¹⁾

次に、DBDB 則を典型的な二値問題である5ビットパリティ問題に適用し、その性能を評価した。従来、こうした2値パターンを用いた学習課題では、評価関数曲面に非常に緩やかな傾きをもつ丘が存在することが知られており、この部分における勾配の値が小さく、重みがほとんど更新されないため収束が遅くなることが報告されている⁹⁾。しかし、こうした問題においても谷は存在しており、慣性項と DBD 則を併用すると抑制できなかった重みの振動が、DBDB 則を用いることにより抑制され、収束性能が向上することを示す。

ここでは、入力-中間-出力層に5(線形関数, バイアスなし)-10(シグモイド関数, バイアスあり)-1(シグモイド関数, バイアスあり)個のユニットをもつ3層 NN に、5ビットの0または1の値をもつパターンを入力し、そのパターンに1が奇数個ある場合は

1を、偶数個ある場合は0を出力するように学習させた。この際、全パターンに対する残差二乗和が0.002に減少するまで学習させた。また、25,000回反復しても収束しない場合は学習を打ち切った。他の実験条件は、先の Sin 関数近似と同様に設定した。前述の6つの学習法を用い、100回の初期値試行の中から、収束した試行に関して、収束に要した計算時間の平均値を求めた結果を図11に示す。

図11から、パリティ問題においても先の Sin 関数近似の結果と同様に、DBD 則と慣性項を併用した場合、収束していないことがわかる(平均計算時間が最大値を示している部分)。これに対して、DBDB 則では、ほぼすべての増減率の組合せに対して、速く収束していることがわかる。また、全平均計算時間、および、最小平均計算時間から、DBDB 則により収束の頑健性が改善される様子、および、Kick Out 法により収束の速さが改善されている様子が見られる。

次に、重みの振舞いを詳細に調べるために、先の例と同様に、Jacobs hybrid 法および Kick Out 法を用いた場合の学習曲線を図12に示す。また、このときの重みの軌跡を最大固有ベクトルと最小固有ベクトルで張られる超平面上に射影した結果を図13に示す。これらの図から、パリティ問題においても、DBD 則

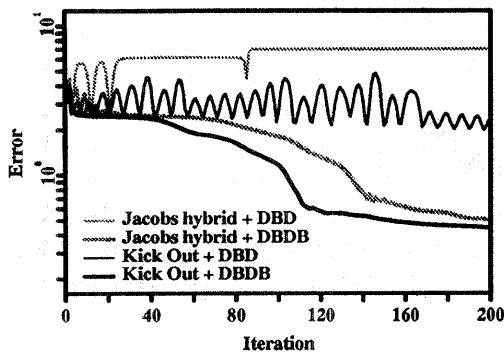


図 12 各学習法による学習曲線 (5 ビットパリティ問題)
Fig. 12 Learning curves for each learning algorithm (5-bit parity problem).

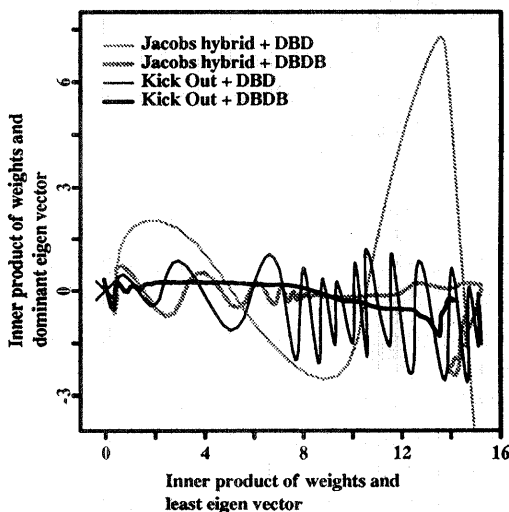


図 13 各学習法による重みの軌跡 (5 ビットパリティ問題)
Fig. 13 Loci for each learning algorithm (5-bit parity problem).

を用いた場合、重みが激しく振動し、収束していないことがわかる。これに対して、DBDB 則を用いると、重みの振動は素早く抑制されており、速く収束していることがわかる。

6. 考察

図 8, 11 に示した数値実験結果から、Jacobs 法を用いた場合、DBDB 則、および DBD 則による加速効果はほぼ同等であることがわかる。このことから、DBDB 則は、慣性項を含まない重みの更新則と併用した場合でも、重みの振動を十分抑制できていると考えられる。これは、DBDB 則による谷での振動抑制

効果を考えると同様である。すなわち、DBDB 則を用いた場合、図 2 の k 反復目の点において、平滑化微分の谷の直角方向成分は勾配の同一成分と同符号であることから、学習率の更新基準は $\overline{\delta_{k-1,2}} \cdot \overline{\delta_{k,2}} < 0$ となり、重みの振動は抑制されることがわかる。同様のことは、重みの更新則に補正項を含む Kick Out 法を用いた場合についても言える。また、DBD 則は、Jacobs 法に対してのみ、さまざまな増減率の組合せにおいて同程度の速さで収束するという頑健性を備えているのに対して、DBDB 則は、さまざまな重みの更新則に対してもこのような頑健性を備えており、実用的な更新則であると言える。

本論文で提案した DBDB 則は、学習過程において評価関数曲面の形状が変化しない一括学習を用いることを前提としている。したがって、NN に学習パターンを提示する度に重みを更新する逐次学習を用いた場合の有効性は明らかではなく、今後の検討課題である。

7. おわりに

本論文では、NN の学習に Jacobs 法を用いる際、学習率の更新則：DBD 則に含まれる増減率の値を調整しなければ、十分な加速効果が得られないという問題について考察した。その結果、DBD 則と慣性項を併用すると、谷の部分で発生する重みの振動を抑制できなくなるため、加速効果が得られないことを明らかにした。さらに、この問題は谷における学習率の更新方法が不適切であるために生じることを指摘し、これを解決した学習率の新しい更新則：DBDB 則を提案した。数値実験により DBDB 則の有効性を調べた結果、増減率を調整しなくても、最も速く収束する場合と同程度の収束速度が得られることを示した。

参考文献

- 1) Rumelhart, D. E. and McClelland, J. L.: *Parallel Distributed Processing*, Vol. I, Chap. 8, MIT Press, Cambridge, MA (1986).
- 2) Sutton, R. S.: Two Problems with Backpropagation and Other Steepest-Descent Learning Procedures for Networks, *Proc. of the 8th Annual Conf. of the Cognitive Science Society*, pp. 823-831 (1986).
- 3) Hush, D. R., Horne, B. and Salas, J. M.: Error Surface for Multilayer Perceptrons, *IEEE Trans. Syst. Man and Cybern.*, Vol. 22, No. 5, pp. 1152-1161 (1992).
- 4) Jacobs, R. A.: Increased Rates of Conver-

- gence through Learning Rate Adaptation, *Neural Networks*, Vol. 1, pp. 295-307 (1988).
- 5) Ochiai, K., Toda, N. and Usui, S.: New Accelerated Learning Algorithm to Reduce the Oscillation of Weights in Multilayered Neural Networks, *IJCNN-92* (Baltimore), Vol. I, pp. 914-919 (1992).
- 6) 落合慶広, 戸田尚宏, 臼井支朗: 重みの振動を抑制する階層型ニューラルネットワークの学習法—Kick Out 法—, *電学論(C)*, Vol. 113-C, No. 12, pp. 1154-1162 (1993).
- 7) Akaike, H.: On a Successive Transformation of Probability Distribution and Its Application to the Analysis of the Optimum Gradient Method, *Ann. of the Institute of Statistic and Mathematics*, Tokyo, Vol. 11, pp. 1-16 (1959).
- 8) Minai, A. A. and Williams, R. D.: Acceleration of Backpropagation through Learning Rate and Momentum Adaptation, *IJCNN-90* (San Diego), Vol. I, pp. 676-679 (1990).
- 9) 浅川和雄, 渡部信雄, 川村 旭, 益岡竜介, 棚橋純一, 山田 博: 階層ニューラルネットワークの機能と学習の高速化, *電学論(C)*, Vol. 110-C, No. 3, pp. 141-146 (1990).
- 10) 落合慶広, 臼井支朗: 新しい学習率の更新則 ($\delta - \delta$ 則) による Kick Out 法の改良, *信学技報*, NC 92-94, pp. 111-118 (1992).
- 11) Ochiai, K. and Usui, S.: Improved Kick out Learning Algorithm with Delta-Bar-Delta-Bar Rule, *ICNN-93* (San Francisco), Vol. I, pp. 269-274 (1993).

(平成5年6月29日受付)

(平成6年2月17日採録)



落合 慶広

1965年生. 1994年豊橋技術科学大学大学院博士課程(システム情報工学)修了. 工学博士. 同年NTTヒューマンインタフェース研究所. 非線形最適化法, ニューラルネットワークの学習に関する研究に従事. 1993年電気関係学会東海支部連合大会奨励賞, 日本神経回路学会第1回研究賞受賞. 電子情報通信学会, 日本神経回路学会各会員.



戸田 尚宏

1959年生. 1990年豊橋技術科学大学大学院博士課程(システム情報工学)修了. 工学博士. 同年同大情報工学系助手. 1992年舞鶴工業高等学校専門学校電気工学科講師, 1994年豊橋技術科学大学情報工学系講師. 生体信号等の非線形, 非定常解析に関する研究に従事. 電子情報通信学会, 計測自動制御学会各会員.



臼井 支朗 (正会員)

1943年生. 1974年カリフォルニア大学大学院博士課程(電気工学計算機科学科)修了. Ph. D. 同年, 名古屋大学工学部助手, 1979年豊橋技術科学大学講師, 助教授を経て, 1986年教授. 生体情報工学, 生理工学などの研究に従事. 著書「生体信号処理の基礎」, ほか. IEEE Fellow. 電子情報通信学会, 電気学会, 計測自動制御学会, 日本ME学会, 日本生理学会, 日本神経科学学会, 日本神経回路学会, INNS 等各会員.