ニューラルネット学習における非線形最適化手法の効果

Takagi, Hideyuki
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd.

Sakaue, Shigeo
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd.

Togawa, Hayato
College of Science and Technology, Nihon University

出版情報：The transactions of the Institute of Electronics, Information and Communication Engineers. D. J74-D2 (4), pp.528-535, 1991-04-25. 電子情報通信学会
バージョン：

論　文

ニューラルネット学習における非線形最適化手法の効果

正員　高木　英行†　　正員　阪上　茂生†　　非会員　戸川　隼人††

Evaluation of Non-linear Optimization Methods for the Learning Algorithm of Artificial Neural Networks

Hideyuki TAKAGI†, Shigeo SAKAUE†, *Members and* Hayato TOGAWA††, *Nonmember*

あらまし　ニューラルネット(NN)学習に非線形最適化手法を組み込み，更に四つの改良を加えてその高速化への寄与を評価した．はじめにbackpropagation学習法の課題を指摘し，この対策として非線形最適化手法の有効性を述べる．次に各非線形最適化手法をNN学習の観点から考察し，Davies，Swann，Campeyの直線探索法とFletcher-Reeves法による共役こう配法を選択し，これら二つの基本手法を組み込んだNN学習アルゴリズムを定式化する．更にこの二つの非線形最適化手法をもとに，NN学習に適した四つの改良手法を提案する．提案する改良手法は，(1)直線探索時のforward計算をメモリの増加と引換えに短縮する方法，(2)直線探索初期の局所最小点へのトラップを避ける方法，(3)並列処理に向いた直線探索方法，(4)こう配方向の切換えによる共役傾斜方法，の4手法である．排他的論理和問題による評価実験を通じて以上の手法の評価を行う．基本的非線形最適化手法はもちろんのこと，本論文で提案する四つの改良手法がいずれも高速化に有用であることを示す．

1. まえがき

ニューラルネットワーク(NN)の研究はここ数年で急速に増え，研究者の増加と相まって応用面と理論面の双方に大きな進展が見られた．理論面に限れば，学習の高速化，なかでもbackpropagation学習アルゴリズム(BP法)の高速化をテーマにした研究が多い．これはBP法がパターン認識/処理に向いたfeed-forward型NN学習アルゴリズムであり，この分野はデータ数が多くモデル規模も大きくなるという応用面からの要求も強いからである．

Rumelhartらが提案した基本BP法[1]は，教師データとNN出力の誤差を小さくする重み係数を，重み係数空間での微分方向に探索するという最急降下法に基づいている．非線形最適化手法の中でもこの最急降下法は最も基本的な手法の一つであり着実に学習を進めるのであるが，代表的な課題が二つある．一つは降下方向を与えるアルゴリズムであって降下量(学習率)には何も関与していないことで，もう一つは局所的な最適降下方向を与えるアルゴリズムであるが大局的な最適降下方向の保証がないことである．この結果，例えば誤差曲面が狭い谷間を形成するような場合は収束が非常に遅くなることが知られている．

この対策には二つ考えられる．一つは誤差を最小にする学習率決定アルゴリズムを導入することと，最急降下方向以外の降下法を検討することである．この観点から，非線形最適化手法あるいは経験則に基づいた各種の手法を組み込んだNN学習アルゴリズムが検討されてきた．

学習率決定方法の代表的な方法は直線探索である．探索する区間幅の設定方法でいろいろな方法があり，R. L. Watrous[2]，E. D. Dahl[3]，瀬戸山ら[4]，高木ら[5]が各様に取り組んでいる．また，降下方向の変化量に着目した経験的あるいは探索的な手法がM. A. Franzini[6]，J. P. Carter[7]，R. A. Jacobs[8]，T. P. Voglら[9]，P. Haffnerら[10]，中村ら[11]から提案されている．

最急降下方向以外の降下方向決定アルゴリズムの代表はニュートン法と共役こう配法である．これらの方

† 松下電器産業株式会社中央研究所，守口市
Central Research Laboratories, Matsushita Electric Industrial Co., Ltd., Moriguchi-shi, 570 Japan
†† 日本大学理工学部数学科，東京都
College of Science and Technology, Nihon University, Tokyo, 101 Japan

法にもいろいろな種類があり，R. L. Watrous[2]，D. B. Parker[12] や A. J. Owens ら[13] は Hessian 行列を解くニュートン法を，E. D. Dahl[3]，瀬戸山ら[4]，高木ら[5]，S. Markram-Ebeid ら[14]，吉田[15] は後者の共役こう配法を各様に検討している．

本論文ではこれら非線形最適化手法を NN 学習の観点から検討し，改良手法を提案し，基本手法と改良手法の効果を評価する．**2.**ではいろいろな種類の非線形最適化手法の中から解探索の安定さと計算コストの観点で BP 法に適した手法を選択し，**3.**で BP 法に組み込んだ基本アルゴリズムを提示する．**4.**では**2.**で選択した基本手法をもとに，NN 学習に向いた四つの改良手法を提案する．最後にこれら手法の高速化への寄与を評価する．

2. 非線形最適化手法の選択

2.1 学習率の動的決定法の選択

一般に誤差曲面の状況は一定ではないので，BP 学習において動的な学習率を用いることの効果は明らかであり，いかに動的に決定するかに多くの研究が報告されている．**1.**でも述べたように，動的決定方法として直線探索法とヒューリスティクスに基づいた方法とがある．しかし後者の場合，発見的探索法の宿命でもあるのだが，常に最適値が設定されている保証がないことや，性能向上のための解析が困難な場合が多いなどの問題がある．従って，ここでは前者の直線探索を評価の検討対象とする．

直線探索には，(a)黄金分割による探索[16]，(b)Fibonacci 探索[16]，(c)M. J. Powell の方法[17]，(d)Davies, Swann, Campey の方法（DSC サーチ）[18]，などがある．このうち，(a)は黄金分割比で，(b)は Fibonacci 数比で，(c)は2次曲線の当てはめで，閉区間内にある単峰性関数の最小点の存在範囲を順次狭くして求める方法である．これらは最小点を閉区間内へ囲い込む方法ではないので，NN 学習のように降下方向のどの範囲に最小点があるかわからない場合は十分広い区間を探索開始点とする必要がある．しかも十分広くしたとしても最小点を囲い込んでいる保証にはならない．一方，DSC サーチははじめに最小点が存在する範囲を探し出し，その後2次曲線の当てはめで最小点を特定する．もちろん DSC サーチの囲い込みアルゴリズムの後で(a)～(c)を行えば問題はない．

収束率や精度の観点からは，(a)，(b)はほとんど同じで，(c)，(d)は(a)，(b)より良いことが証明される[16]．以上より本論文では DSC サーチを動的学習率決定方法の代表として扱う．

2.2 こう配方向探索法の選択

最急降下法以外のこう配決定の代表的な手法はニュートン法と共役こう配法である．前者は誤差曲面の曲率（2次微分）を利用するもので Hessian 行列の計算が中心になる．後者は「前回までに進んでいない方向へ学習を進める」という考えに基づき，連続する m 回の学習方向が互いに共役直交になるよう方向を決めるものである．n 次元重み係数空間上の誤差曲面が2次形式で表せる場合，n 回以内に収束することが保証されていることが強みである．

一般にこの n のオーダは音声や画像の認識/処理に用いる NN の場合数千～数万になる．ところが，ニュートン法の Hessian 行列の演算にはメモリが n^2 必要であり，現実的な計算機資源からすると NN 学習に採用するのは負担が大きい．このことよりこう配方向決定には共役こう配法を採用するものとする．

共役こう配法で非線形問題に使われている手法としては，(a)Fletcher-Reeves 法（FR 法）[19]，(b)Davidon-Fletcher-Powell 法（DFP 法）[20]，(c)Broyden-Fletcher-Goldforb-Shanno 法（BFGS 法）[21]，(d)Powell 法（PARTAN：PARallel TANgent とも言う）[17]，などいろいろバリエーションがある．

このうち(b)，(c)は準ニュートン法系の共役こう配法で Hessian 行列が絡んでくる．そのため計算所要時間およびメモリの点で，大きな NN モデルの学習に適用するには難がある．小さな NN モデルを用いた筆者らの予備実験では(b)と(a)の性能差はほとんど見られなかった．(d)は2次形式で表されたときの最終段階での収束率が大きいと言われている反面，NN のように次元数 n が大きい場合は新しい探索方向の選択が厳しく，良いと思われる方向でも選ばれない場合があると指摘もされている[22]．以上より，本論文では FR 法をこう配方向決定法の代表として扱う．

3. 非線形最適化手法の組込みと改良

3.1 記号の説明

2.で求めた二つの基本的非線形最適化手法を BP 法に組み込む．はじめに，以後で用いる記号を定義する．BP 法で求める i 回目の重み係数修正量 $\Delta\vec{w}$ を慣性項まで含めて定義すると，次式で表される．

$$\Delta\vec{w}_i=\varepsilon\cdot\vec{g}_i+\alpha\cdot\Delta\vec{w}_{i-1} \tag{1}$$

ここで ε は学習率，α は慣性項の定数である．
学習率が ε で重み係数ベクトルが $\vec{w}_i$ のときの，k 番目の学習データに対する NN 出力 $\vec{y}_k$ と教師データ $\vec{t}_k$ との 2 乗誤差を $E(\vec{w}_i, \varepsilon)$ とすれば最急降下方向 $\vec{g}_i$ は

$$\vec{g}_i = -\frac{\partial E(\vec{w}_i, \varepsilon)}{\partial \vec{w}_i} \qquad (2)$$

で与えられる．基本的な BP 法では一般に学習データごとに

$$E(\vec{w}_i, \varepsilon) = \| \vec{y}_k - \vec{t}_k \|^2 \qquad (3)$$

を求めて backward の計算を行うが，m 個の学習データセットの平均誤差

$$E(\vec{w}_i, \varepsilon) = \sum_k^m \| \vec{y}_k - \vec{t}_k \|^2 \qquad (4)$$

に対して backward の計算を行う場合もある．**3.2**，**3.3** の非線形最適化手法はアルゴリズム上は 2 乗誤差の定義を式(3)か式(4)かのいずれかに限定するものではない．しかし式(3)の 2 乗誤差を **3.2** の DSC サーチに用いた場合，学習データごとに適した学習率で大きく重み係数を修正するため，他の学習データの学習には不都合が生じる場合があるので，**4.**での評価では m を全学習データ部とした式(4)の定義を用いた．

3.2 DSC サーチの組込み

従来の経験値に代わって出力層での誤差 $E(\vec{w}, \varepsilon)$ を最小化する学習率 ε を積極的に探し出す方法が直線探索法で，本節では DSC サーチを BP 法に組み込む．DSC サーチは学習率の決定に関する探索手法であるから学習方向には依存しない．式(2)の最急降下法とも **3.3** の共役こう配法とも組み合わせることができる．DSC サーチの手順は

(a) $\varepsilon_{j+1} = \varepsilon_0 + 2^j \cdot h$
と指数的に探索範囲を拡大していき $E(\vec{w}, \varepsilon_{j+1}) > E(\vec{w}, \varepsilon_j)$ になったら(b)に移る（ここでは j は探索回数，h は基本進み幅）

(b) 「$E(\vec{w}, \varepsilon)$ の最小点近傍では 2 次曲線で十分近似可能」の仮定に基づいて，2 次曲線の当てはめによって最小誤差 $E(\vec{w}, \varepsilon)$ を与える ε を求める

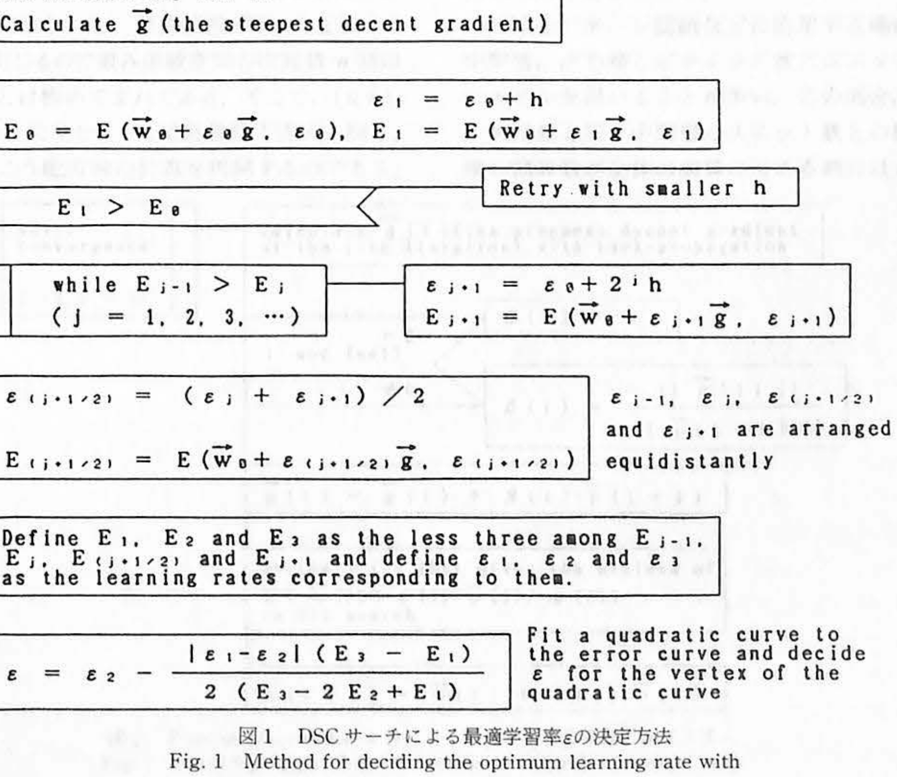


図1 DSC サーチによる最適学習率εの決定方法
Fig. 1 Method for deciding the optimum learning rate with DSC search.

であるが，極小点が必ず存在する閉区間を見つけ出す点に特長がある．DSC サーチアルゴリズムを組み込んだ BP 法を図 1 に示す．メモリは，比較のための 2 乗誤差値と学習率を記憶する 8 個程度が増加するのみである．

更に精度を上げる場合はこうして求まった ε を出発値として小さな進み幅 h で図 1 の DSC サーチを繰り返せばよい．なお，排他的論理和 (EXOR) 問題で実験を行った場合は 1 回の DSC サーチで十分であった．

3.3 FR 法の組込み

FR 法で共役こう配方向を求めるには，

(a) あるこう配方向上の最小点が求められること

(b) その最小点での最急降下方向が陽に求められること

が必要である．一般に行われているような学習率 ε を定数にした学習では，(a)の条件を満足しないが，**3.2** の DSC サーチを導入することで FR 法が可能になる．また，(b)の条件も sigmoid 関数のような微分可能な関数を用いたときの基本 BP 法により陽に求められる．

FR 法を組み込んだ BP 法を図 2 に示す．図中の β は最急降下こう配を共役こう配へ修正する変数である．現実には重み係数空間上の誤差曲面は単純な 2 次形式をしているわけではないし，計算誤差 (特に上記(a)の最小点検出) が生じるので重み係数空間の次元数 n 回以内に収束することは極めてまれである．そこで，$(n+1)$ 回ごとに $\beta(i)$ を 0 にリセットして最急降下方向へ戻し，その点から共役こう配方向の計算を再開するのである．

メモリとしては β 以外に，共役こう配方向ベクトルの現在の値と 1 回前の値を保持する必要があるので，重み係数の個数の 2 倍のメモリ増加が必要である．

3.4 四つの改良

3.2 や **3.3** の基本的手法を更に高速にするために，四つの改良手法を提案する．

［改良 1］ 直線探索法の改良 1　　直線探索では複数の ε_n に対する誤差を求めるため，何度も forward 計算をする必要がある．この計算のうち，メモリの増加と引換えに，第 1 中間層の出力計算の演算を 1 回ですますことができる．sigmoid 関数 (またはしきい値関数) を $f(\ )$ とすれば，u 番目の学習入力データ $\vec{x}^u$ を入力とする第 1 中間層の第 j ユニット出力は

$$f_j(\sum(w_{ij}+\Delta w_{ij})x_i^u)$$
$$=f_j(\sum(w_{ij}+\varepsilon_k\cdot g_{ij})x_i^u)$$
$$=f_j(\sum w_{ij}x_i^u+\varepsilon_k\sum g_{ij}x_i^u)$$
$$\therefore\ =f_j(A_j^u+\varepsilon_k B_j^u)$$

すなわち，直線探索中の $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_k$ の計算のときには，同じ値の A_j^u, B_j^u を用いるので DSC サーチを始める前に 1 回だけ計算しておけばよい．但し，入力ベクトル $\vec{x}^u$ は学習データ数個あるので，A_j^u, B_j^u もそれだけ用意しておく必要がある．

NN をパターン認識などに応用する場合，入力層，中間層，出力層とピラミッド状にユニット数を減らしたモデルを用いることが多い．この場合，入力ベクトル次元数と第 1 中間層のユニット数との積で表される積和演算数が全体の演算に占める割合は大きいので，

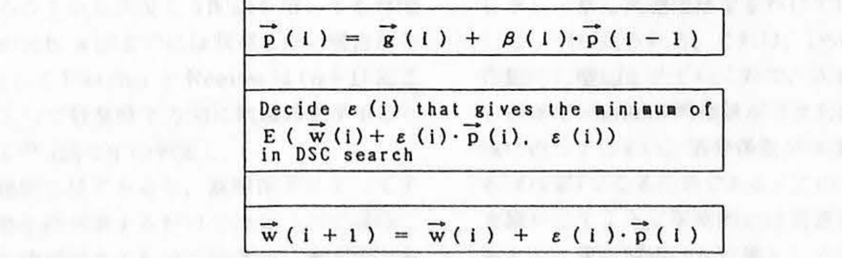


図 2　Fletcher-Reeves 法によるニューラルネットの学習アルゴリズム
Fig. 2 Learning algorithm of artificial neural network with Fletcher-Reeves method.

本方式の効果は大きい．

［改良2］ 直線探索の改良2　　DSC サーチは 3.2 で述べたような逐次計算で探索直線上の最小点を探す．学習率 ε の探索範囲を指数的に拡大していくので ε＝定数の場合に比べれば局所的な最小点を乗り越えやすいが，探索初期や大きな局所的窪みがあるときは，最初の窪みにトラップされやすい．［改良2］はこの問題の解決を目的としている．

基本的な DSC サーチは図1の反復とその処理に示してあるように，$E(\vec{w}, \varepsilon_j)$ と $E(\vec{w}, \varepsilon_{j+1})$ の計算と比較を逐次に行い誤差の局所最小点を探す．これを DSC_{org} と表現しよう．これに対し，$DSC_r(r>3)$ を行うのが第2の改良手法である．すなわち，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r$ を計算した後，$E(\vec{w}, \varepsilon_i)$ $(1 \leqq i \leqq r)$ の中の最小値 $E_{\min}(\vec{w}, \varepsilon_n)$ を与える ε_n の近傍3点を選んで2次曲線の当てはめを行う方法である．r 番目の $E_{\min}(\vec{w}, \varepsilon_r)$ が最小値の場合は更に先に最小値がある可能性が大であるから，次の r 個の $\varepsilon_{r+1}, \varepsilon_{r+2}, \cdots, \varepsilon_{2r}$ について $E(\vec{w}, \varepsilon_i)$ を求め同様に繰り返す．

演算量［DSC_r］≧演算量［DSC_{org}］

であり DSC サーチによっても最小点が得られる保証はないが，最小点に達する確率は

確率［DSC_r］≧確率［DSC_{org}］

であり，ひいては全体の収束が早くなる確率も同様である．

［改良3］ DSC サーチの改良　　並列処理が第3の改良手法である．直線探索の基本形は誤差計算と比較を繰り返す1点ずつの逐次計算であり，並列処理は難しい．しかし，［改良2］の直線探索は r 個の独立演算によって誤差 $E(\vec{w}, \varepsilon_1), E(\vec{w}, \varepsilon_2), \cdots, E(\vec{w}, \varepsilon_r)$ を求めるので，［改良2］と組み合わせることで並列処理が可能になる．

［改良4］ FR 法の改良　　NN のような多次元空間での非線形誤差曲面が2次形式で近似されるのは極小点の近傍付近であろうから共役こう配法を用いても実際には重み空間次元数 n 回までには収束しない場合が多い．この対策として Fletcher と Reeves は $(n+1)$ 回ごとに $\beta(n)=0$ として最急降下方向に軌道修正することを提案している[19]（図2中の判定）．

ところが実験的に見てみると，直線探索によってすらも必ずしも最小点が求まるわけではないので共役こう配方向も十分信用できるわけではない．そこで，ある程度学習が進んだら，局所的には最も信用できる最急降下方向へ早目に軌道修正した方が，共役こう配方向を放棄するデメリットはあるにしても，全体としては安全であるという考えも成り立つ．基本的な FR 法を FR_{n+1} と表すならば，例えば $(n+1)/2$ 回ごとに $\beta(n)=0$ とする $FR_{(n+1)/2}$ が［改良4］である．

4. 評価実験

4.1 実験条件

EXOR 問題に適用し，以上の手法を評価する．NN モデルの各層のユニット数は $(2\times2\times1)$ で，そのほかにしきい値用の定数1のユニットをもつ．教師データは (ON, OFF)＝(0.9, 0.1) とした．

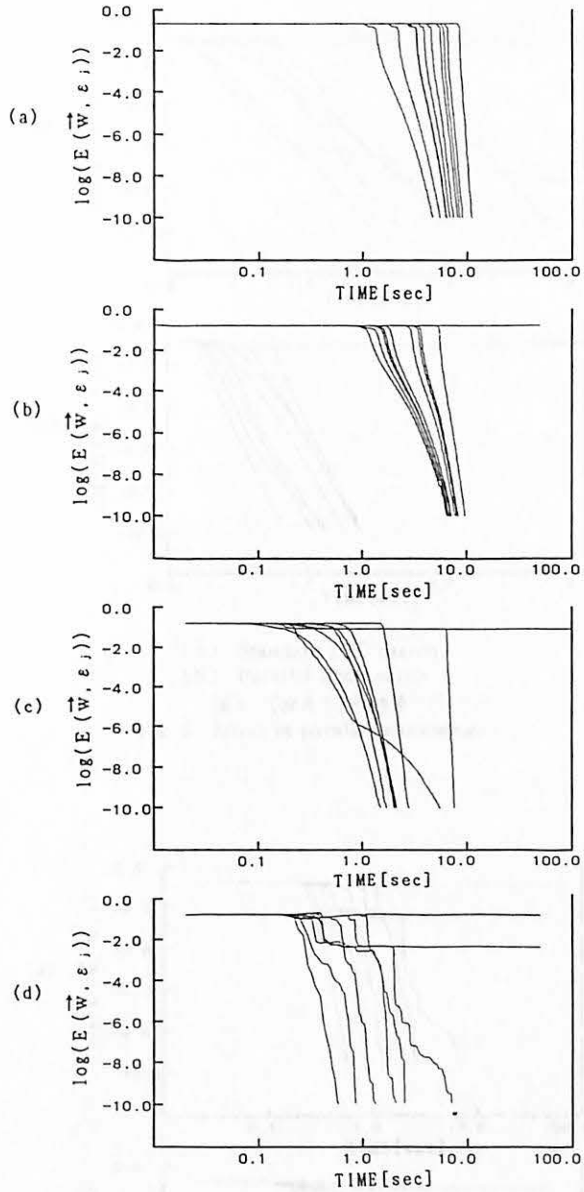
実験は Sequent 社の密結合並列処理計算機 S81 を用いた．但し，［改良3］以外の実験はすべて1 CPU で行った．実行時間は CPU 時間である．NN の学習に占める浮動小数点演算の割合が大きいので，今回の実験結果から他の計算機での実行時間を推定する場合には CPU 以外に浮動小数点演算の性能も考慮する必要がある．今回の実験では Weitech 社のアクセラレータが用いられている．

なお，以下の実験の DSC サーチはすべて［改良1］を加えたものである．また，実験結果図3～6は縦軸が対数2乗誤差値，横軸は CPU 時間である．すべての実験はあらかじめ用意した10種類の乱数値で重み係数を初期化してから学習を始めた．収束曲線が10本見えない場合は収束しなかった誤差曲線が図中で重なっているためである．

4.2 実験結果

4.2.1 非線形最適化手法の効果

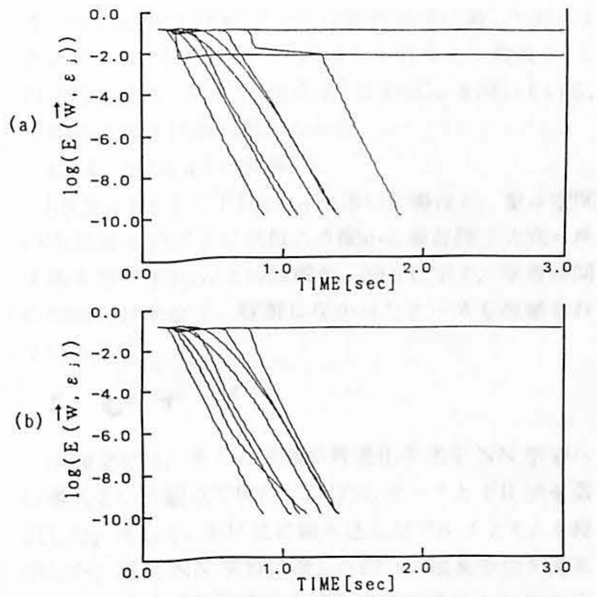
(a)基本 BP 法，(b)BP＋慣性項，(c)BP＋DSC サーチ，(d)BP＋DSC サーチ＋FR 法の結果を図3に示す．DSC サーチは［改良1］を加えている．(a), (b) の BP 法の学習率は予備実験で求めた最適値を用いており固定学習率では最高の性能状態であるが，非線形最適化手法を用いると更に高速化が達成できている．しかし，常に高速化になるわけではなく，まれに収束しない例が見られる．これは，DSC サーチは学習率を指数的に増加させていくので，大きく重み係数を動かし過ぎて sigmoid 特性値が0または1に非常に近い領域に達してしまい，微分係数がほぼ零（すなわち学習量もほぼ零）になるためである．このままでも図3以降の実験が示すように平均的には高速化の効果は明らかであるが，更に積極的な対策として白紙再開基準[5]のような異常状態回避機能を組み込むとよい．DSC サーチの探索中は重み係数を仮に変更するだけで，真の重み



(a) Backpropagation with the fixed learning rate
(b) Backpropagation with the fixed learning rate and the momentum term
(c) Backpropagation with DSC search
(d) Backpropagation with DSC search and FR method

図3　非線形最適化手法の効果
Fig. 3　Effect of the standard non-linear optimization method.

係数は探索終了まで更新されない．従ってたとえ探索途中で異常を検出しても，容易に探索直線上の任意の点に戻ることが可能であり，異常状態回避機能は特にDSC サーチに適している．



(a) Standard DSC search with parallel processing
(b) Improved DSC search avoiding local minima

図4　[改良2]の効果
Fig. 4　Effect of DSC_{10} compared with DSC_{org}.

[改良1]は収束特性は変えないが，モデル規模に応じて一律に学習時間を短縮する．文献(5)の平均収束期待値で評価した場合，入力層ユニットがわずか2個のEXOR 問題でも2割高速になった．

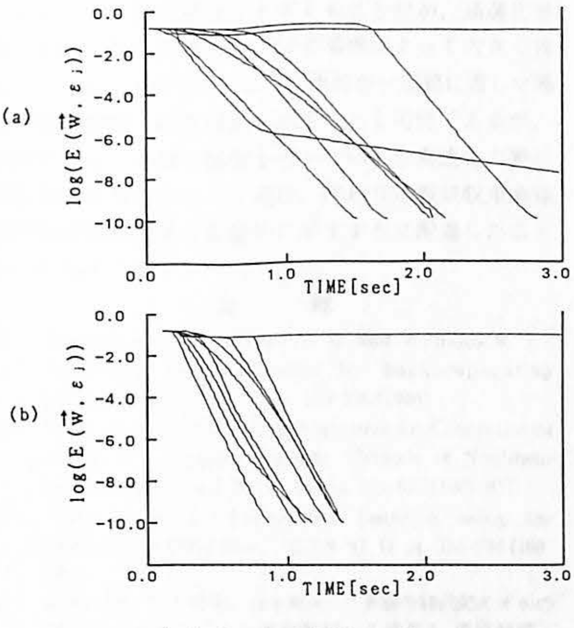
4.2.2 [改良2]の効果

[改良2]ははじめに r の学習率を計算し終えてから探索直線上の極小値を探すので，当然計算時間が増える．単純に基本 DSC サーチの逐次探索との収束時間比較をしては，局所最小値を乗り越えることの効果が評価できない．そこで，[改良1＋改良3]と[改良1＋改良2＋改良3]を用い，直線探索回数・演算量を同じにして比較する．その差は，基本的 DSC サーチの問題点である最初の局所最小値にとらわれやすいか，[改良2]の目的である局所最小値乗越えができるか，の違いである．

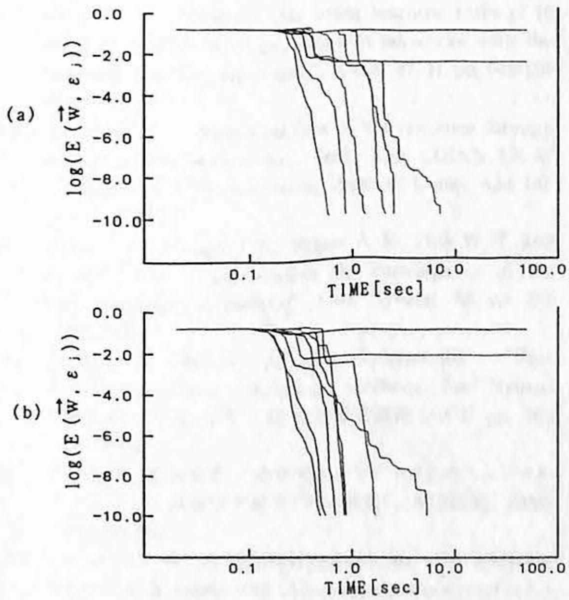
[改良2]として DSC_{10} を用いた場合の結果を図4に示す．10回の実験中2回に顕著な効果が現れており，直線探索上の局所最小点を乗り越えることが全体収束を早めることにつながると言える．

4.2.3 [改良3]の効果

[改良2]は基本形 DSC サーチに比べ，局所最小点を乗り越えることで高速化に寄与していることがわかったが，1回当りの学習時間は多くなる．[改良2]のも



(a) Standard DSC search
(b) Parallel DSC search
図5 [改良3]の効果
Fig. 5 Effect of parallel processing.



(a) Standard FR conjugate gradient method
(b) Improved FR method with earlier switching to the steepest decent gradient.
図6 [改良4]の効果
Fig. 6 Effect of $FR_{(n+1)/2}$ compared wih FR_{n+1}.

う一つの長所はDSCサーチを並列処理に適した形にする点である．[改良1]と[改良1＋改良2＋改良3]との比較を図5に示す．[改良2]はDSC_{10}を用いている．平均的には2倍弱高速になった．

4.2.4 [改良4]の効果

[改良4]として$FR_{(n+1)/2}$を用いた場合と，重み空間の次元数n回ごとに共役こう配から最急降下方向へ戻す基本形[19] FR_{n+1}との比較を，図6に示す．学習時間の短縮だけでなく，収束しなかったケースも改善されている．

5. むすび

本論文では，多くの非線形最適化手法をNN学習への導入という観点で検討し，DSCサーチとFR法を選択した．そして，BP法に組み込んだアルゴリズムを提示した．更にNN学習に適した四つの改良手法を提案し，基本的非線形最適化手法と提案改良法の性能を実験で示した．

評価実験では，従来のBP法は予備実験を通じて最適な学習率を求めておいた．従来法で行う限りこのような最適値は試行錯誤によってしか得られない．たとえ，4.の実験で従来法と非線形最適化手法との性能差がなかったとしても，実際には数倍の試行錯誤の時間がかかっているわけで，一般的に結論づけるならば，非線形最適化手法の導入は効果があると言えよう．

更に四つの提案改良手法の効果を示した．[改良1]は入力層と中間層の結合数に応じて一律に1回当りの学習時間を短縮できる．収束回数には影響を与えないので，4.の実験ではすべてに[改良1]を含めて評価した．[改良2]はDSC_rの直線探索回数rを増やすことで更に性能向上が期待できるので，並列計算機のCPU資源との兼合いを考慮の上，[改良3]と組み合わせてrを大きくとるとよい．[改良4]は本実験では$FR_{(n+1)/2}$を用いたが，この最適回数は問題に依存する．そこでこう配方向の切換えを動的に行えば更に効果が期待できる．例えば，この切換え基準として「共役こう配方向への重み係数の修正量$\Delta\vec{w}$が零になれば，共役こう配方向から最急降下方向へ切り換える」が考えられる．このように本論文で提案した改良手法は，いっそうの性能向上の可能性ももっている．

以上より，非線形最適化手法の選択指針，基本的手法の効果，提案改良手法の効果を示した．NNが実用化技術になり得るかどうかの一つに学習時間の問題があるが，本論文ではこの解決の一助になるものである．

なお，どの評価実験でも言えることだが，高速化手法の性能は評価問題および評価基準によって大きく異なる．例えば文献(5)のように誤差が一定値に達した場合の収束時間・収束回数を示すことも可能であるが，本論文ではこの種の誤解を招かないよう高速化の断片的な定量表現をなるべく避け，代わりに誤差収束曲線で収束傾向の全体を定量的に示すように配慮したことを付け加えておく．

文　　献

(1) Rumelhart D. E., Hinton G. E. and Williams R. J. : "Learning representations by backpropagating errors", Nature, 323-9, pp. 533-536 (1986).

(2) Watrous R. L. : "Learning Algorithm for Connectionist Networks : Applied gradient Methods of Nonlinear Optimization", ICNN '87, **II**, pp. 619-627 (1987-07).

(3) Dahl E. D. : "Accelerated Learning using the Generalized Delta Rule", ICNN '87, **II**, pp. 523-530 (1987-07).

(4) 瀬戸山徹，川人光男，鈴木良次："多層神経回路モデルの新学習アルゴリズムと運動制御への応用"，信学技報，**MBE**88-170 (1989-03).

(5) 高木英行，阪上茂生，戸川隼人："非線形最適化手法を用いたニューラルネットワーク学習アルゴリズムの高速化"，1989 信学春季全大，SD-1-12.

(6) Franzini M. A. : "SPEECH RECOGNITION WITH BACK PROPAGATION", 9th Annual Conf. IEEE Engn. Medic. Biol. Soc., pp. 1702-1703 (1987-11).

(7) Carter J. P. : "Successfully using learning rates of 10 (and greater) in back-propagation networks with the heuristic learning algorithm", ICNN '87, **II**, pp. 645-651 (1987-07).

(8) Jacobs R. A. : "Increased rate of convergence through learning rate adaptation", Tech. Rep. COINS TR 87-111, Univ. of Massatusetts, Dept. of Comp. and Inf. Scien. (1987).

(9) Vogl T. P., Mangis J. K., Rigler A. K., Zink W. T. and Alkon D. I. : "Accelerating the convergence of the back-propagation method", Biol. Cybern., **59**, pp. 257-263 (1988).

(10) Haffner P., Weibel A. and Shikano K. : "Fast Back-Propagation Learning Methods for Neural Network in Speech", 昭 63 春季音講論 2-P-1, pp. 203-204 (1988-10).

(11) 中村雅己，鹿野清宏："英文テキストデータからニューラルネットによる単語列予測モデルの検討"，信学技報，**SP88**-26 (1988-06).

(12) Parker D. B. : "Optimal algorithms for adaptive networks : Second order backpropagation second order direct propagation, and second order hebbian learning", ICNN '87, **II**, pp. 593-600 (1987-07).

(13) Owens A. J. and Filkin D. L. : "Efficient Training of the Back Propagation Network by Solving a System of Stiff Ordinary Differential Equations", IJCNN '89, **II**, pp. 381-386 (1989-06).

(14) Markram-Ebeid S., Sirat J. -A. and Viala J. -R. : "A Rationalized Error Back-Propagation Learning Algorithm", IJCNN '89, **II**, pp. 373-380 (1989-06).

(15) 吉田利信："共役勾配法を用いた多層神経回路網の高速学習法の評価"，信学技報，**NC89**-34 (1989-12).

(16) Kowalik J. and Osborne M. R. (山本，小山共訳)："非線形最適化問題"，培風館 (1970).

(17) Powell M. J. : "An Efficient Method of Finding the Minimum of a Function of Several variables without Calculationg Derivatives", Computer J., **7**, p. 155 (1964).

(18) Swann W. H. : "Report on the Development of a New Direct Search Method of Optimization", I. C. I. Ltd, Centr. Inst. Lab. Research Note 64/3 (1964).

(19) Fletcher R. and Reeves C. M. : "Function Minimization by Conjugate Gradients", Compt. J., **7**, 149 (1964).

(20) Fletcher R. and Powell M. J. D. : "A rapidly convergent decent method for minimization", Compt. J., **6**, pp. 163-168 (1963).

(21) Broyden C. G. : "The convergence of a class double-rank minimization algorithms", J. Inst. Mathes. Applics., **6**, pp. 76-90, pp. 222-231 (1970).

(22) Fletcher R. : "Function Minimization without Evaluating Derivatives—A Review", Compt. J., **8**, p. 33 (1965).

(平成 2 年 9 月 13 日受付，11 月 8 日再受付)

高木　英行

昭 56 九州芸工大修士課程了．同年松下電器産業株式会社入社．同社中央研究所において，C コンパイラ・音声 WP の開発を経て，現在，音声認識・聴覚モデル・神経回路網・ファジー理論等の研究に従事．昭 63 本会篠原記念学術奨励賞受賞．日本音響学会，日本ファジィ学会，神経回路学会，INNS 各会員．

阪上　茂生

昭 60 京大修士課程了．同年松下電器産業株式会社入社．同社中央研究所において，光ディスクの信号処理部の開発を経て，現在，画像処理，神経回路網等の研究に従事．

戸川　隼人

昭 33 早大第 1 理工・数学卒．同年科学技術庁航空宇宙技術研究所入所．昭 47 京都産業大学教授，現在，日本大学教授．理博．数値解析，計算機科学の研究に従事．情報処理学会，日本数学会，応用統計学会，IEEE 等各会員．