

複写可

修士論文

SpikePropにおける学習を阻害する要因  
に関する研究



平成20年度修了  
三重大学大学院工学研究科  
博士前期課程 電気電子工学専攻

藤田 優

# 目次

第1章	はじめに	1
第2章	SpikeProp	3
2.1	SpikePropの概要	3
2.1.1	スパイクングニューロン	3
2.1.2	ネットワーク構造	4
2.1.3	学習法	5
2.2	学習法の改良のための従来手法	7
2.2.1	最急降下法	7
2.2.2	共役勾配法	8
2.2.3	探索方向の極小値の求め方	9
第3章	学習を阻害する要因に関する考察	14
3.1	非線形最適化手法を用いた学習と学習を阻害する要因の調査	14
3.2	誤差曲面の調査	15
第4章	実装方法が学習に与える影響	17
4.1	Bhoteらの実装方法における非線形最適化手法の効果	17
4.2	Bohteらの実装方法での誤差曲面の調査結果	19
4.3	誤差曲面に凹凸ができる原因	19
4.4	課題	20
第5章	SpikePropのネットワークモデルが学習に与える影響	28
5.1	出力時刻を正確に計算した場合での非線形最適化手法の効果	28
5.2	極小解が発生する原因の考察	29
5.3	内部電位の変化のしかたが学習に与える影響の考察	29

5.4	内部電位に山ができる原因 . . . . .	30
5.5	課題 . . . . .	31
<b>第6章</b>	<b>まとめ</b>	<b>37</b>
	<b>謝辞</b>	<b>38</b>
	<b>参考文献</b>	<b>39</b>
	<b>発表論文</b>	<b>40</b>

# 第1章 はじめに

近年，脳の情報処理機構を模倣したスパイクングニューラルネットワークの研究が盛んにおこなわれている [1][2][3]．脳ではニューロン間の情報伝達にスパイクというパルス状の信号を用いている．スパイクングニューラルネットワークは，ネットワーク内部の情報表現に，脳と同じようにスパイク (パルス) を用いるニューラルネットワークである．

このニューラルネットワークの大きな特徴として，時系列処理を行うことができることが挙げられる．そのため，ネットワークの入力に音声信号を用いて到達する音声信号のずれから音源の位置の特定を行う応用研究 [4] や，入力にセンサー信号を用いて自律ロボットのコントローラを構築する応用研究 [5] が行われている．特に [5] では従来のシグモイド型ニューラルネットワークに比べて，スパイクングニューラルネットワークで構成されたコントローラのほうが構築しやすく，性能が優れていることを明らかにしている．また，スパイクングニューラルネットワークの学習則としては，STDP(スパイクタイミング依存可塑性) 学習則 [1] や遺伝的アルゴリズム [5] が用いられることが多い．しかし，STDP についてはネットワークを構築するのに多数のニューロンが必要であるという欠点があり，遺伝的アルゴリズムについては計算量が多くなるという欠点がある．そこで，別のアプローチとして従来のシグモイド型ニューラルネットワークの学習則の一種である誤差逆伝搬 [6] をスパイクングニューラルネットワークにも適用する研究が行われている [7]．

この誤差逆伝搬をスパイクングニューラルネットワークに用いる研究の成果として，SpikeProp[8][9] と呼ばれるモデルがある．これは，スパイクのタイミングにより情報を表現し，特殊なユニット間の結合を用いることで学習を容易にしたモデルである．また学習には勾配を用いた降下法を用いている．なお，一般に勾配を用いた降下法は学習が収束するまで時間がかかるという欠点があり [10]，SpikeProp

でも学習に時間がかかる。そこで、この欠点を解消するために、従来のシグモイド型ニューラルネットワークで用いられてきた非線形最適化手法を用いた学習が SpikeProp でも検討されている [11][12] が、目立った成果は見られない。そこで筆者が実際に非線形最適化手法を用いて学習を行ってみたところうまく学習ができないことがわかった。本研究では、非線形最適化手法を用いた学習がなぜうまく働かない原因について調査する。

以下に本論文の構成を示す。2章では研究の対象としている SpikeProp と学習法の改良のための従来手法について述べる。3章では学習を阻害する要因に関して考察し、4章では実装方法が学習に与える影響、5章では SpikeProp のネットワークが学習に与える影響について議論する。最後に6章で本研究のまとめをする。

## 第2章 SpikeProp

本章では SpikeProp の概要と，その実装方法，およびその学習の高速化手法について説明する．

### 2.1 SpikeProp の概要

SpikeProp はユニットにスパイクニューロンを用いた階層型ニューラルネットワークである．まず 2.1.1 節でユニットに用いるスパイクニューロンについて，2.1.2 節で SpikeProp のネットワーク構造について，2.1.3 節で SpikeProp の学習法について説明する．

#### 2.1.1 スパイクニューロン

スパイクニューロン（以下，ニューロン）とは脳のニューロンをモデル化したものである．ニューロン  $o$  がシナプス前ニューロン  $h$  からのスパイクを  $d$  だけ時間遅れしたうえで受け取る場合を考える．このとき，ニューロン  $o$  の内部電位  $x_o(t)$  は式 (2.1) で定義されるスパイク応答関数  $\varepsilon(t)$  (図 2.1) に結合荷重  $w_{ho}$  を乗じた分だけ変化する．この関係は式 (2.2) で表される．

$$\varepsilon(t) = \begin{cases} 0 & (t < 0) \\ \frac{t}{\tau} e^{1-t/\tau} & (t \geq 0) \end{cases} \quad (2.1)$$

$$x_o(t) = \sum_{h \in \Gamma_o} w_{ho} \varepsilon(t - t_h - d) \quad (2.2)$$

ここで  $\Gamma_o$  はニューロン  $o$  の前ニューロン  $h$  の集合， $t_h$  はシナプス前ニューロン  $h$  の発火時間である．そして内部電位  $x_o(t)$  が最初に発火しきい値  $\theta$  を超えたときにニューロンは発火し，この時刻を出力時刻とする．

## 2.1.2 ネットワーク構造

SpikeProp では, 図 2.2 に示すようなニューロン間に特殊な結合を持つ階層型ネットワークを使用する. この特殊な結合は, 1 本の主結合につき  $m$  本の副結合により構成されている. この副結合それぞれに決まった時間遅れ  $d^k$  と結合荷重  $w_{ho}^k$  が定められている. 副結合の時間遅れ  $d^k$  には  $1, 2, 3, \dots, m$  のような一定間隔の定数を用い, すべての主結合において同じ値を用いる. こうすることによって, 結合荷重を調整するだけで出力時刻の調整ができるようになり学習を容易に行うことができる. ニューロン  $o$  の内部電位  $x_o(t)$  は式 (2.3) となる.

$$x_o(t) = \sum_{h \in \Gamma_o} \sum_{k=1}^m w_{ho}^k \varepsilon(t - t_h - d^k) \quad (2.3)$$

実際の内部電位の変化は図 2.3 のようになる.

### 2.1.3 学習法

学習の目的は、入力層  $I$  に属する入力ニューロン  $i \in I$  に入力パターンの集合  $\{P[t_1 \dots t_i]\}$  を与えたときに出力層  $O$  に属する出力ニューロン  $o \in O$  での目的の出力時刻  $\{t_o^*\}$  を出力するような結合荷重  $\mathbf{w} = \{w_1 \dots\}$  を見つけることである。しかし、方程式を解いて最適な結合荷重を求めるのは、ニューラルネットワークの入出力の関係が非線形なため、困難である。一般には、反復計算によって最適な結合荷重を求めることが多い。SpikeProp では、この反復計算の方法として勾配を用いた降下法を用いる。勾配を用いた降下法では誤差関数を定義し、その誤差関数の値が減るように結合荷重を更新する。この誤差関数には平均二乗誤差関数を用いる。希望する出力時刻を  $t_o^*$ 、実際の出力時刻を  $t_o$  とすると、誤差関数は次式で定義される。

$$E = \frac{1}{2} \sum_{o \in O} (t_o - t_o^*)^2 \quad (2.4)$$

そして誤差関数  $E$  の値が減るように結合荷重を更新する方法として勾配を用いた降下法を用いる。これは勾配の方向と逆向き（最急降下方向）に少しずつ進むことを繰り返すことで誤差を最小化することをめざしている。Bhote らは誤差関数の勾配の計算式を導出する際に、結合荷重を変更したことによって変化した内部電位の変化量と、それによる出力時刻の変化量は比例するという前提を用いている。模式図を 2.4 に示す。

勾配を用いた降下法での 1 回の学習における結合荷重  $w$  の変更量  $\Delta w$  と更新式は以下の式となる。

$$\Delta \mathbf{w} = -\eta \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} \quad (2.5)$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

ここで  $\eta$  は学習率であり定数である。中間層-出力層間の結合荷重  $w_{ho}^k$  と入力層-中間層間の結合荷重  $w_{ih}^k$  について式 (2.5) で変更量を計算するとそれぞれ式 (2.6)、式 (2.7) となる。



$$\begin{aligned}\Delta w_{ho}^k &= -\eta \frac{\partial E}{\partial w_{ho}^k} \\ &= -\eta y_h^k(t_o) \delta_o\end{aligned}\tag{2.6}$$

$$\begin{aligned}\Delta w_{ih}^k &= -\eta \frac{\partial E}{\partial w_{ih}^k} \\ &= -\eta y_i^k(t_h) \delta_h\end{aligned}\tag{2.7}$$

ただし、式 (2.6) , 式 (2.7) 中の  $\delta_o$  ,  $\delta_h$  はそれぞれ式 (2.8) , 式 (2.9) である.

$$\delta_o = \frac{-(t_o - t_o^*)}{\sum_{h \in \Gamma_o} \sum_{k=1}^m w_{ho}^k \left\{ \frac{\partial \{y_h^k(t)\}}{\partial t} (t_o) \right\}}\tag{2.8}$$

$$\delta_h = \frac{\sum_{o \in \Gamma^h} \left\{ \delta_o \sum_{k=1}^m \left\{ w_{ho}^k \frac{\partial \{y_h^k(t)\}}{\partial t} (t_o) \right\} \right\}}{\sum_{i \in \Gamma_h} \sum_{k=1}^m \left\{ w_{ih}^k \frac{\partial \{y_i^k(t)\}}{\partial t} (t_h) \right\}}\tag{2.9}$$

この勾配を用いた降下法を用いた学習の学習過程は図 2.5 のようになる. この方法では、徐々に結合荷重を更新するために、学習が収束するまで時間がかかるという欠点がある.

## 2.2 学習法の改良のための従来手法

前節では、勾配を用いた降下法は学習が収束するまで時間がかかるという欠点があると説明した。従来のシグモイド関数型ニューラルネットワークに適用した場合にも同様の欠点があられる。この欠点を解消するために、非線形最適化手法がよく用いられている [10][13][14]。以下で各方法を説明する。

### 2.2.1 最急降下法

最急降下法は、最急降下方向において誤差が極小値となるように結合荷重を更新することを繰り返す方法である。以下で詳しく説明する。

まず初期値  $w_0$  を与え、誤差関数の最急降下方向  $-\nabla E$  の直線上で誤差関数の値が極小になる点に進む。その点で再び  $-\nabla E$  を計算し、またその直線上での極小値を探索する...ということを収束するまで繰り返す方法である。学習における更新式は以下のようなようになる

$$\Delta w = -\alpha \nabla E(w) \quad (2.10)$$

$$w \leftarrow w + \Delta w \quad (2.11)$$

ここで  $\alpha$  は最急降下方向で誤差関数の値が極小になるように選ぶ。この方法は単純で効果的な方法だが、以下のような欠点がある [10]。

- 最小値以外の極小値がある場合、初期値の与え方によっては最小値でない極値に到達してしまう可能性がある。
- 極小値が多数ある場合には極小値に陥る可能性が極めて高くなる。
- たとえ極値が一つしかないとしても、その極値が細長い谷の底にあるときには、なかなか極値に収束しない。

## 2.2.2 共役勾配法

共役勾配法は、最急降下法と同様に直線方向の最小値探索を繰り返す方法である。最急降下法では最急降下方向への探索を行っていたが、共役勾配法では共役勾配の方向に探索を行う。この方法は、従来のシグモイド型ニューラルネットワークでの学習の高速化に効果をあらわしている [13]。共役勾配法では以下の式で結合荷重の更新を行う。

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w} \quad (2.12)$$

$$\begin{aligned} \Delta \mathbf{w} &= \alpha(t) \left( -\nabla E(t) + \beta(t) \frac{\Delta \mathbf{w}(t-1)}{\alpha(t-1)} \right) \\ &= \alpha(t) \mathbf{d}(t) \end{aligned} \quad (2.13)$$

ここで  $-\nabla E(t)$  は各学習パターンの二乗誤差に対する最急降下方向の和である。 $\alpha(t)$  の値は共役勾配方向で誤差関数が極小になるように選ぶ。 $\beta(t)$  は以下の Polak-Ribiere-Polyak の公式で計算する。

$$\beta(t)_{\text{PRP}} = \frac{\nabla^T E(t) \{ \nabla E(t) - \nabla E(t-1) \}}{\| \nabla E(t-1) \|^2} \quad (2.14)$$

### 2.2.3 探索方向の極小値の求め方

最急降下法・共役勾配法では，探索方向  $\mathbf{d}$  の極小値，つまり最適な  $\alpha$  をどのように算出するのが重要になる．一般にこの値は探索により求めることが多い．その代表的な手法の一つである直線探索法の手順を以下に示す．

1.  $\mathbf{w}$  の初期値を与え，探索のステップ幅を  $\alpha \leftarrow \alpha_0$  とする．

2. 次のように置く．

$$\mathbf{W} \leftarrow \mathbf{w}, \quad \mathbf{W}' \leftarrow \mathbf{w} + \alpha \mathbf{d}$$

3. もし  $E(\mathbf{W}) > E(\mathbf{W}')$  であれば次の計算を行う．

(a)  $E(\mathbf{W}) \leq E(\mathbf{W}')$  となるまで次の計算を繰り返す．

$$\alpha \leftarrow 2\alpha, \quad \mathbf{W} \leftarrow \mathbf{W}', \quad \mathbf{W}' \leftarrow \mathbf{W} + \alpha \mathbf{d}$$

(b)  $\mathbf{w} \leftarrow \mathbf{W}$ ,  $\alpha \leftarrow \alpha/2$  と置く．

4. そうでなければ次の計算を行う．

(a)  $E(\mathbf{W}) \geq E(\mathbf{W}')$  となるまで次の計算を繰り返す．

$$\alpha \leftarrow \frac{\alpha}{2}, \quad \mathbf{W}' \leftarrow \mathbf{W}' - \alpha \mathbf{d}$$

(b)  $\mathbf{w} \leftarrow \mathbf{W}'$ ,  $\alpha \leftarrow 2\alpha$  と置く．

5. ステップ2に戻り，これを  $\alpha \leq \varepsilon$  となるまで繰り返す．ここで  $\varepsilon$  はあらかじめ設定しておいた探索の終了条件である．

6. 得られた  $\alpha$  を返す．

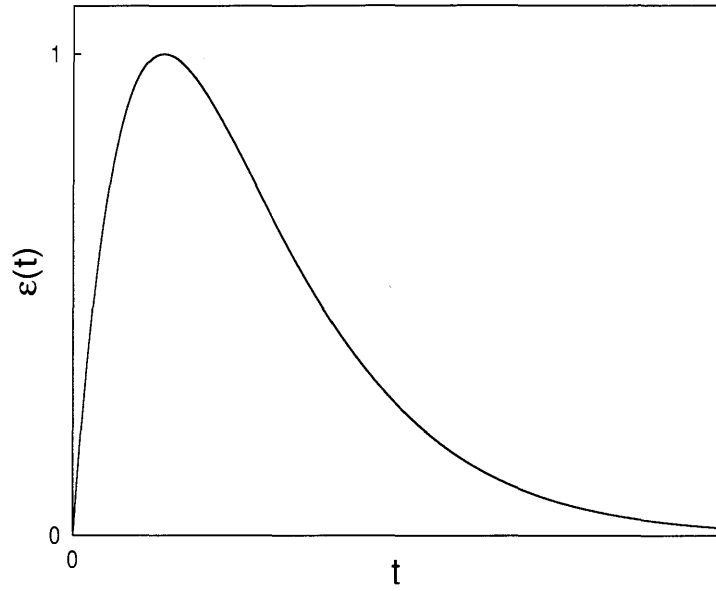


図 2.1: Spike Response Function

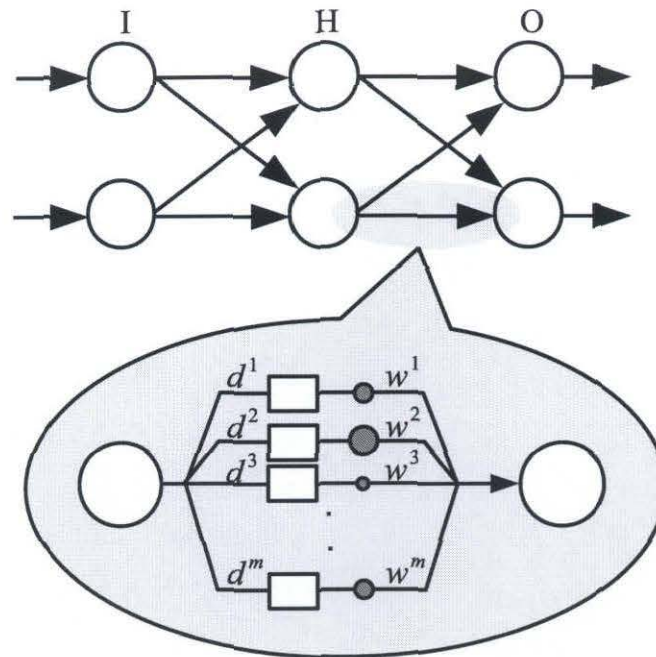


図 2.2: SpikeProp のネットワーク構造

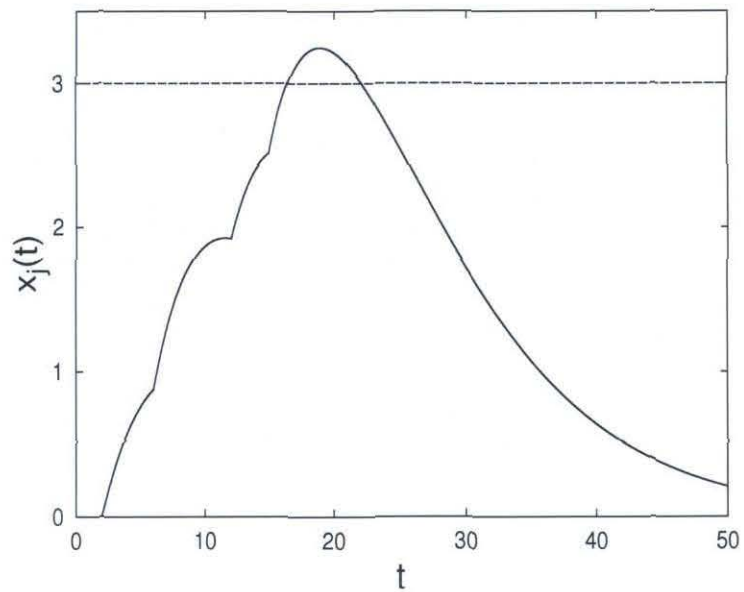


図 2.3: ニューロンの内部電位の変化

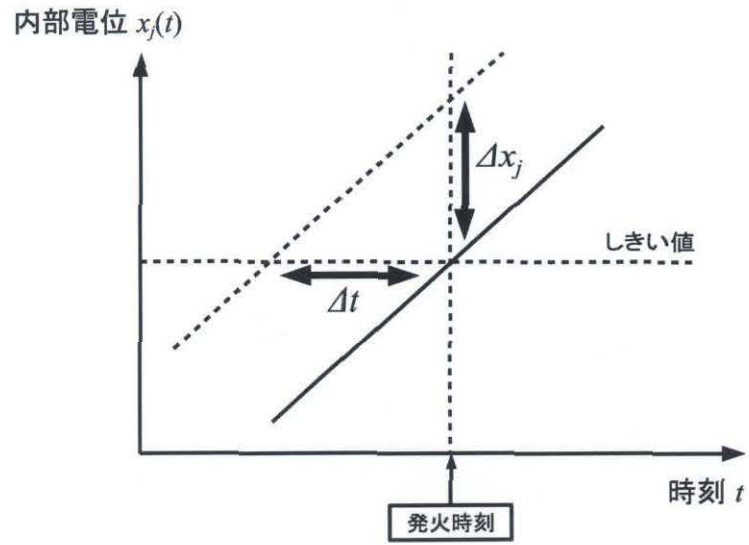


図 2.4: Bohte らの想定

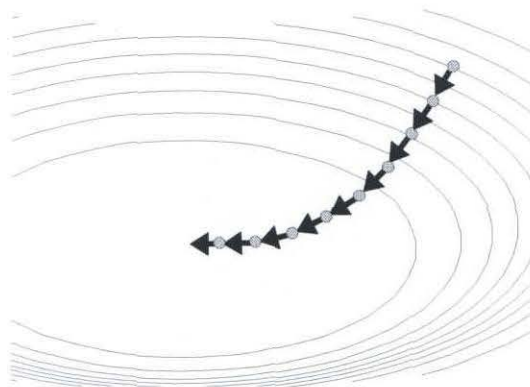


図 2.5: 勾配を用いた降下法

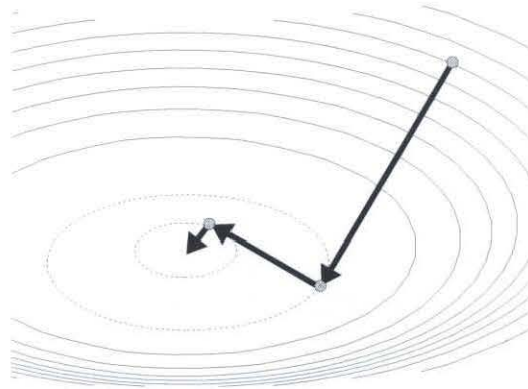


图 2.6: 最急降下法



# 第3章 学習を阻害する要因に関する 考察

2章では SpikeProp と学習法の改良について説明した。しかし、SpikeProp に非線形最適化手法を用いた場合、学習に成功する可能性が極めて小さくなる。本章では、SpikeProp の学習を阻害する要因について検討する。

## 3.1 非線形最適化手法を用いた学習と学習を阻害する要因の調査

非線形最適化手法を用いた学習は、極小解があると極小解に陥りやすいという欠点がある。従来のシグモイド型ニューラルネットワークにおいて非線形最適化手法を用いた場合、学習ができないというほど極小解が学習に与える影響はなかった。では SpikeProp ではどうだろうか。そこで SpikeProp で非線形最適化手法（共役勾配法）を学習に用いてみたところ、シグモイド型ニューラルネットワークと比べ学習に成功する場合が少なくなることがわかった。その学習の過程を分析することで、共役勾配法の更新ステップ  $\alpha$  がきわめて小さく加速の効果があらわれず、すぐに極小解で学習が終了していることがわかった。

ここでまず着目したのが Bohte らの実装方法である。この方法では出力時刻を正確に計算していないため、これにより何かしらの影響があるのではと考えた。このことについては4章で詳しく説明する。また、この不具合を解消した状態で、再び非線形最適化手法を用いて学習を行ってみたところ、序盤は一回のステップで大幅に誤差を減らすことができたが、学習終盤で極小解にはまり学習に失敗してしまうことがわかった。この極小解の発生原因については5章で詳しく説明する。またこれらの学習を阻害する要因について調査するのに誤差局面を用いた。この

誤差局面を用いた調査については次節で詳しく説明する。

## 3.2 誤差曲面の調査

極小解がどのように存在しているのか調査するのに、誤差曲面がよく用いられる [15][16]. 式 2.4 より、誤差関数  $E$  は、出力時刻  $t_j$  を変数とした関数である。また、式 2.2 と内部電位  $x_j(t)$  が発火しきい値  $\theta$  を最初に越えた時刻が出力時刻であることより、出力時刻  $t_j$  は結合荷重  $\mathbf{w}$  を変数とした関数である。したがって誤差関数  $E$  は結合荷重  $\mathbf{w}$  を変数とする関数である。この誤差関数  $E(\mathbf{w})$ 、各結合荷重  $w_{ij}$  の値を軸とした多次元の空間内の超曲面が誤差曲面である。

しかし、多次元の超曲面のままでは解析が難しいので、本研究では、結合荷重の変化を探索方向のみに固定し、2次元で表す (図 3.1)。これは探索方向に結合加重を変化させたときの誤差を表すことになるので、学習はこの誤差曲面上の極小点のうち探索の許容誤差を下回るものを探索することになる。そのため誤差曲面の形状は学習の成否に大きく関係する。例えば、探索方向の誤差曲面が図 3.2 のような形状をしていた場合で直線探索を行うと、極小解に陥り、最小解を見つけることが出来ない。このように誤差曲面を調査することは学習に与える影響を明らかにできるので重要である。そこで、本研究では誤差曲面を調査することによって、SpikeProp のネットワークとしての性質が学習にどのような影響を与えるのか明らかにする。

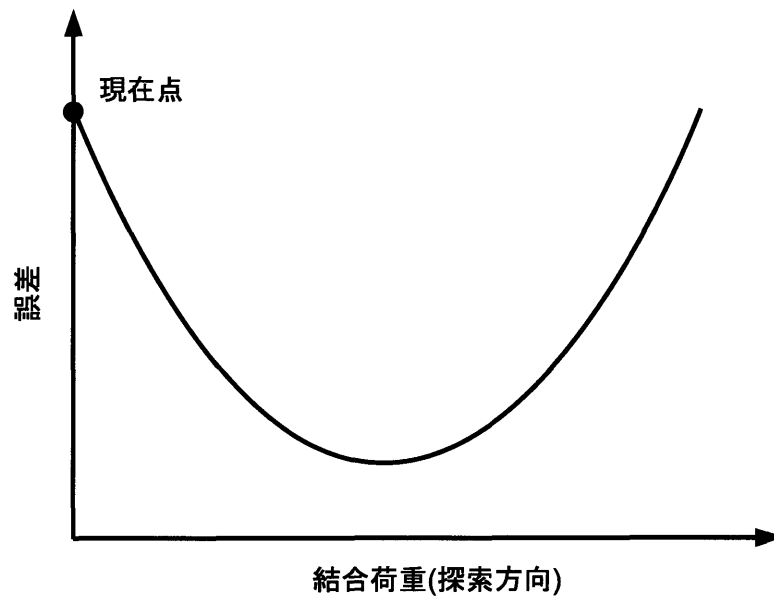


図 3.1: 探索方向に限定した誤差曲面

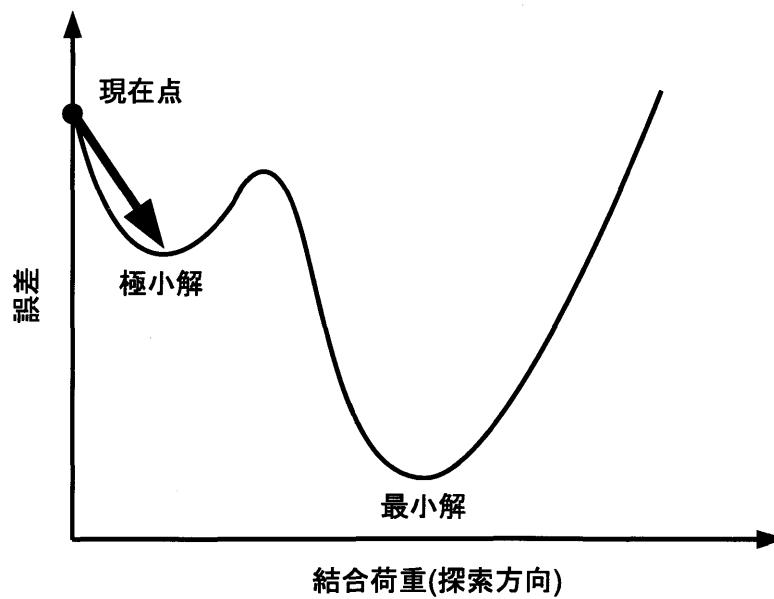


図 3.2: 直線探索がよい結果を得られない誤差曲面の例

## 第4章 実装方法が学習に与える影響

学習に影響を与える要因のひとつとして考えられるのはニューロンの実装方法によるものである。それは、Bohteらの実装方法ではニューロンの内部電位と発火しきい値の交点を正確に計算していないためであり、この方法が何かしらの影響があるのではと考える。この章ではSpikePropの実装方法が学習に与える影響について調査・検討する。

### 4.1 Bohteらの実装方法における非線形最適化手法の効果

2.2節では非線形最適化手法を用いた学習法について説明した。今節では、SpikePropでの誤差曲面が非線形最適化手法を行う場合でもうまくはたらく形状をしているのかを実際の実験結果をもとに検討する。具体的には勾配を用いた降下法で学習を行い、学習の過程で各探索点の近傍における誤差曲面がどのような形状をしているのか調査する。ネットワークのパラメータには以下に示すようにBohteらが行った実験と同じ条件で作成した。

#### 調査に用いる学習対象

調査では簡単な非線形な問題である時間版XOR問題を学習させる。時間版XORの入出力関係を表4.1に示す。表にあるように、同じ時刻にスパイクを入力した場合は時刻16にスパイクを出力し、違う時刻にスパイクを入力した場合は時刻10にスパイクを出力する。

表 4.1: 時間版 XOR 問題

入力時間 1	入力時間 2		出力時間
0	0	→	16
0	6	→	10
6	0	→	10
6	6	→	16

## ネットワークの実装方法

ネットワークの実装方法には、Bohte らが SpikeProp の評価実験で用いた方法 [8] を使って調査を行う。この実装方法は、まず動作時刻を離散化し、この離散化された時刻上で時々刻々と発火しきい値とユニットの内部電位を比較し、内部電位が発火しきい値を超えた時刻を出力時刻とする方法である (図 4.1)。

## パラメータの設定

実験に必要なパラメータの値を以下のとおりにした。

- ネットワーク構造としては、入力層ニューロンが 3 つ (1 つはバイアスニューロン)、中間層ニューロンが 5 つ、出力層ニューロンが 1 つのものを用いる。入力層でのバイアスニューロンとは必ず時刻 0 で発火をするニューロンである。
- 中間層ニューロン 5 つのうち 1 つは抑制性シナプス結合をもつニューロンとする。
- 結合荷重は正の値のみとする。
- スパイク応答関数の時定数は  $\tau = 7$  とする。
- 副結合は 16 本で副結合の時間遅れは 1 から 16 の整数値を用いる。
- 結合荷重の初期化値は 1 から 10 の一様乱数を用いる。
- 発火のしきい値は  $\theta = 50$  とする。

- 勾配を用いた降下法の学習率は  $\eta = 1$  とする.

## 4.2 Bohte らの実装方法での誤差曲面の調査結果

学習初期段階（学習 1 回目）における誤差曲面を図 4.2 に示す. 横軸は勾配方向にどれだけ結合荷重を変化させたかを表し, 縦軸は誤差を表している. 結果をみると, 探索方向の誤差曲面には非常に多くの凹凸が発生し, 極小解が多いことがわかる. このような状態だと, 最急降下法, 共役勾配法で用いている直線探索がすぐに終了してしまうことになる. 実際に最急降下法, 共役勾配法を用いて学習を行ってみたが, 学習開始直後に学習に失敗してしまった.

また時刻の離散化間隔を変更した場合の誤差曲面についても調査した. 離散化間隔を 0.1 とした場合の誤差曲面を図 4.3 に, 離散化間隔を 0.001 とした場合の誤差曲面を図 4.4 に示す. 結果をみると離散化間隔を小さくすれば凹凸の幅が小さくなることがわかる. このことから離散化間隔を小さくすることで, これらの凹凸を事実上問題がない大きさまで小さくすることは可能だと予想できるが, Bohte らの実装方法では離散化間隔を  $1/n$  にすることは計算量が  $n$  倍になる. つまり, Bohte らの実装方法では実用的な範囲では直線探索を含む手法において, 誤差曲面上に存在する極小点のために, 学習が失敗するといえる.

## 4.3 誤差曲面に凹凸ができる原因

誤差曲面に凹凸ができる原因として考えられるのが, Bohte らの実装方法である. この実装方法では, 動作時刻を離散化して時々刻々とニューロンの内部電位と発火しきい値を比較し, この離散時刻上で初めて内部電位が発火しきい値を超えた時刻を出力時刻としている. しかし実際に内部電位が発火しきい値を超えた時刻とはずれているため離散化誤差が生じることになる (図 4.5). この離散化誤差によって誤差曲面に凹凸が発生しているのではないかと考え, 出力時刻の計算を正確に計算した場合の誤差曲面も調査した. 正確に出力時刻を計算した場合の学習初期段階における誤差曲面を図 4.6 に示す. 図 4.2 と比べると, 図 4.6 での誤

差曲面には凹凸がなくなっている。これより、実装方法が原因であることが明らかになった。

しかし、ここでひとつ疑問点がある。それは、誤差は式 2.4 で定義されるように 2 乗誤差なので、離散化誤差を含む場合には、誤差曲面が図 4.7 のように階段状になると予想されるのに対し、実際には凹凸が発生している点である。この原因を明らかにするために、学習対象とした時間版 XOR 問題の各パターンについて誤差曲面はどのような形状をしているのかと思い調査した。調査結果を図 4.8 に示す。結果をみると、各パターンごとでは誤差曲面が階段状になっている。しかし、パターンによって誤差の増減の違いがあり、階段のステップの幅が違う。このため、重ねあわさることによって凹凸のある誤差曲面となることがわかった。

このような凹凸のある誤差曲面だと、極小解にすぐに陥り学習に失敗してしまう。

## 4.4 課題

前節までの結果より、学習に非線形最適化手法を用いる場合、出力時刻に離散化誤差を含む実装方法をするとうまく働かないということがわかった。これを解決するには、出力時刻に離散化誤差を含まないように実装すればよいが、この実装方法では、ニューロンの内部電位と発火しきい値の交点を正確に計算しなければいけないため計算のコストがかかるという欠点がある。これでは非線形最適化手法を用いることによって学習にかかるコストを減らしたところで、あまり意味がない。解決するには、計算コストがあまりかからず出力時刻に離散化誤差を含まない実装方法か、離散化誤差による誤差曲面の凹凸を許容できる学習法が必要となる。

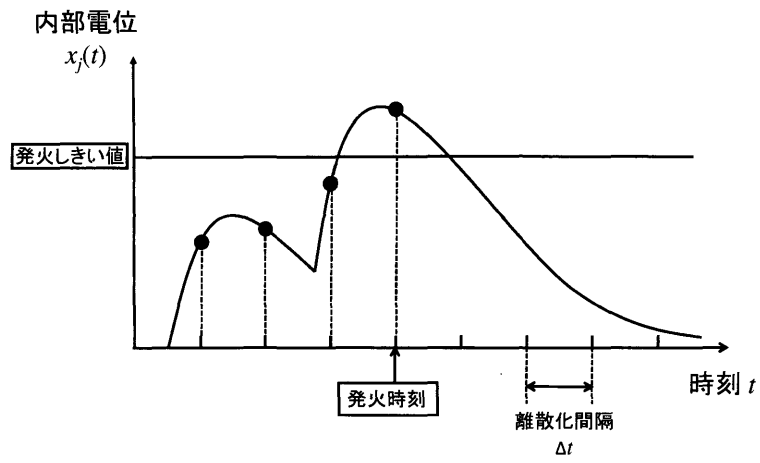
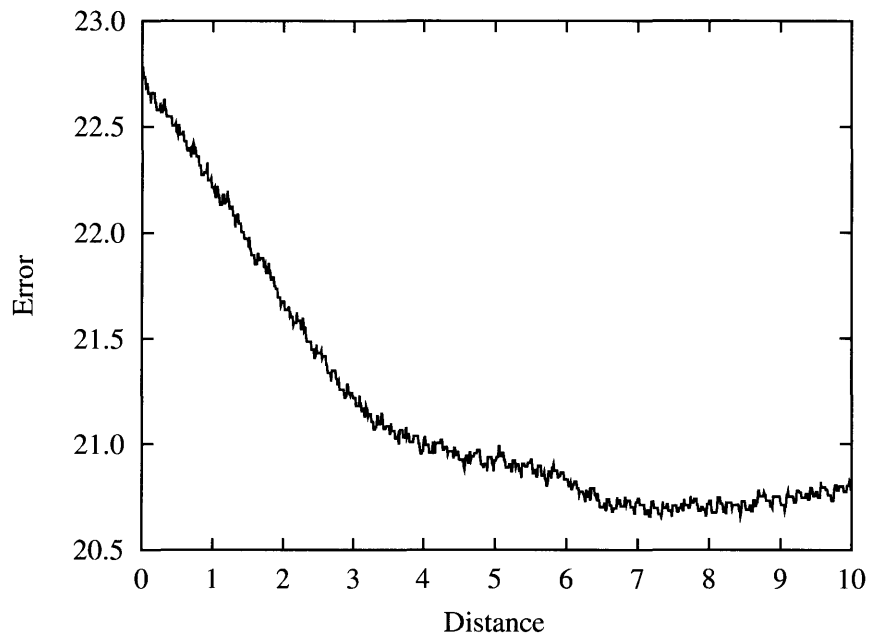
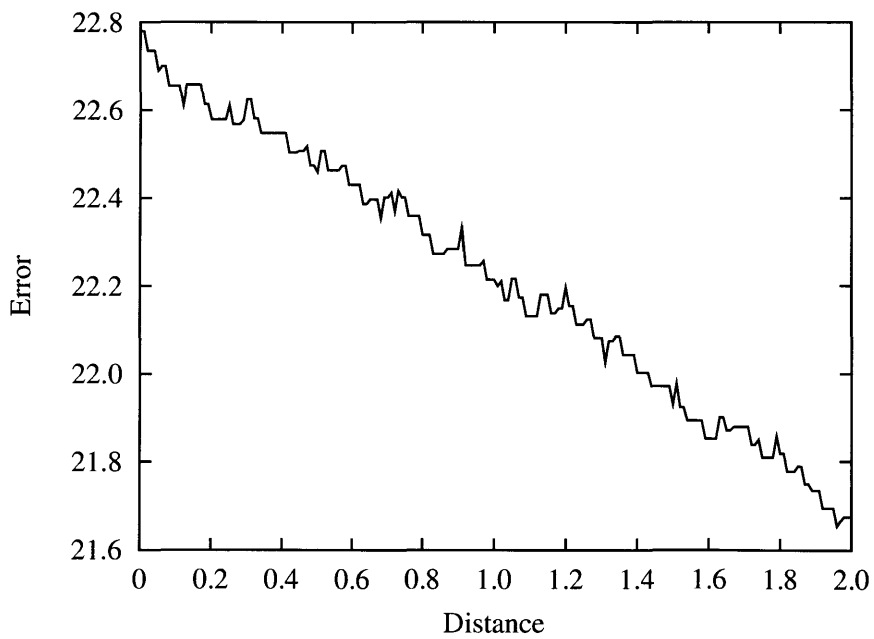


図 4.1: 内部電位と発火しきい値を時々刻々と比較する方法





(a)



(b) 拡大図

図 4.2: 学習初期段階における誤差曲面 (離散化間隔 0.01)

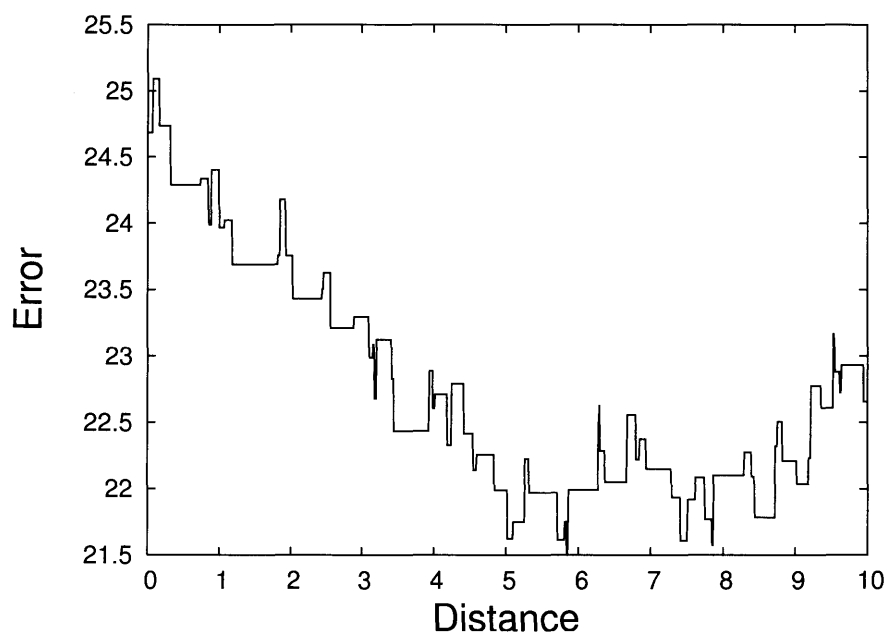
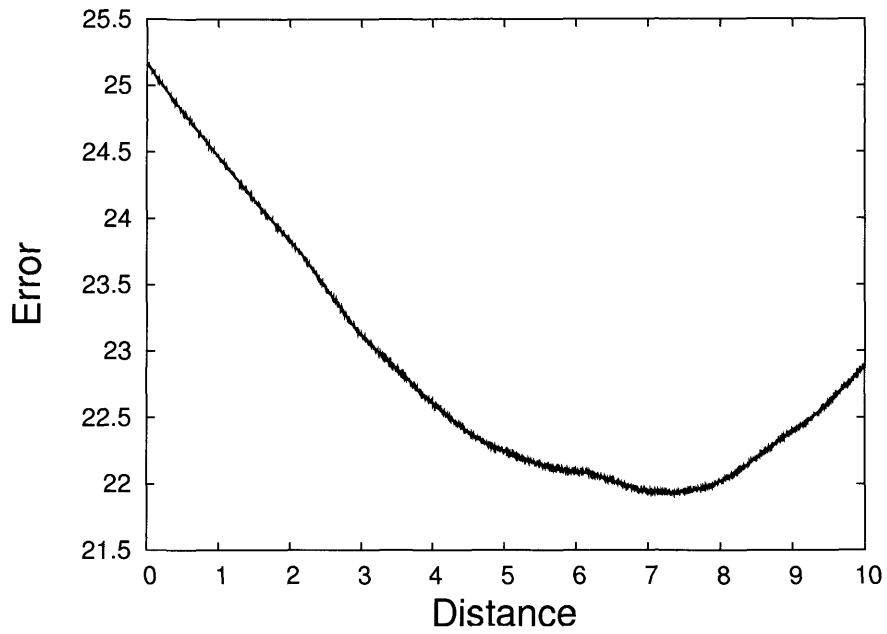
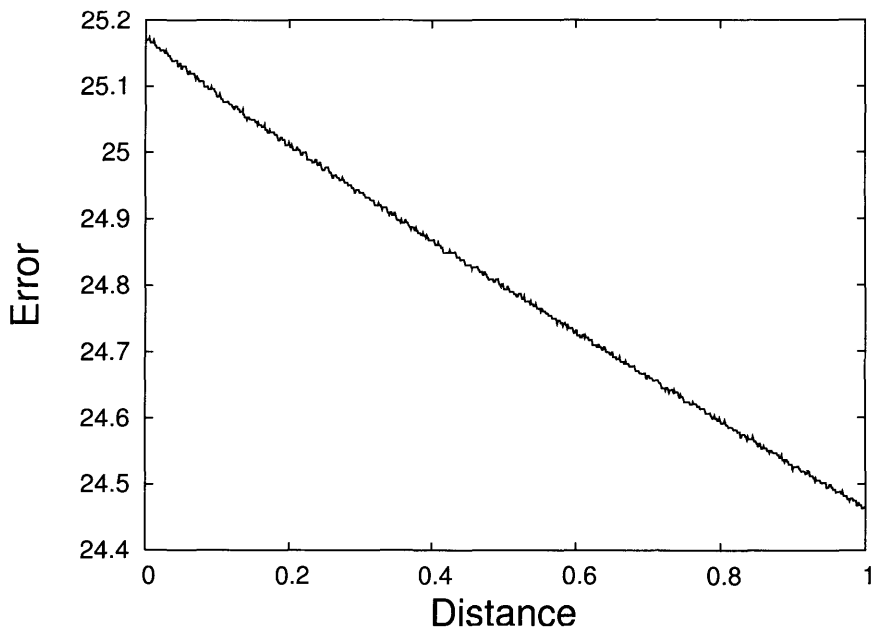


図 4.3: Bohte らの実装方法での学習初期段階における誤差曲面 (離散化間隔 0.1)



(a)



(b) 拡大図

図 4.4: Bohte らの実装方法での学習初期段階における誤差曲面（離散化  
間隔 0.001）

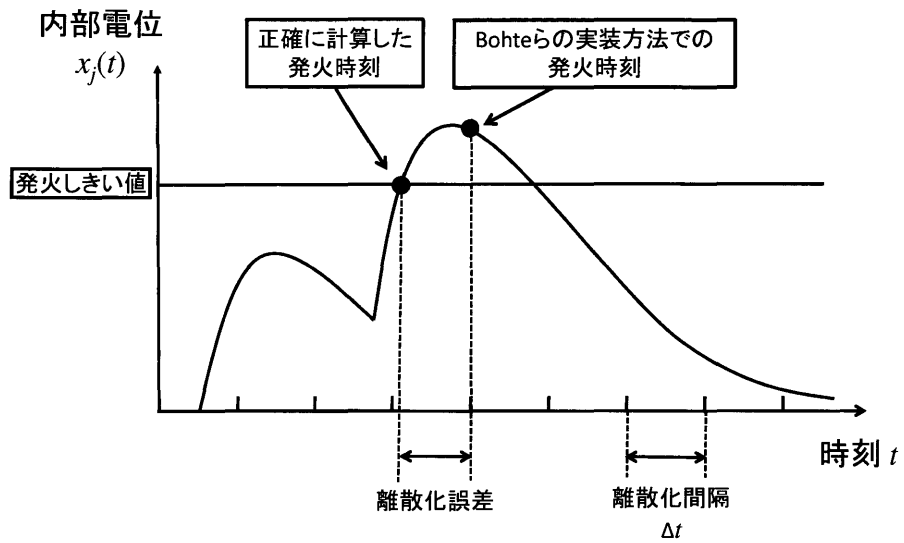


図 4.5: 離散化誤差

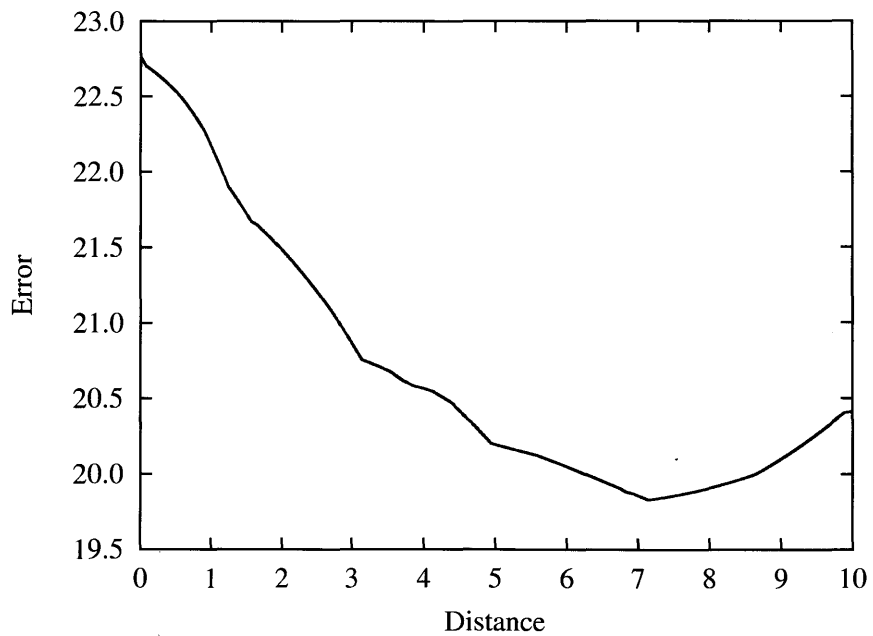


図 4.6: 正確に出力時刻を計算した場合の学習初期段階における誤差曲面

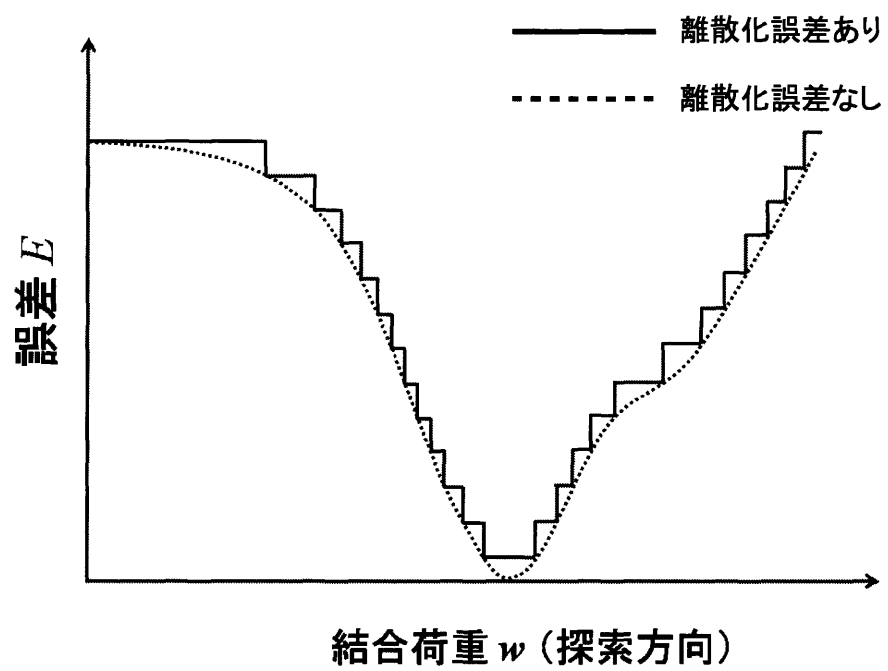
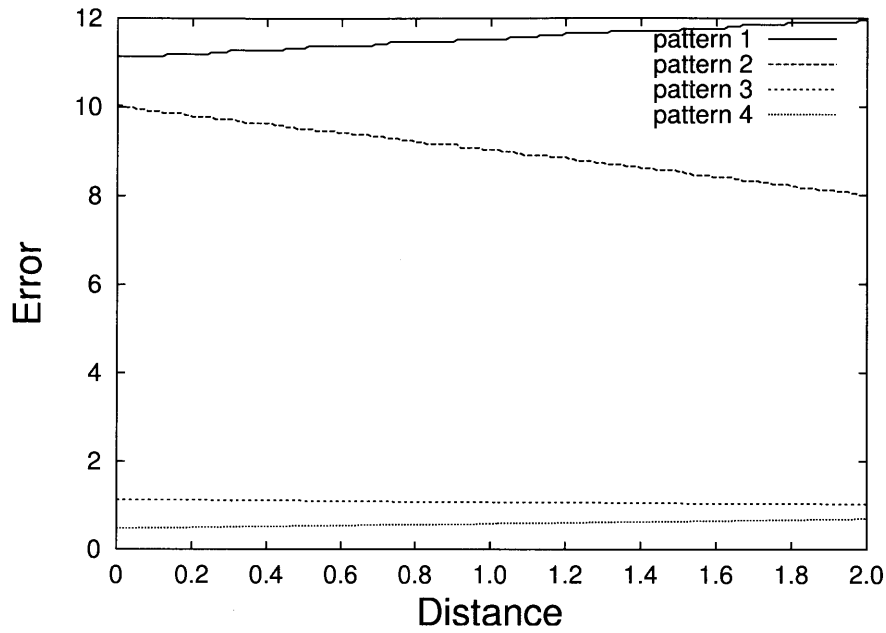
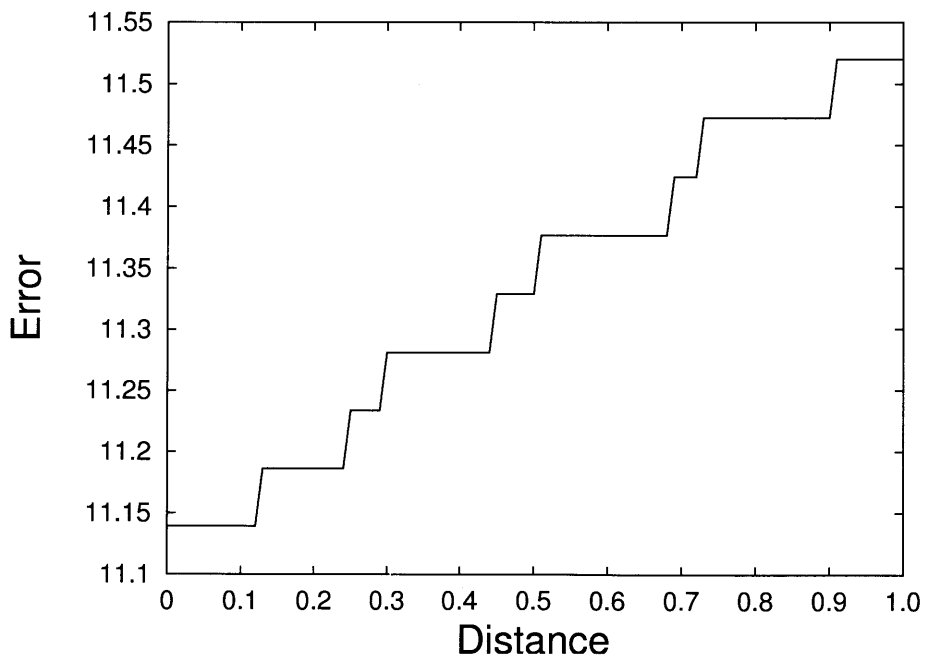


図 4.7: 離散化誤差を含む場合の予想される誤差曲面



(a) すべてのパターン



(b) pattern 1 のみを拡大

図 4.8: 学習パターンごとの誤差曲面

# 第5章 SpikePropのネットワークモデルが学習に与える影響

この章では SpikeProp のネットワークモデルが非線形最適化手法を用いた学習に与える影響について議論する。実際に非線形最適化手法を用いた学習を行った結果、学習終盤に極小解にはまり学習に失敗してしまうことがわかった。そこで学習失敗時の誤差曲面を調査し、学習終盤で極小解が発生する原因について調査する。

## 5.1 出力時刻を正確に計算した場合での非線形最適化手法の効果

出力時刻を正確に計算した場合での非線形最適化手法の効果を調査するために、実際に最急降下法、共役勾配法を用いて時間版 XOR の問題（表 4.1）を学習した。実験のパラメータは 4.1 節で行った実験と同一のものを用い、100 種類の結合荷重の初期乱数シードを用いて、学習が何回成功するのか調査した。学習成功率と平均学習回数を表 5.1 に、学習過程の一例を図 5.1 に示す。図 5.1 中の BP は勾配を用いた降下法、CG は共役勾配法、GDM は最急降下法である。結果をみると、勾配を用いた降下法に比べ、最急降下法、共役勾配法ともに学習回数は減少しているが、多くの場合、極小解にはまって学習に失敗し学習の成功率が低い。特に共役勾配法については 100 回中 1 回しか学習に成功していない。また、学習過程を見ると、勾配を用いた降下法においても振動しながら収束している。これらのことより極小解が多く発生しているのではないかと考える。

表 5.1: 非線形最適化手法での結果

学習法	学習成功率 [%]	平均学習回数 [回]
勾配を用いた降下法 (BP)	100	1337
最急降下法 (GDM)	20	272
共役勾配法 (CG)	1	188

## 5.2 極小解が発生する原因の考察

5.1 節の実験の結果より極小解が多いのではないかと考えた。そこで実際に学習に失敗した時点での誤差曲面を調査した。結果を図 5.2 に示す。結果をみると、探索方向に結合荷重を少し変化させると、わずかに減少し、その後誤差が急激に増加しているのがわかる。実際には、誤差はわずかながら減少するので極小解にはなっていない。しかし、この図のようにわずかに結合荷重を変化させると誤差が急激に増加してしまう場合、直線探索の最小探索精度によっては探索方向の最小値を見つけることができずに学習に失敗してしまうことになる。

では、なぜ結合荷重を少し変化させただけで誤差が急激に増加してしまうのか。そこで着目したのは内部電位の変化の仕方である。2.4 より誤差関数  $E$  は出力層ニューロンの出力時刻  $t_o$  を変数とした関数である。また、出力層ニューロンの出力時刻は  $t_o$  は内部電位  $x_o(t)$  が発火しきい値  $\theta$  を超えた時刻なので、出力時刻  $t_o$  が急激に変化する原因は内部電位の変化の仕方であると考えられる。

## 5.3 内部電位の変化のしかたが学習に与える影響の考察

2.1.3 節で説明したとおり、Bohte らは学習法を考案する上で、結合荷重を変更したことによって変化した内部電位の変化量と、それによる出力時刻の変化量は比例するという前提を用いている。この前提が成り立つ内部電位の変化は図 5.3 のような単調増加の変化をしなければならない。しかし実際の内部電位の変化はどうだろうか。まずは、共役勾配法での学習における学習開始直後での出力層ニューロンの内部電位の変化を図 5.4 に示す。実線が内部電位の変化、破線が発火しきい値である。このときの内部電位の変化の仕方は Bohte らの想定のとおり単調増加



となっている。次に、学習に失敗する直前での出力層ニューロン内部電位の変化を図5.5に示す。内部電位は最初は増加し、その後減少、再び増加するという変化をする。このような山がある内部電位の変化の仕方をしてる場合、結合荷重の変更により最初の山の高さが発火しきい値をこえてしまうと、発火時刻が不連続かつ大幅に変化することになる。実際に、この内部電位の変化の仕方の影響をみるために、共役勾配法で学習に失敗した時点から、勾配を用いた降下法で1回学習を行ったときの出力層ニューロンの内部電位の変化を見てみる。結果を図5.6に示す。結果を見ればわかるとおり、勾配を用いた降下法での学習前は最初の山が発火しきい値以下であったのに対し、勾配を用いた降下法で1回学習を行うと最初の山が発火しきい値を超えて出力時刻が大幅に変化しているのがわかる。このように、Bohteらの想定していたことが成り立たない内部電位の変化をしている。

## 5.4 内部電位に山ができる原因

では、図5.5のような内部電位の変化をする原因は何なのだろうか。この原因として、時間がたつにつれ減衰するスパイク応答関数の重ね合わせによるものと抑制性シナプス結合によるものではないかと考えた。以下で詳しく説明する。

- **時間がたつにつれ減衰するスパイク応答関数の重ね合わせによるもの**

ニューロンの内部電位は式2.2で表されるとおり、時間がたつにつれ減衰するスパイク応答関数の重ね合わせによって定義されている。模式図で描くと図5.7のようになる。

- **抑制性シナプス結合によるもの**

Bohteらの実験では中間層に1つだけ抑制性シナプス結合を持つニューロンを持たせている。この抑制性結合をもつニューロンからの出力されたスパイクを出力層ニューロンが受け取ると、図5.8のように内部電位が減少することになる。

5.5の内部電位の変化を見てみると、スパイク応答関数の減衰時定数は7なので、内部電位に山ができる原因がスパイク応答関数によるものであるとは考えにくい。しかし、抑制性シナプス結合をもつ中間層ニューロンの出力時刻は9.36512であ

り，副結合の時間遅れの幅が1～16であることを考えると，内部電位に山ができるのは抑制性シナプス結合が原因であることがわかる．

## 5.5 課題

いままでSpikePropのネットワーク構造が学習に与える影響について議論した．結果より，抑制性シナプス結合が原因でSpikePropの勾配の計算法の前提が崩れることによる極小解の発生が学習に大きな影響があるということがわかった．しかし，抑制性シナプス結合をなくすと，結合荷重空間が小さくなることが原因でネットワークの出力の多様性が無くなり，最適解が見つからない（学習出来ない）ということになる．実際に抑制性シナプス結合がない場合，時間版XORですら学習できないことになる．このため，抑制性シナプス結合を無くすわけにはいかないので，学習法を図5.5のような変化の仕方をする内部電位に対応させる必要がある．

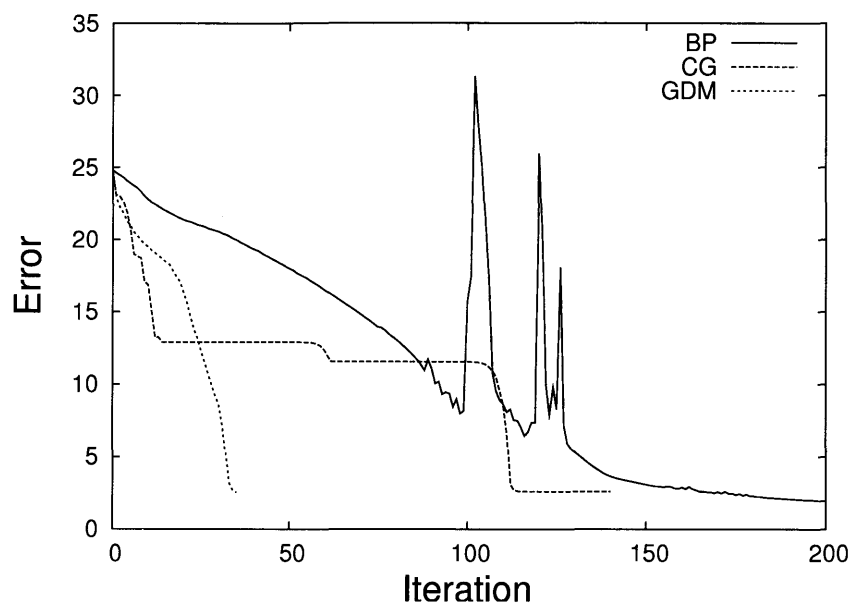
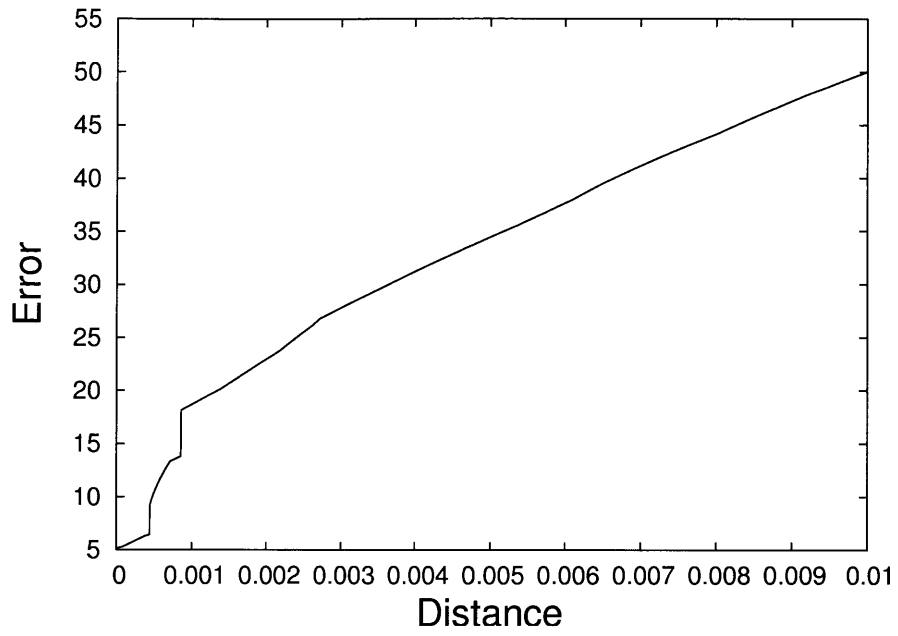
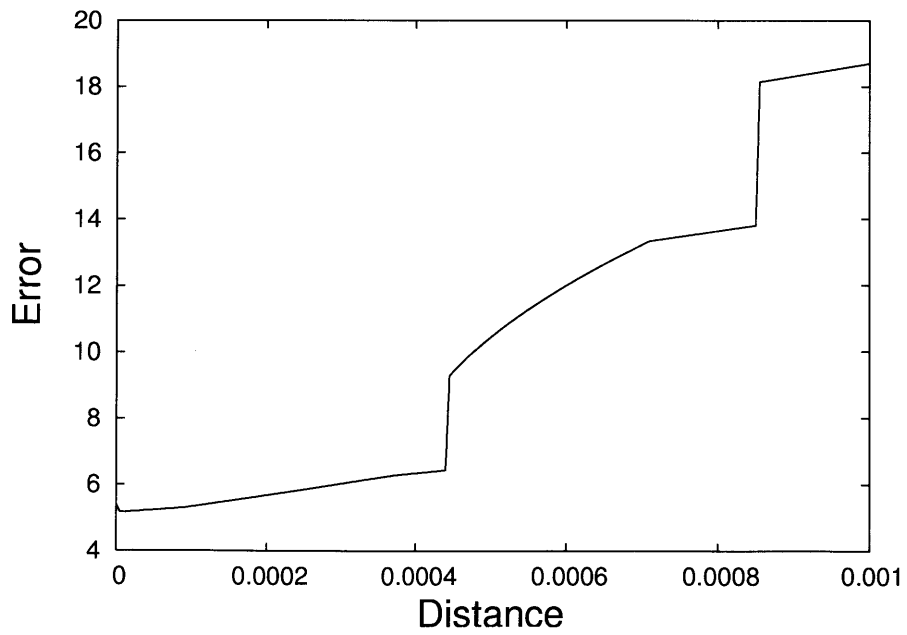


図 5.1: 学習過程の一例



(a)



(b) 拡大

図 5.2: 学習に失敗する直前における現在点付近の誤差曲面

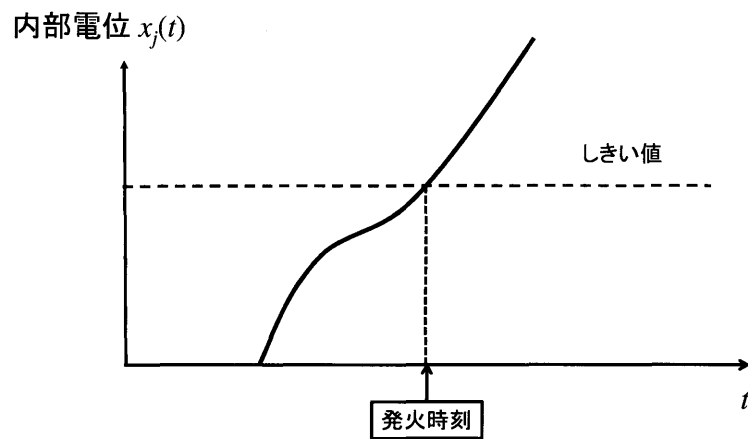


図 5.3: 想定が成り立つ内部電位の変化

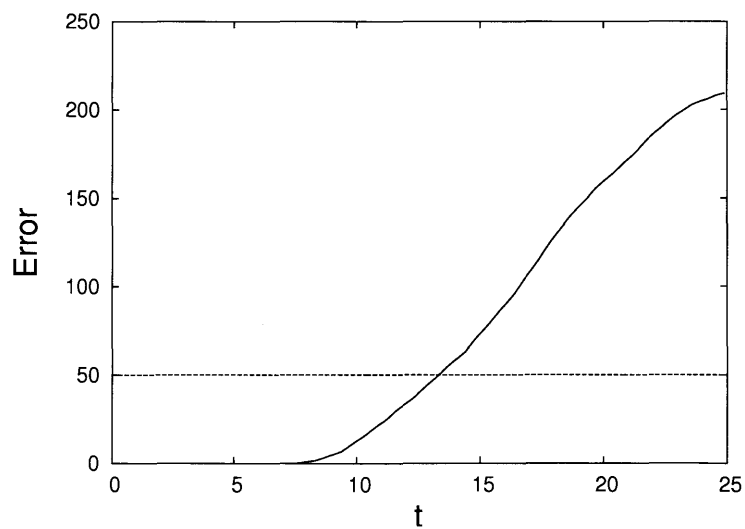


図 5.4: 学習開始直後の内部電位の変化

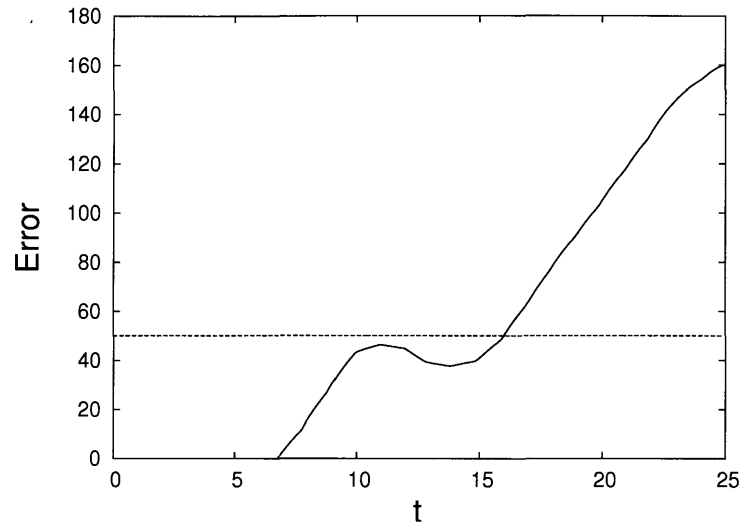


図 5.5: 学習失敗時の内部電位の変化

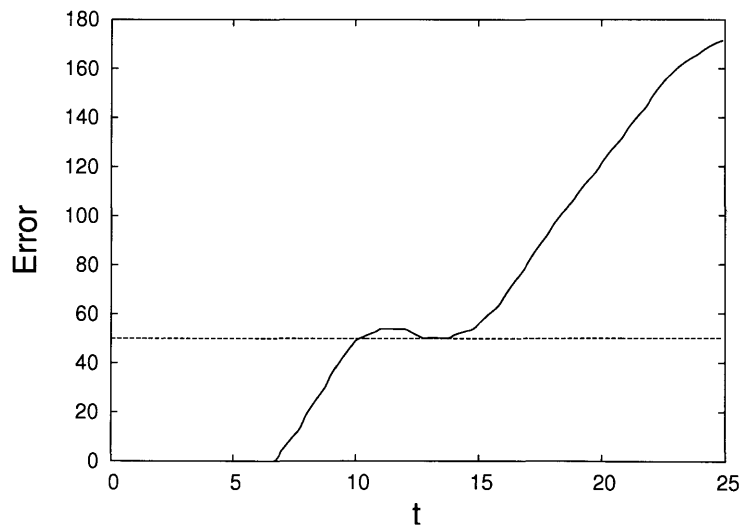


図 5.6: 学習失敗時から一回学習後の内部電位の変化

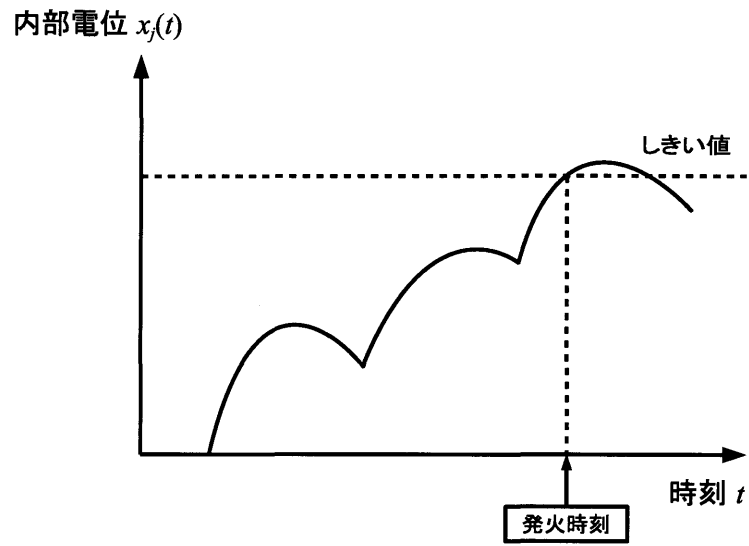


図 5.7: スパイク応答関数の重ね合わせによるもの

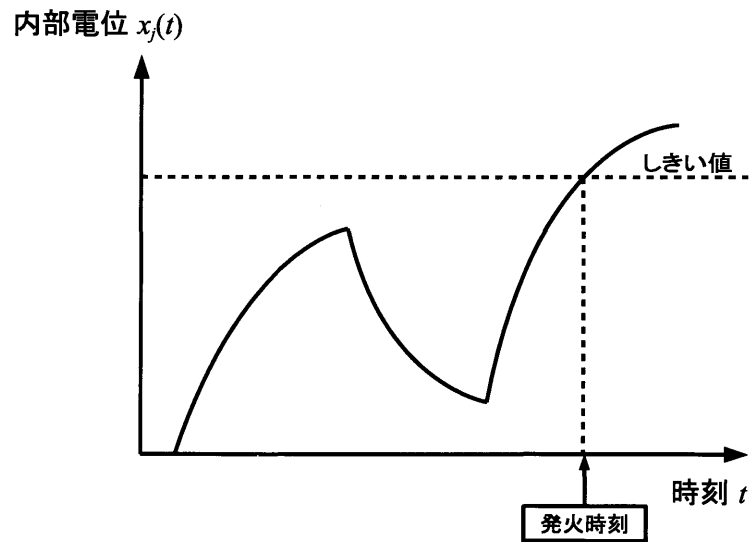


図 5.8: 抑制性結合によるもの

## 第6章 まとめ

本研究では SpikeProp における学習を阻害する要因を調査した。

SpikeProp では勾配を用いた降下法で学習を行っているため、学習が収束するまでに時間がかかるという欠点がある。従来のシグモイド型ニューラルネットワークでも同様の問題があり、この問題の解決法として非線形最適化手法を用いた学習が行われている。しかし、非線形最適化手法を用いた学習を SpikeProp で行った場合、学習がうまくいかない。そこで誤差曲面を用いて、学習を阻害している要因について調査を行った。調査の結果、学習を阻害しているのが極小解によるものであることがわかった。そして極小解ができる原因が以下の2つであることをつきとめた。

1. 出力時刻に離散化誤差を含む実装方法によるもの
2. SpikeProp のネットワーク構造、特に抑制性シナプス結合によるもの

1つ目については、正確に出力時刻を計算すれば極小解の発生はなくすことができるが、正確に計算する場合、計算コストがかかるという問題がある。2つ目については、抑制性シナプス結合をなくすと学習できるものの種類が極めて狭くなり、実用性がなくなる。そのため抑制性シナプス結合は必要であり無くすことはできない。

今後、非線形最適化手法を用いた学習を SpikeProp に適用する場合はこの2点を考慮に入れて学習法を考える必要がある。



# 謝辞

本論文は、著者が三重大学大学院工学研究科博士前期課程時に行った研究をまとめたものである。本論文を進めるにあたり、懇切丁寧な御指導と御督励を賜った三重大学の林照峯教授，北英彦准教授，高瀬治彦准教授に感謝いたします。また，日頃熱心に討論していただいた計算機工学研究室の皆様方にお礼申し上げます。

最後に，本論文をまとめるにあたり，助言，討論，その他お世話になったすべての方々に感謝いたします。

## 参考文献

- [1] Wolfgang Mass and Christopher M. Bishop: Pulsed Neural Networks, The MIT Press, pp.3–85 (2001)
- [2] 川人光男: 小脳内部モデルとモザイクの計算論的神経科学, 人口知能学会誌, Vol.17, No.6, pp.720–731 (2002)
- [3] 銅谷賢治, 伊藤浩之, 藤井宏, 塚田稔: 脳の情報表現 — ニューロン・ネットワーク・数理モデル, 朝倉書店 (2002)
- [4] 黒柳奨, 奥村享司, 岩田彰: パルスニューロンモデルを用いた音源定位聴覚系モデル —位相情報抽出のためのニューロン層の構築—, 電子情報通信学会技術研究報告, Vol.96, No.134, pp.163–169 (1997)
- [5] 元木誠, 小坂成一, 平田廣則: ニューラルネットワークの違いに着目した自律移動ロボットコントローラの性能比較, 電子情報通信学会技術研究報告, Vol.105, No.457, pp.25–30 (2005)
- [6] 吉富康成: シリーズ非線形科学入門 ニューラルネットワーク, 朝倉出版, pp.27–49 (2002)
- [7] 黒江康明: スパイクングニューラルネットワーク — 学習法を中心として, システム制御情報学会誌, Vol.48, No.2, pp.57–62 (2004)
- [8] S.M. Bohte, J.N. Kok, and J.A.L. Poutre: SpikeProp: backpropagation for networks of spiking neurons”, Proceedings of ESANN’2000, pp.419–424 (2000)
- [9] Simon Christian Moore: Back-propagation in Spiking Neural Networks, Master’s thesis, University of Bath (2002)

- [10] 金谷健一: これなら分かる最適化数学—基礎原理から計算手法まで, 共立出版 (2005)
- [11] Sam McKennoch, Dingding Liu, and Linda G. Bushnell: Fast Modifications of the SpikeProp Algorithm, Proceedings of 2006 International Joint Conference on Neural Networks, pp.7701–7708 (2006)
- [12] Sergio M. Silva, Antonio E. Ruano: Application of the Levenberg-Marquardt method to the training of spiking neural networks, Proceedings of 2006 International Joint Conference on Neural Networks, pp.7709–7713 (2006)
- [13] 吉田隆弘, 後藤正幸, 俵信彦: 共役勾配法における探索効率向上法に関する一考察, 日本経営工学会論文誌, Vol.48, No.5, pp.257–263 (1997)
- [14] 後藤正幸, 開沼泰隆, 俵信彦: 変傾共役勾配法による BP 学習の安定化と高速化, 日本経営工学会論文誌, Vol.46, No.2, pp.152–158 (1995)
- [15] 能勢容幸, 原一之: 大域情報を用いた階層形ニューラルネットワークの学習法, 電子情報通信学会技術研究報告, Vol.99, No.193, pp.63–68 (1999)
- [16] 山本浩一, 郷原一寿, 内川嘉樹: ニューラルネットにおける誤差曲面の解析, 情報処理学会全国大会講演論文集, 第38回平成元年前期, No.1, pp.494–495 (1989)
- [17] 太田宜衛, 河村嘉顯: 2次逆伝播に基づく層別ニュートン法を用いた多層パーセプトロンの学習, 電子情報通信学会技術研究報告, Vol.94, No.98, pp.65–72 (1994)
- [18] T. NatschlAager, B. Ruf: Spatial and temporal pattern analysis via spiking neurons. Network, Computation in Neural Systems, Vol.9, No.3, pp.319–332 (1998)

## 発表論文

- [1] 藤田優, 高瀬治彦, 北英彦, 林照峯: SpikeProp の誤差曲面に対する一考察, 情報科学技術レターズ, pp.193-194 (2007)
- [2] FUJITA Masaru, TAKASE Haruhiko, KITA Hidehiko, and HAYASHI Terumine: Shape of error surfaces in SpikeProp, Proceedings of 2008 International Joint Conference on Neural Networks, pp.841-845 (2008)
- [3] FUJITA Masaru, TAKASE Haruhiko, KITA Hidehiko, and HAYASHI Terumine: Influence of the discrete time computation in SpikeProp, Proceedings of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems 2008, pp.930-935, (2008)