ニューラルネットワーク学習の正則化における複素解析関数利用の効果

H-8 Effects of a holomorphic approach for regularizing the training of neural networks

下川 信祐† 大田原 一成†
SHIMOGAWA, Shinsuke OHTAWARA, Kazushige

1 まえがき

私たちはモノ・人・社会など様々な環境との関わりの多くで記号作用を媒介とします [1]. 例えば，インターネットや携帯電話などの情報通信ネットワーク・IT の発展に記号作用の性質が深く関わっています．またブランドなどを通じて経済に運動量をもたらしています．その結果が記号作用をますます拡大させ，人々の思考・行動様式や価値観にも影響を与えています．

記号作用は人間と環境が関わって人に生じる複雑な作用ですが，環境に縛られない自由度（あそび，恣意性）を生み出す出すという一定の性質があり，環境にも一定の特徴を伴い，記号を生み出します．こうしてモノ・技術を用いて記号作用に関わることが可能になっています．

データから記号作用を捉えてゆく方法論は，文字や画像の認識にとどまらず，私たちを取り巻く生活や社会についての理解を支援し，人とモノの関係の評価として人々に受け入れられるシステムやサービスのデザインに役立ちます．また，社会の様々な変数について予測する手法にもつながります．

データに記号を対応させる方法論については，(i) 記号抽出の正当性，(ii) 計算コストなどの課題がありますが，本研究では，計算のコストに関する問題に注目します．具体的には，ニューラルネットに記号を学習させるのに要する最適化問題の計算効率です．

多層ニューラルネットの学習では，多重 basin (極小解) やプラトー（学習の長い停滞）が計算量を増大させる深刻な要因になります（図 1）．このような評価関数のランドスケープの複雑さを生み出す本質は，記号の不連続で離散的な性質が探索アルゴリズムの連続的性質と対立していることにあります．そこで本研究では，不連続性・離散性と連続性の関係を協同的な関係にすることで，ランドスケープをシンプルにすることを目指します．

2 ニューラルネットにおける連続性と不連続性の対立とその対処

記号の離散的な性質を反映して，ニューラルネットは元々階段関数から作られていました．しかし，階段関数の不連続性が 評価関数のランドスケープを不連続にするため，探索が効率化に利用できる局所的な手がかりが失われ大規模化できません．そこで，階段関数の不連続性をシグモイド関数

$$\sigma(x) = 1 - \frac{1}{1 + e^{\lambda x}} \qquad (1)$$

に置き換えることで，連続性を導入し，勾配などが意味を持つ（正則化）ことになりました [2]．しかし，元々階段関数の不連続部分をなめらかにしただけなので，小さい導関数がプラトーを引き起こし，また多数の記号データ相互の離散性が，多数の極小解を生みだして，評価関数のランドスケープを複雑なものにしています（図 1）．

ところで，連続的なものと不連続・離散的なものは，一見対立するように見えます．しかし現実は，両者何れも人間が環境との関わりから体得してきたベーシックな記号作用・感覚であり，協同して人間の環境との関わりを巧みに媒介しています [3]．このとき，記号の不連続性・離散性は，連続的な感覚の境界面として働いています．つまり，諸記号の不連続性は，連続的なものの境界面に捉えることで協同的な関係となると期待されます．これは，階段関数をシグモイド関数にするのではなく，階段関数を「より高次元の空間に定義された性質の良い（つながりに優れた）関数が低次元の境界面に定義する関数」として捉えることを示唆します．

3 複素解析関数の利用

より高次元の空間に定義された性質の良い概念を用いて，低次元の複雑な概念をとらえることは，数学では resolution/reduction として知られます．複雑な実関数を複素解析的な関数の境界として捉えることは，佐藤の超関数論の考え方です．

たとえば階段関数，$H(x) = 0$ if $x < 0$, $H(x) = 1$ if $x > 0$ は複素対数関数 $\log(z)$ を用いて，

$$H(x) = -\frac{1}{2\pi i}(\log(-x - i0) - \log(-x + i0)) \quad (2)$$

と表記されます（$i = \sqrt{-1}$）．複素数 $z = x + iy$ の偏角関数 $\arg(z)$: $z = \exp(\log|z| + i\arg(z))$ によって階段関数のギャップが繋がります．

複素解析関数は，定数関数でなければ微分が消滅しないためプラトーの低減が期待されます．また，零点集合が大きく，ランドスケープを簡単にするため，極小解の問題に対処しやすいことが期待されます．

4 評価関数のランドスケープの構造

ここでは，feed foward 型ベクトル値ニューラルネットの各層を接続する関数関係 $G(x) = (g_{ij}^k(x))$ のパラメータ x が，実空間の全域で実解析的（例えば多項式写像）で与えられるものを考えます．シグモイド関数 σ が中間層に閾値関数として設定されるとします． パラメータ x, 入力 a の時の出力を，

†（株）国際電気通信基礎技術研究所 適応コミュニケーション研究所, ATR Adaptive Communications Research Labratories

$N(a,x)$ と記します．教師データ $\{(a_i,b_i)\}_{i=1}^N$，に対して，平均二乗誤差

$$\sqrt{V(x)} = \sqrt{\sum_{i=1}^{N} ||N(a_i,x)-b_i||^2/N} \qquad (3)$$

を与えられた基準値より小さくするパラメータ x を探索することが問題です．ここで $||\cdot||^2$ は，ベクトルの長さの２乗を表します．x の次元を n，実数体を $\mathbb{R}$，複素数体を $\mathbb{C}$ と記します．前提から $V(x)$ の定義域は $\mathbb{R}^n$ です．これを $\mathbb{C}^n$ において $\mathbb{R}^n$ の近傍の複素解析関数に拡張したものを $V(z)$ と記します．$z=x+iy$, $y \in \mathbb{R}^n$ と表記されます．適当に大きな $C>0$ を取れば，$V(x)$ を最小化する問題と $|V(z)|+C||y||^2$ を最小化する問題を置き換えることが可能です．

$V(z)=c$ が解けやすいことに注目すると $V(z) \to V(x)$ as $y \to 0$ のランドスケープの構造が見えてきます．これを非線形探索手法に利用することで，極小解やプラトーなどの問題を回避できる可能性が高まります．

複素数値で山の形を考えることは出来ませんが，“等高面”（$V(z)=c$ の族）を考えることができます．等高面は以下の性質を持ちます（$n \le 2$）．

(a) $V(z)=c$ は，$\mathbb{C}^n$ の $n-1$ 次元集合を定める．

(b) $V(z)=c$ 上のベクトル場 $\mathrm{grad}_{V(z)=c}||y||^2$ は，実点 $(V(x)=c)$ を除き正則点で非退化．

この２つの性質から，$V(z)$ は $y \to 0$ に沿って図**2**のような構造を持つことが解ります．$G(x)$ が簡単（線形写像など）で３層場合，$V(z)=c$ 上の勾配ベクトル場に沿うことで $V(x) \le c$ を解くことが可能になります．($0 \le c \le V(x_0)$ for some $x_0 \in \mathbb{R}^n$)．図**3**に計算例を示します．

5 まとめと課題

記号をデータから抽出するときに問題となる，不連続性と連続性の対立を解消する方法として，複素解析関数の利用を検討しました．複素領域から眺めると，問題のランドスケープの構造が解りやすく，極小解を避けた探索が可能になります．勾配ベクトル場による探索手順の詳細と証明，極小解を脱出する従来手法との連携，大規模化での計算効率評価などが今後の課題です．

謝辞

本研究は通信・放送機構の研究委託「自律分散型無線ネットワークの研究開発」により実施したものである．

参考文献

[1] 丸山　圭三郎，ソシュールの思想．岩波書店．1981．

[2] M．ミンスキー，他，パーセプトロン．中野　他訳，パーソナルメディア．1993．

[3] 木田　元，ハイデガーの思想．岩波書店．1993．

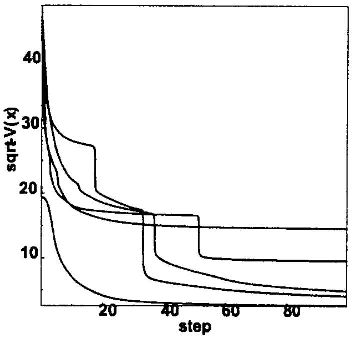

図 **1.** 極小解捕捉とプラトー．

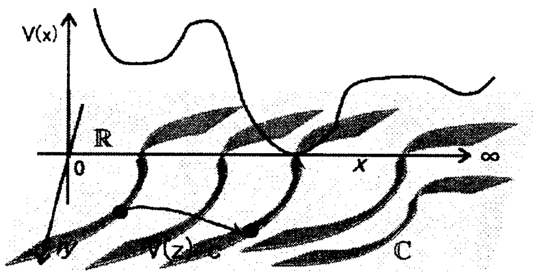

図 **2.** 複素解析関数として見た評価関数．

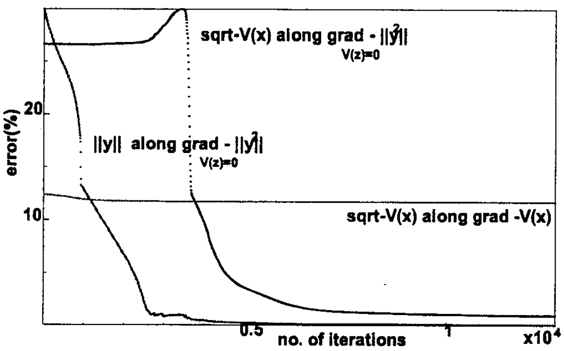

図 **3.** 複素解析関数利用の効果．