

## 科学研究費助成事業 研究成果報告書

令和 5 年 5 月 28 日現在

機関番号：32686

研究種目：基盤研究(C)（一般）

研究期間：2016～2022

課題番号：16K00347

研究課題名（和文）ニューラルネットワークの特異点の解消

研究課題名（英文）Resolution of singularities in neural networks

研究代表者

新田 徹 (NITTA, Tohru)

立教大学・人工知能科学研究科・特任教授

研究者番号：20357726

交付決定額（研究期間全体）：（直接経費） 3,500,000 円

研究成果の概要（和文）：ニューラルネットワークの特異点の特性の一端を明らかにした。つまり、深層ニューラルネットワークが数多くの「階層構造に基づいた特異点」を持っていることを数理的に明らかにし、そのような特異点を持たないための十分条件を明らかにした。また、深層複素数型ニューラルネットワークと等価な深層実数型ニューラルネットワークにおいて、いくつかの特異点が複素数に基づく性質により自ずと解消されていることを数理的に示した。さらに、非線形な深層ニューラルネットワークが（学習性能に悪い影響を与える）悪い局所解（学習誤差の大きい局所解）を持たないための十分条件を明らかにした。

研究成果の学術的意義や社会的意義

深層学習技術では、大量のパラメータの調整が必要であるため、学習が行える適切な条件を特定するのに大変な労力がかかっている。ニューラルネットワークには、学習に悪い影響を与える多くの特異点が存在するからである。本研究では、ニューラルネットワークの特異点の特性を明らかにし、特異点を持たないための十分条件を導いた。また、特異点を持たないタイプのニューラルネットワークを提示した。これらの研究成果は深層ニューラルネットワークの学習性能向上に、引いては現在世界的に進められている深層学習を利用した社会実装に資するものと考えられる。

研究成果の概要（英文）：We have revealed some of the characteristics of singularities in neural networks. In other words, we mathematically clarified that deep neural networks have a number of "hierarchical structure-based singularities" and derived the sufficient conditions for not having such singularities. In addition, we mathematically showed that some singularities in deep real-valued neural networks, which are equivalent to deep complex-valued neural networks, are naturally resolved due to their complex number-based nature. Furthermore, we derived the sufficient conditions for nonlinear deep neural networks not to have bad local minima with large learning errors which have a bad influence on learning performance.

研究分野：情報科学

キーワード：ニューラルネットワーク 特異点 深層学習 ディープラーニング 危点 学習

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) 情報化の進展により、産業や自然科学などのさまざまな分野において、大規模で多様なデータが集積されてきていた。そのようなデータを解析するために、統計学や計算科学の知見を取り入れた機械学習の手法が発展し、実問題への応用も進められていた。その中でも、多数の層を持つニューラルネットワークにおける学習(深層学習、deep learning)は注目を集めていた。既存の機械学習に比べて、汎化能力が格段に高く、より人間に近い識別能力を持つからである。しかし、深層学習技術では、大量のパラメータの調整が必要であるため、学習が行える適切な条件を特定するのに大変な労力がかかっていた。これは次のような理由に因っている。一般に、ニューラルネットワークには多くの特異点が存在する(特異点とはパラメータの微小変化に対してシステム応答が不変となる点のこと)。特異点はニューラルネットワークの学習に悪い影響を与えることが明らかにされてきていた。特に、層が深いほど、特異点が学習の進行を妨げるため、深い層を持つニューラルネットワークで学習を行う際のボトルネックとなっていた。深層学習技術では、各層毎に学習を行うなどの工夫によりこの問題に対処し、成果を挙げていたが、その反面、大量のパラメータの調整が必要となり、学習が行える適切な条件を特定するのに大変な労力がかかってしまっていた。

(2) 一般に、ニューラルネットワークにおける特異点の存在証明は非常に難しい。たとえば、XOR(排他的論理和)問題にはローカルミニマム(特異点)が存在すると長い間信じられていた。しかし、実際には特異点が存在しないことが証明されたのは、1998年になってからである(文献 )。研究代表者は2013年頃からニューラルネットワークの特異点の研究に取り組んでいた。その研究活動の中で、ニューラルネットワークを複素数に拡張することによりローカルミニマム(特異点)を解消できることを数理的に明らかにした(文献 )。この経験から、一般に、ニューラルネットワークを複素数や四元数といった高次元の代数系に拡張することが、特異点の解消に結び付くのではないかとの着想を得た。高次元の代数系には、クリフォード代数や Cayley-Dickson 数など様々な種類が存在している。

### 2. 研究の目的

本研究の目的は、ニューラルネットワークのすべての特異点を解消することである。

ニューラルネットワーク、特に深い層(複数個のニューロンから成るグループを層という)を持つ深層ニューラルネットワーク上の学習においては、たとえば何十億個といった大量のパラメータの調整が必要であるため、学習が行える適切な条件を特定するのに大変な労力がかかる。場合によっては何か月もの間、数十台もの高性能コンピュータを使って、ようやく学習できるといった状況もある。このようなことが起きるのは次のような理由に因っている。一般に、ニューラルネットワークには多くの特異点が存在する(特異点とはパラメータの微小変化に対してシステム応答が不変となる点のこと)。層数が多くなればなるほど特異点の数は増える。特異点はニューラルネットワークの学習に悪い影響を与えることが明らかにされている。特に、層が深いほど、数多くの特異点が学習の進行を妨げるため、深い層を持つニューラルネットワークで学習を行う際のボトルネックとなっていた。

本研究では、ニューラルネットワークのすべての特異点を解消することによって、このような状況を改善することを意図している。

### 3. 研究の方法

ニューラルネットワークにおけるすべての特異点を特定し、解消するために次の方法を取る。

#### (1) 特異点の特定

特異点の解消に先立って、まずは解消すべき特異点を洗い出し、特定する。そのために、実験による特定、数理解析による特定という2種類のアプローチを取る。実験による特定では、単純な場合から始めて、最終的には具体的な応用対象(たとえば、医療データ、ソーシャルデータ、金融データ、気象データなど)を絞って、実験を実施する。特定できた特異点は随時、3次元可視化ソフトウェアを用いて、可能なものについては可視化していく。数理解析による特定では、「階層構造に基づかない特異点」を対象にして数理解析手法により特定する。「階層構造に基づいた特異点」は既に理論的に特定されているため(文献 )、対象としない。

#### (2) 特異点の解消

次のように場合分けして特異点を解消する：階層構造に基づいた特異点の場合、階層構造に基づかない特異点の場合。階層構造に基づいた特異点の場合は、研究代表者がニューラルネットワークを複素数に拡張することによって特異点を解消した経験を基にして、クリフォード代数などの代数系を使ってニューラルネットワークをより高次元に拡張することにより特異点の解消を試みる。また、新しいニューロンモデルの考案による解消も試みる。階層構造に基づかない特異点の場合、新たな学習アルゴリズムを考案することにより特異点を解消する。その際、

単一複素ニューロンでの特異点による学習への悪影響を除外した経験(文献 )を手掛かりとする。また、特異点には、ニューラルネットワークの階層構造に起因するものが存在する。そのような特異点は既に文献 において明らかにされている。そこで示された特異点の条件を手掛かりにして、特異点を持たないような新しいニューロンモデルの考案を試みる。さらに、研究代表者は Savitha 誤差関数を使った学習アルゴリズムで、極座標表示された複素ニューロンを学習したところ、特異点の悪影響が除去されることを実験により明らかにした(文献 )。この知見を利用して特異点の解消を試みる。

#### 4. 研究成果

(1) 深層ニューラルネットワークが数多くの「階層構造に基づいた特異点」を持っていることを数理的に明らかにした。「階層構造に基づいた特異点」は中間ニューロン数と層数に応じて増加する。そして、深層ニューラルネットワークがそのような「階層構造に基づいた特異点」を持たないための十分条件を明らかにした。また、求めた十分条件を実現する具体的な手法を2つ考案した。一つは、「階層構造に基づいた特異点」の近傍に入ったら、強制的に抜け出す仕組みを取り入れた学習アルゴリズムである。もう一つは、ネットワークの構造上、「階層構造に基づいた特異点」を持たないニューラルネットワークを用いる手法である。一般に、深層ニューラルネットワークの学習においては、特異点が学習性能に悪い影響を与えている。本研究で得られた成果は、「階層構造に基づいた特異点」を持たない深層ニューラルネットワークを提供するものであり、特に勾配ベースの学習アルゴリズムを用いる場合の学習性能に良い影響をもたらすものと考えられる。

(2) 通常に比べて特異点が少ない深層実数型ニューラルネットワークを提案した。まず、深層複素数型ニューラルネットワークと等価な深層実数型ニューラルネットワークにおいて、いくつかの特異点が複素数に基づく性質により自ずと解消されていることを数理的に示した。そのような深層実数型ニューラルネットワークでは、特異点に起因するローカルミニマムやプラトーンに陥ることが少なくなると考えられる。次に、コンピュータ・シミュレーションにより、極端な非線型性を持つ Two spirals 問題(ベンチマーク問題)の求解を通じて上記結果を検証した。

(3) 線形な深層ニューラルネットワークが(学習性能に悪い影響を与える)悪い局所解(学習誤差の大きい局所解)を持たないことが2016年に証明された。それに対して、非線形な深層ニューラルネットワーク(活性化関数はReLU関数)の場合は、2つの仮定が置かれた上で悪い局所解を持たないことが証明された。そこで、(これまで未解決となっていた)当該2つの仮定が成り立つための十分条件を探索的に調べた。その結果、重み等のパラメータが従う確率分布の確率密度関数が偶関数(縦軸に関して対称)であるならば、当該2つの仮定が満足されることを数理的に証明した。そして、その性質を満足するようなパラメータをもたらす初期化法(even initialization)を提案した。これによって、最急降下法を使う場合は、少なくとも学習開始直後において、悪い局所解が存在しない状態を作り出すことが可能になった。提案した even initialization がもたらすパラメータが取る値は、これまでに発見的に使われてきた一様分布を使う方法、および Kaiming He らが提案した(正規分布を使った)He 初期化法がもたらすパラメータが取る区間の一部に含まれることも明らかにした。

(4) (3)において述べた研究成果を踏まえて、学習開始当初だけでなく、学習のすべての過程に渡って、当該2つの仮定が成り立つことを証明した。証明には Jacot らの研究成果を用いた。つまり、幅が十分広い(つまり、中間ニューロン数が十分多い)深層ニューラルネットワークでは、重み等のパラメータが従う確率分布は初期値の確率分布からほとんど変化することは無いという事実を使って証明した。

(5) (4)の解析対象であった深層ニューラルネットワークは無限個の中間ニューロンを持つ幅が無限であるようなモデルであったが、それを(本質的な特性を保持しつつ)実用に供することのできる程度のモデルに近似する手法等について検討した。その結果、四元数、特に可換四元数を利用するというアイデアに辿り着いた。四元数は構造化された4次元数であるため、良質の特異点を持つと思われる(これは複素ニューラルネットワークの特異点からの類推である)。そこで、可換四元数ニューロンを定式化し、その基本的性質を調べ、可換四元数ニューロンと実ニューロンおよび複素ニューロンとの関係や決定境界等の性質を明らかにした。

また、もう一つのアイデアとして双対数に注目し、双対ニューラルネットワークを定式化し、その基本的性質を調べた結果、剪断に関する汎化能力を持つことを明らかにした。双対数は虚部の基底を2乗すると0になる数であり、良質な特異点を持つと予想される。

さらに、複素数の完全な統計的情報を複素ニューラルネットワークに導入するというアイデアに基づく複素ニューラルネットワークを定式化し、コンピュータ実験を行った結果、これまでの複素ニューラルネットワークが持っている2次元アフィン変換学習能力とは異なった能力を持つことを明らかにした。つまり、縮小変換を学習させただけで、縮小変換に加えて対称変換を汎化する複素ニューラルネットワークが実現された。このような効率的なネットワークは良質な特異点を持つと予想される。

## 引用文献

Hamney, L. G. C.: "XOR has no local minima: a case study in neural network error surface analysis", *Neural Networks*, Vol.11, No.4, pp. 669-682 (1998).

Nitta, T.: "Local minima in hierarchical structures of complex-valued neural networks", *Neural Networks*, Vol.43, pp.1-7 (2013).

Fukumizu, K. and Amari, S: "Local minima and plateaus in hierarchical structures of multilayer perceptrons", *Neural Networks*, Vol.13, No.3, pp.317-327 (2000).

Nitta, T.: "Learning dynamics of a single polar variable complex-valued neuron", *Neural Computation*, Vol.27, No.5, pp.1120-1141 (2015).

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Tohru Nitta, Masaki Kobayashi, Danilo P. Mandic	4. 巻 67
2. 論文標題 Hypercomplex Widely Linear Estimation Through the Lens of Underpinning Geometry	5. 発行年 2019年
3. 雑誌名 IEEE Transactions on Signal Processing	6. 最初と最後の頁 3985 - 3994
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TSP.2019.2922151	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Nitta Tohru	4. 巻 41
2. 論文標題 Resolution of Singularities via Deep Complex-Valued Neural Networks	5. 発行年 2018年
3. 雑誌名 Mathematical Methods in the Applied Sciences	6. 最初と最後の頁 4170 ~ 4178
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/ma.4434	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nitta Tohru, Kuroe Yasuaki	4. 巻 29
2. 論文標題 Hyperbolic Gradient Operator and Hyperbolic Back-Propagation Learning Algorithms	5. 発行年 2018年
3. 雑誌名 IEEE Transactions on Neural Networks and Learning Systems	6. 最初と最後の頁 1689 ~ 1702
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TNNLS.2017.2677446	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 新田 徹	4. 巻 28
2. 論文標題 Resolution of Singularities Introduced by Hierarchical Structure in Deep Neural Networks	5. 発行年 2017年
3. 雑誌名 IEEE Trans. Neural Networks and Learning Systems	6. 最初と最後の頁 2282 ~ 2293
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TNNLS.2016.2580741	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Nitta Tohru	4. 巻 14
2. 論文標題 Proposal of fully augmented complex-valued neural networks	5. 発行年 2023年
3. 雑誌名 Nonlinear Theory and Its Applications, IEICE	6. 最初と最後の頁 175 ~ 192
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/nolta.14.175	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計15件 (うち招待講演 2件 / うち国際学会 6件)

1. 発表者名 Yuto Okawa, Tohru Nitta
2. 発表標題 Learning Properties of Feedforward Neural Networks Using Dual Numbers
3. 学会等名 13th Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) (国際学会)
4. 発表年 2021年

1. 発表者名 大河勇斗、新田徹
2. 発表標題 双対数を用いたニューラルネットワークとその学習特性
3. 学会等名 東北大学 電気通信研究所 共同プロジェクト研究研究会：「高次元・時空間ニューロダイナミクスとそれに基づくシステム構築への展開」
4. 発表年 2022年

1. 発表者名 Tohru Nitta, Hui Hu GAN
2. 発表標題 Fundamental Structure of Orthogonal Variable Commutative Quaternion Neurons
3. 学会等名 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems, SCIS & ISIS2020 (国際学会)
4. 発表年 2020年

1. 発表者名 新田 徹, Hui Hu GAN
2. 発表標題 直交変数可換四元数ニューロンの基本構造
3. 学会等名 東北大学電気通信研究所共同プロジェクト研究研究会：「高次元ニューロダイナミクスとそのニューロハードウェア構築への展開」
4. 発表年 2021年

1. 発表者名 新田 徹
2. 発表標題 双曲勾配オペレータと階層型双曲ニューラルネットワーク
3. 学会等名 福岡大学情報数理セミナー（招待講演）
4. 発表年 2019年

1. 発表者名 新田徹、小林正樹、Danilo, P. Mandic
2. 発表標題 On the Equivalence between Hypercomplex Widely Linear Estimation and its Real Vector Counterpart
3. 学会等名 東北大学電気通信研究所共同プロジェクト研究研究会「高次元ニューロダイナミクスとそのニューロハードウェア構築への展開」
4. 発表年 2018年

1. 発表者名 新田徹
2. 発表標題 Resolution of Singularities via Deep Complex-Valued Neural Networks
3. 学会等名 Empowering Novel Geometric Algebra for Graphics & Engineering Workshop (国際学会)
4. 発表年 2017年

1. 発表者名 新田徹、黒江康明
2. 発表標題 階層型双曲ニューラルネットワークの学習特性
3. 学会等名 東北大学電気通信研究所共同プロジェクト研究研究会「高次元ニューラルネットワークにおける情報表現の最適化」
4. 発表年 2017年

1. 発表者名 新田徹
2. 発表標題 On the Singularity in Deep Neural Networks
3. 学会等名 The 23rd International Conference on Neural Information Processing (国際学会)
4. 発表年 2016年

1. 発表者名 新田徹
2. 発表標題 深層複素ニューラルネットワークの学習特性
3. 学会等名 計測自動制御学会 第9回コンピュテーショナル・インテリジェンス研究会
4. 発表年 2016年

1. 発表者名 新田徹
2. 発表標題 複素ニューラルネットワークによるTwo Spirals Problemの求解
3. 学会等名 東北大学電気通信研究所共同プロジェクト研究研究会「高次元ニューラルネットワークにおける情報表現の最適化」
4. 発表年 2016年



1. 発表者名 新田 徹
2. 発表標題 深層複素ニューラルネットワークの学習特性
3. 学会等名 計測自動制御学会 システム・情報部門 学術講演会
4. 発表年 2016年

1. 発表者名 Tohru Nitta
2. 発表標題 Fully Augmented Complex-Valued Neural Networks
3. 学会等名 2022 International Symposium on Nonlinear Theory and its Applications (NOLTA2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Y. Okawa, S. Kanoga, T. Hoshino, T. Nitta
2. 発表標題 Sequential Learning on sEMGs in Short- and Long-term Situations via Self-Training Semi-Supervised Support Vector Machine
3. 学会等名 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2022) (国際学会)
4. 発表年 2022年

1. 発表者名 新田 徹
2. 発表標題 高次元ニューラルネットワーク
3. 学会等名 東京女子大学学会 数学会 (招待講演)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

Tohru Nitta, "Spurious Local Minima of Deep ReLU Neural Networks in the Neural Tangent Kernel Regime", arXiv preprint, arXiv:1806.04884v3 (2022)

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------