

研究終了報告書

「深層学習の潜在的な正則構造の理解に基づく学習法の安定化と高速化」

研究期間：2019年10月～2023年3月

研究者：二反田 篤史

1. 研究のねらい

深層学習モデルの性能を十分に発揮する為には適切なネットワーク構造の決定、学習アルゴリズムのチューニングが必要であるが、これらの指針となる理論は乏しく膨大な試行錯誤によって補われているという現実がある。深層学習の一試行には多大な時間が掛かることから、結果として甚大な総計算コストが費やされる。今後、深層学習モデルがさらに発展し次世代技術としての立場を確立する為には、深層学習の理論的理解を深めそれに基づき構築された理論保証付きの学習法によってチューニング要素の削減・効率化をする必要がある。このような状況を踏まえ本研究では以下の研究目的を設定する。

目的1：深層学習の理論保証付き学習手法の開発。

目的2：深層ニューラルネットワークのネットワーク構造の自動学習手法の開発。

深層ニューラルネットワークの汎化性能と最適化手法の間に密接な関係性があるという経験的事実にもとづき、本研究では機械学習モデルに対する最適化手法の理論的性質を詳しく調べるというアプローチをとる。具体的には以下の研究項目を設定する。

研究1：深層学習モデルの最適化手法の最適解への収束性（大域収束性）の理解。

研究2：深層学習モデルの最適化手法の潜在的な正則化（陰的バイアス）の理解。

研究3：正則化を陽に利用した最適化手法の開発。

研究4：ネットワーク構造最適化の為の効率的ブースティング法の開発。

研究1・2では深層学習モデルが何故高い性能を発揮するのか、その理由の解明を目指す。そしてその理解に基づく学習法およびネットワーク構造の最適化手法の開発（研究3・4）に取り組む。

2. 研究成果

(1) 概要

深層学習モデルの最適化は非凸問題に帰着するにも関わらず、確率的勾配降下法などの勾配法ベースの手法は最適解に収束する。またその解の性質も様々で中には過学習を引き起こすものも存在する。それにも関わらず回避することが実験的に広く観測されている。この問題は深層学習の理論研究のコミュニティにおける共通課題であり、その解決に向けてニューラルタンジェントカーネル(NTK)やニューラルネットワークの平均場理論の研究が進んだ。本研究もこの両理論の進展に大いに寄与した。具体的には NTK 理論を精緻化し二層ニューラルネットワークを理論上最適な効率で学習可能な条件を与えた。

また平均場理論が解析対象とする平均場ニューラルネットワークはデータへの適応性に優れ、深層学習の成功要因である表現学習を体現するモデルであるが、収束性の解析は困難とされていた。そのような中、本研究では平均場ニューラルネットワークの最適化ダイナミクスに対する凸解析理論を創出することで、その解析手段を与えた。さらにこの理論に基づく効率的最適化

手法を複数開発した。NTK およびこれらの成果は深層学習の最適化ダイナミクスの基礎を与えるものであり、**研究1, 2, 3**に寄与する。

残差ネットワークの構造を関数空間における勾配法とみなすことで、残差ネットワークの勾配ブースティング法を提案し、フランク・ウルフ法に類似した収束証明を与えた。この手法の反復は残差ネットワークに層を逐次追加することで実現されるため、ネットワーク構造そのものの最適化を行う手法とみなせる。従って本成果は**研究4**に貢献する。

機械学習の経験損失最小化問題を高速に求解する分散縮小法を開発した。また平均場ランジュバンダイナミクスと組み合わせることで平均場ニューラルネットワークの高速な最適化も可能であることを証明した。この成果は**研究3**で適切な正則化付き学習法を開発する上で有用である。

(2) 詳細

1. NTK 理論下の高次元二層ニューラルネットワークの最適収束率

ニューラルネットワークの最適化問題はそのパラメータについて非凸最適化であり非常に困難な問題であるが、ニューラルネットワークそのものを変数とすれば、凸最適化問題と見なせる。すなわち、パラメータについての最適化法から関数としての更新則を特定出来れば、この凸性を最大限に活かした理論解析が可能となる。そのため

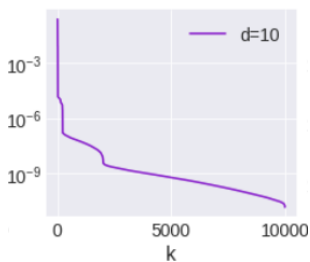


図 2 : NTK の固有値減衰

の道具としてニューラルタンジェントカーネル (NTK) および個人研究が提案した平均場ニューラルネットワークの二つの理論がある。NTK の理論ではニューラルネットワークの横幅を無限大に大きくした極限での勾配法の挙動を NTK が定める再生核ヒルベルト空間内で記述する。本研究(代表的論文[1])では NTK の固有値減衰を考慮に入れた解析により特定条件下で平均化確率的勾配降下法が大域的収束することと汎化誤差についてのミニマックス最適

レートを達成することを示した。これは NTK 理論を精緻化するものであり、**研究1**に貢献するものである。また NTK 理論の研究を通じて、高次元ニューラルネットワークの学習がより効率化されるための条件が見えつつあり、**研究2**に関して重要な知見をもたらすと考えている。

2. 平均場ニューラルネットワークの効率的最適化手法と収束率解析

大域収束性を持つモデルとして平均場ニューラルネットワークがある。このモデルは NTK 理論が対象とするニューラルネットワークに比べて、表現学習の性質を備えることから重要な研究対象になっている。簡易的な実験で平均場ニューラルネットワークのパラメータが真の関数に効率的に収束することが確認できる(図 3)。しかし同時に最適化が難しいモデルでもあり効率的な収束性の担保には強い条件が必要と考えられていた。実際、一般的な条件下における収束効率については不明な部分が多い。本研究では、KL 情報量による正則化を課せば平均場ニューラルネットワークを劣線形の効率性で最適化できることを有限次元の凸最適化法を自然に拡張することで示した(代表的論文[2])。より標準的な最適化法であるノイズ付き勾配降下法に対

しても平均場ランジュバンダイナミクスの理論を通じて収束率解析を行い(代表的論文[3]), さらには分散縮小法との組み合わせにより高速化されることを示した. これらは**研究 1-3**に貢献する成果である.

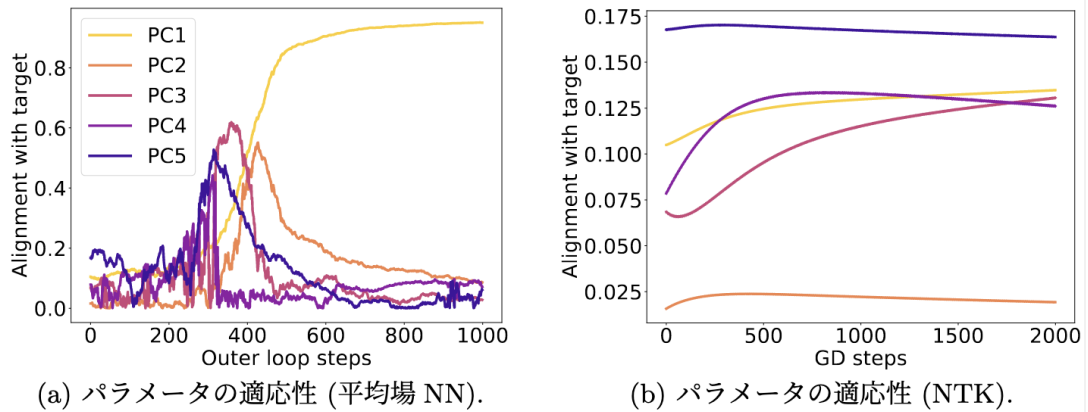


図 3. 真の関数とパラメータの類似度の推移.

3. 関数勾配法に基づく逐次的な残差ネットワークの学習法

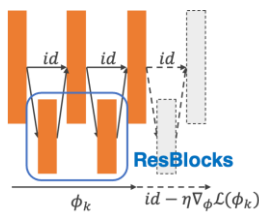


図 1 : ResNet 最適化

残差ネットワークは残差ブロック (あるいはスキップ接続) という層構造を備える深層学習モデルで, とりわけ高いパフォーマンスを発揮することが知られている. 個人研究者は残差ブロックを関数空間における最適化の一ステップとみなすことで関数空間におけるフランク・ウルフ型最適化法が残差ネットワークの構築に適用可能であることを示し, 同時に汎化性能保証もあたえた. 学習後に大域的最適性のための条件の充足性が判定される既存手法に比べると本手法は残差ブロックの凸結合で分類可能なデータセットが学習可能であることが事前に保証される. 提案手法の最適化過程は入力層のノード追加および層の追加で実現されるため (図 1), データに適応的なネットワーク構造最適化手法ともみなすことができ, **研究 4**に貢献するものである.

4. 凸経験損失最小化問題に対する確率的最適化手法の研究

ミニバッチを用いた確率的最適化手法は深層学習を含めた機械学習モデルの代表的学習手法である. ミニバッチは確率的勾配 (経験損失関数の勾配の確率的近似) の分散を縮小させ収束に掛かる反復回数を軽減する目的で用いられる. 本研究では, ミニバッチ法をさらに効率化させた分散縮小法を研究し, 収束に掛かる最小反復回数およびその際必要となる最小ミニバッチサイズを具体的な確率的最適化手法を提案するとともに示した. 分散縮小法は正則化付き経験損失最小化法の高速度を実現する技術であるため, **研究 3**で適切な正則化付き学習法を開発する上で有用である.

3. 今後の展開

平均場ニューラルネットワークの研究を通じて、表現学習能力を引き出すためには最適化のステップサイズを大きくとることが重要であることが分かった。一般の深層学習モデルにおいても大きなステップサイズを用いた平均化確率的勾配降下法が汎化性能に優れた解へ強く誘導することが現在進行中の研究で判明しつつある。この現象に対する理論と実際の実験を通してその有効性を広く伝えることを検討している。成果はすでに出つつあるため2~4年を目処に結果を出したい。またニューラルネットワークの平均場理論はより一般の問題にも適用可能であり潜在的に多くの応用例の存在が期待される。実際、本研究が開拓した平均場理論は他研究グループの貢献もあり拡がりを見せつつある。今後も5~10年にわたり平均場理論およびその応用研究の開拓に取り組む計画である。

4. 自己評価

研究目的の達成状況。

研究1, 2の目標に関しては限定的ではあるがニューラルネットワークに対する NTK 理論および平均場理論を研究することで十分に達成できたと考える。一般のニューラルネットワークへの拡張は今後の課題として残るが、これらの理論によって最適化法が最適解へ収束することが説明できる。また NTK 理論によって汎化の意味で最適性が達成される条件および平均場理論により表現学習が引き起こされる条件が部分的ではあるが解明された。特に後者によって最適化における大きなステップサイズの重要性が見出された。これは**研究3**を今後さらに推進するための新たな研究指針となる重要な知見である。また**研究4**についても残差ネットワークの逐次的最適化法の開発という成果を得た。

2. 研究の進め方(研究実施体制および研究費執行状況)。

2021年度よりさががけ研究補助員を5名雇い、提案手法や理論の実験的検証の補助を依頼した。具体的には本研究費で購入した GPU サーバーの整備やネットワーク環境の構築、および深層学習モデルの大規模実験を分担して行ったことで研究を加速することができた。また各研究補助員が不都合なく作業を進めるために必要な環境整備(コンピュータや周辺機器の整備)にも取り組んだ。

3. 研究成果の科学技術への波及効果。

深層学習モデルに対する大きなステップサイズを用いた平均化確率的勾配降下法の有効性は十分に高く、深層学習の標準的な学習法の一つになる可能性は十分にあると考えている。そのようなことになればその影響の大きさは相当なものである。またニューラルネットワークの平均場理論の応用範囲は潜在的に広く多くの応用事例の発見が期待される。さらに本理論はニューラルネットワークのノイズ付き最急降下法(平均場ランジュバンダイナミクス)に対して従来のランジュバンダイナミクスの凸拡張としての特徴付けを与えることが出来た。それにより非常にシンプルな解析手段を見出せたことは数学的観点からも大きな貢献であると考えている。事実、すでに関連研究者の興味を惹きフォローアップする研究も出てきている。今後この分野が発展していったときその礎として位置付けられるであろう。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数:15件

1. Atsushi Nitanda and Taiji Suzuki. Optimal Rates for Averaged Stochastic Gradient Descent under Neural Tangent Kernel Regime. The 9th International Conference on Learning Representations (ICLR2021), 2021. (Outstanding Paper Award)
NTK を用いることで二層ニューラルネットワークに対する平均化確率的勾配降下法の収束解析を行った。そしてニューラルネットワークは非常に複雑な関数系にも関わらず汎化すること、およびその最適性を特定条件下で理論的に示した。本成果は国際会議 ICLR2021 において Outstanding Paper Awards を受賞した。
2. Atsushi Nitanda, Denny Wu, and Taiji Suzuki. Particle Dual Averaging: Optimization of Mean Field Neural Networks with Global Convergence Rate Analysis. In <i>Advances in Neural Information Processing Systems 34</i> (NeurIPS2021), pp.19608--19621, 2021.
平均場二層ニューラルネットワークは深層学習において重要な表現学習を体現するモデルであるが、その最適化は理論上容易ではないと考えられていた。そこで本研究ではカルバック・ライブラ情報量を正則化として加えることで効率的な最適化が可能となることを具体的な最適化手法の提案とその収束解析を通じて示した。
3. Atsushi Nitanda, Denny Wu, and Taiji Suzuki. Convex Analysis of the Mean Field Langevin dynamics. The 25th International Conference on Artificial Intelligence and Statistics (AISTATS2022), <i>Proceedings of Machine Learning Research</i> , 151:9741--9757, 2022.
確率測度が成す空間上の凸解析に基づき、平均場二層ニューラルネットワークのノイズ付最急降下法に対して従来のランジュバンダイナミクスの凸拡張としての特徴付けを与えた。その凸性を利用し収束性と収束率を明らかにした。

(2) 特許出願

特になし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

学会発表

1. Atsushi Nitanda, Denny Wu, and Taiji Suzuki. Convex Analysis of the Mean Field Langevin Dynamics. Conference on the Mathematical Theory of Deep Neural Networks (DeepMath), San Diego, CA, USA, November 17th, 2022. (selected to 20mins talks)
2. 二反田篤史. 確率的最適化法の収束解析. 第 32 回 RAMP 数理最適化シンポジウム (RAMP 2020), オンライン開催, 2020 年 10 月 27 日.

著作物

3. 二反田篤史. 深層学習の基礎. 深層学習からマルチモーダル情報処理へ, 12-72, サイエンス社, 2022 年 11 月. (著者: 中山英樹, 二反田篤史, 田村晃裕, 井上中順, 牛久祥孝)
4. 二反田篤史. ニューラルネットワークの最適化理論. オペレーションズ・リサーチ,

65(12):643–649, 2020.

5. 二反田篤史. 確率的勾配降下法とニューラルネットワーク. 数理科学「統計的思考法のすすめ」, 58 (9):22–28, サイエンス社, 2020 年 9 月 .