量子ニューラルネットワークにおける Barren Plateau と古典ニューラルネットワークの局所的最小値に関する考察

高橋 太平

2025 10/13

Abstract

現在の量子ニューラルネットワークが直面する『Barren Plateau』という深刻な問題は，期待されている新薬開発や新材料発見といった量子コンピューティングの早期応用を根本から妨げる大きな障害である．既存の Barren Plateau を説明する理論は静的なモデルに対する解析が主であるが，私は経験的に成功していた「適応的アルゴリズム」に BP 問題の解決の糸口があると考えた．モデルは動的であるため，既存理論では理論的な解析が困難であった．学習開始時のみの構造をもとに分析する従来手法では捉えられなかった動的アルゴリズムの性能を捉えるため，私はステップごとに使用できる新しい理論的枠組みを構築した．これにより，これまで誰も手を付けられなかった適応的アルゴリズムの解析が可能になった．この研究で最も困難であった点は，量子 NN の内部に存在するよい学習 landscape を持つ一般化線形モデルの存在の発見である．この発見以前では量子 NN 自体の純粋な理論的な帰結 (量子 NN が常に良い landscape を持つとは限らない) に囚われており研究は一度停滞したが，そこで敢えて元の分野から一歩離れ，古典的な機械学習の基本原理に立ち返って問題の構造を捉え直した．この多角的な視点からの粘り強い思考が，最終的に量子 NN の内部に潜む核心部分の発見に繋がり，困難を突破する原動力となった．この分野の垣根を越えた着眼点が，ブレークスルーの鍵となった．量子コンピュータ実用化には複数の課題が存在するが，この成果は，その中でも学習アルゴリズムの根幹に関わる『勾配消失』というボトルネックを理論的に解消する道筋を示したものであり，今後の応用研究を加速させるための不可欠な一歩である．

1 Introduction

量子ニューラルネットワークの学習不能性，すなわち Barren Plateau (BP) の問題は，量子機械学習のスケールアップを阻む主要な課題であり，その克服に向けた先行研究は (Google Scholar 「Barren Plateau」で検索すれば)37,000 件近く存在する．BP 問題とは，学習中の損失の減少速度が量子ビット数に対して指数関数的に遅くなる問題である (Fig.2 の HEAQNN の学習経過がわかりやすく Barren Plateau に囚われている)．

M. Cerezo や M Larocca らは QNN の BP について，量子最適制御の概念である Dynamical Lie Algebra (DLA) を用いた理論を構築した [1]．彼らは，QNN の勾配の分散が，回路の生成子の Lie 閉包である DLA の次元 $\dim(\mathfrak{g})$ に依存することを示し，問題非依存なアンザッツにおいては，DLA の次元が大きい，すなわち表現力が高いほど，勾配の分散が指数関数的に消失し，BP が発生するというトレードオフの構造を証明した [1]．DLA を用いた研究は盛んに行われ，回路の表現力と学習可能性の間のトレードオフについて知見は得られてきたが，この理論は主にランダム回路や静的な構造に焦点を当てている [1]．これに対し，ADAPT-VQE のような動的手法は，高い表現力と BP 抑制を両立することが経験的に知られている [2, 3]．しかし，従来の DLA ベースの理論では動的なアルゴリズムの BP 抑制を説明することができない．この研究は，勾配消失問題の抑制に成功している古典ニューラルネットワークと失敗している量子ニューラルネットワークの比較を通じて，ユニタリ制約が真の BP の原因であることを明らかにし，ステップごとに適用可能な学習可能性理論の新しい観点を提供する．

2　古典ニューラルネットワークと ”Good Local Minima”

古典的な深層学習の成功の背後には，その損失ランドスケープの特異な性質がある．Kawaguchi et al. (2019) らの研究 [4] によれば，実 NN ではネットワークの幅や深さを増大させることで，非漸近的な領域においても損失関数の局所的最小値の質が向上し，それらの値は大域的最小値に近づくことが知られている．本稿では，このようにパラメータ数の増加に伴って損失値が改善する性質を持つ局所的最小値を「Good Local Minima」と定義する．実古典 NN における (鞍点を除く) 微分可能な局所的最小値 θ における損失 $L(\theta)$ は，以下の式で上から抑えられる．[4]

$$L(\theta) \leq \frac{1}{2}\|Y\|^2 - \frac{1}{2}\sum_{l=1}^{H+1}\sum_{k_l=1}^{d_l}\|P[N_{k_l}^{(l)}]\mathrm{vec}(Y)\|^2$$

この式は，ネットワークの表現力（幅 d_l や深さ H）が増すほど，損失 $L(\theta)$ が 0 に近づくことを意味している．つまり，古典 NN は Good Local Minima しか持たない [4]．これは，古典 NN の表現空間が広大であるため，ほとんどの最適解を表現できるためである [4]．この性質から，古典 NN は容易に経験誤差を最小化できる [4]．一方で，量子 NN は経験誤差の最小化も困難である．それは後述する Barren Plateau が原因である．

3　量子NNにおける ”Bad Local Minima” としてのBarren Plateau

3.1　制約付きNNとしての量子NN

QNN（量子 NN）は，その構造上，特殊な制約と活性化関数を用いる複素古典 NN と見なすことができる．古典 NN の最適化が一般的な行列空間 $\mathbb{C}^{N\times N}$ での探索と見なせるのに対し，量子 NN の最適化は，その部分多様体であるユニタリ行列の空間 $U(N)$ 上での探索に限定される．これが両者の最大の違いである．

3.2　”Bad Local Minima” の起源とBarren Plateau

このユニタリ性という強い制約が，量子 NN 特有の「Bad Local Minima」を生み出している．Bad Local Minima を，パラメータを増やしてもその値が改善されない，あるいは最適化が停滞する局所解であると定義する．(Good local minima は表現力の増加で改善する局所解である) ここで，QNN におけるパラメータ θ_μ に関する勾配は，古典 NN が生成する Good Local Minima へと誘導する理想的な勾配方向 (一般的な行列空間での勾配 $\frac{\partial L}{\partial U}$) を，ユニタリ多様体上の移動可能な方向（$\frac{\partial U}{\partial \theta_\mu}$）へ射影したものに等しい．(勾配の内積構造)

$$\frac{\partial L}{\partial \theta_\mu} = \left\langle \frac{\partial L}{\partial U}, \frac{\partial U}{\partial \theta_\mu} \right\rangle_F$$

Bad Local Minima（Barren Plateau）は，Good local minima へと誘導する理想的な勾配 $\frac{\partial L}{\partial U}$ がゼロではないにもかかわらず，探索可能な全ての方向と直交してしまう状況に対応する．

$$\frac{\partial L}{\partial U} \neq 0 \quad \wedge \quad \left\langle \frac{\partial L}{\partial U}, \frac{\partial U}{\partial \theta_\mu} \right\rangle \approx 0 \quad (\forall \mu)$$

高次元空間では，ランダムなベクトル同士がほぼ直交する確率は非常に高い．ユニタリ制約によって探索空間が限定された結果，量子 NN はこのような「進むべき方向に進めない」という Bad Local Minima に陥りやすくなる．これが，広大なパラメータ空間で勾配が消失する Barren Plateau 現象の本質である．

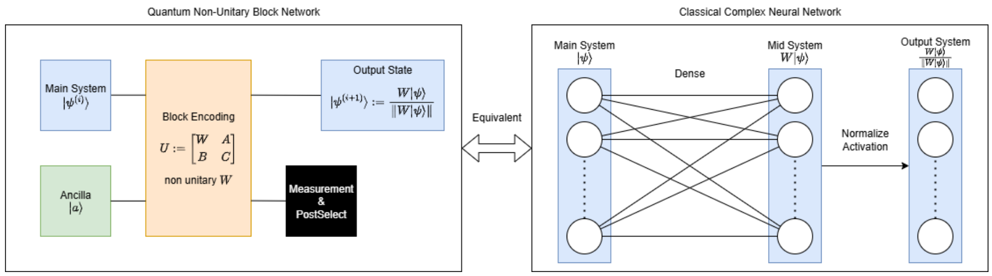

Figure 1: 複素ニューラルネットワーク (右図) と，それと等価な量子ネットワーク (左図)—左の量子 NN から Ancilla 系と postselect を削除した場合 (i.e. ユニタリ制約を付加した場合) 標準的な量子 NN になり，またそれと等価な古典複素 NN も存在する．

3.3 動的なアルゴリズムの評価

ここで動的なアルゴリズムとは U をいくつかの候補から選択し，勾配が大きくなるように変化させていくことができるアルゴリズムを指す．もし U の候補が十分に広く，Bad Local Minima を脱出するために十分なほど $\frac{\partial U}{\partial \theta_\mu}$ を制御できるなら BP を回避できる．実際にはランダムな向きを向く $\frac{\partial L}{\partial U}$ に対してどの程度 overlap を確保できるかで評価する．

4 複素正規化活性化 NN での "Good Local Minima"

4.1 正規化層を持つ非ユニタリ量子 NN

実は前節での議論は不十分である．なぜならば，Kawaguchi[4] らの議論は ReLu 関数を活性化関数とする実 NN でのみ成り立つからである．そこでこの節では量子 NN 内部に存在する古典複素 NN が Good Local Minima へと常に誘導することを示す．各層の演算子を一般の演算子とし，各層の後に状態ベクトルのノルムを 1 に正規化する活性化層を導入したモデルを考える (Fig.1)．いかなる量子 NN もこの複素 NN モデルに対してユニタリ制約を施したものであると考えることができる．このモデルは，以下の定理によって，ReLu を活性化関数とする実古典 NN と同様に Good Local Minima しか持たないことが数学的に示される．

Theorem 4.1. *(Informal)* H 個の非ユニタリ行列 $W := \{W^{(i)}\}_H$ と，各層の後に状態ベクトルのノルムを *1* に正規化する層 $\Lambda(\psi) := \psi/\|\psi\|$ を持つ量子ニューラルネットワークを考える．このネットワークの出力状態を $\psi(W) = \Lambda(\psi^{(H)})$ (ただし $\psi^{(H)} := W^{(H)}\Lambda(\psi^{(H-1)})$) とし，損失関数を $L(W) = \frac{1}{2}\|\psi(W) - \psi_{obj}\|^2$ とする．このとき，微分可能な任意の局所的最小値 W は *"Good Local Minima"* である．すなわち，その損失値 $L(W)$ は，ネットワークの表現力（層の数 H）の増加に伴って大域的最小値（ゼロ）に近づく．

Proof. (概要) 局所的最小値（勾配ゼロ）の条件から，出力状態 $\psi(W)$ が，目標状態 ψ_{obj} をネットワークが表現可能な空間（接空間への射影を含む）に射影したものと一致することが導かれる．

$$\text{vec}(\psi(W)) = P_{\text{tangent}}\left(P_{\text{span}}(\text{vec}(\psi_{\text{obj}}))\right)$$

これにより，損失は目標状態の非射影部分のノルムで記述される．

$$L(W) = \frac{1}{2}\|\psi(W) - \psi_{\text{obj}}\|^2 = \frac{1}{2}\|(I - P_{\text{combined}})\text{vec}(\psi_{\text{obj}})\|^2$$

ネットワークの表現力が増加すると，射影空間（P_{combined}）が拡大し，損失は単調に減少し，ゼロに近づく．したがって，このモデルは Bad Local Minima を持たない． □

この定理は，量子 NN におけるユニタリ制約が，Barren Plateau の原因となる Bad Local Minima を生成し，古典 NN のような望ましい損失ランドスケープ特性を阻害することを理論的に裏付けている．

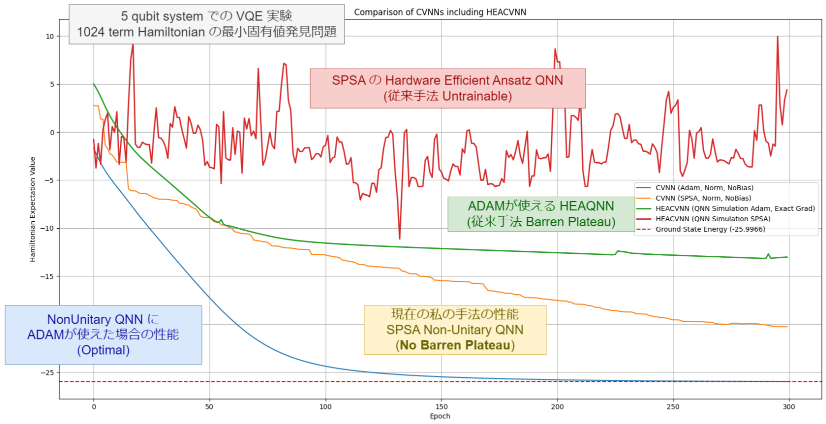

Figure 2: CVNN と QNN 手法の収束比較 — 青線や黄色で示された，本節で議論した non-unitary モデルに相当する手法は，最適化の途中で停滞する従来手法 (赤・緑) とは対照的に，速やかに基底状態エネルギー（破線）へと収束している．これは，ユニタリ制約の緩和が Bad Local Minima（最適化の停滞）を効果的に回避していることを強く示唆している．ただし，指数関数的な収束のためには ADAM のような効率的な Optimizer を採用する必要性があることも同時にわかる

4.2 実験的検証

この理論的帰結は，図 2 の実験結果によって明確に支持されている．

5 結論

本稿では，QNN における Barren Plateau 問題を，ユニタリ制約を持つ特殊なニューラルネットワークという視点から再解釈した．この視点に基づけば，Barren Plateau とは，古典 NN では存在しない「Bad Local Minima」が，ユニタリ制約によって出現したものであると結論付けられる．この問題の核心がユニタリ性の制約にあることを踏まえると，非ユニタリな演算と測定による正規化層を組み合わせたモデルが，Bad Local Minima を回避し，QNN の訓練可能性を向上させるための本質的な解決策であることが示された．そして，適応的なアルゴリズムは Bad Local Minima を直接抜け出す方法を持つことも示された．

References

[1] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo. A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications*, 15(1):1–10, 2024.

[2] H. R. Grimsley, N. J. Mayhall, S. E. Economou, and E. Barnes. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, 10(1):3007, 2019.

[3] H. R. Grimsley, G. S. Barron, E. Barnes, S. E. Economou, and N. J. Mayhall. ADAPT-VQE is insensitive to rough parameter landscapes and barren plateaus. *npj Quantum Information*, 9(1):19, 2023.

[4] K. Kawaguchi, J. Huang and L. P. Kaelbling, Effect of Depth and Width on Local Minima in Deep Learning, *Neural Computation*, vol. 31, no. 7, pp. 1462-1498, July 2019