

量子機械学習におけるデータ符号化由来の勾配消失に関する研究

東京大学 理学系研究科 物理学専攻 寺師研究室
上曾山 健介

2024/01/24

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンプラトール (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンプラトール (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

1

前提知識

- 量子コンピューター
- 変分量子アルゴリズム
- 量子機械学習
- バレンプラトール (BP)

2

研究概要

3

コスト関数の勾配の分散の上界

4

コスト関数の勾配の分散の下界

5

誤差関数の形とコスト関数の勾配の分散

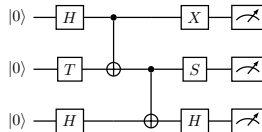
6

まとめ

量子コンピューター

量子計算を行う量子回路は以下の要素から成る。

- 量子ビット：2 準位量子系
- 量子ゲート：量子ビットの状態を変化させるユニタリ
- 測定：量子状態から情報を読み出す操作



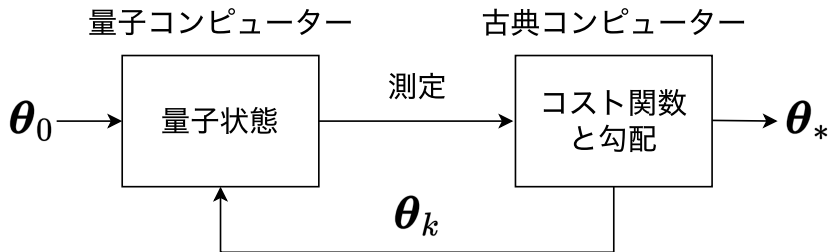
NISQ (Noisy Intermediate-Scale Quantum device) [Preskill2018]

- 数百量子ビット程度の規模の量子コンピューター
- 誤り訂正ができないため、ノイズを無視できない
- 使用できる量子回路の深さに限界がある

変分量子アルゴリズム

[review:Cerezo2021]

- NISQ で実行可能な量子&古典ハイブリッドアルゴリズム
- $U(\theta)$: パラメーター化された量子回路 (変分量子回路)
- $\ell_{i,k}(\theta) = \text{Tr}[U(\theta)\rho_i U^\dagger(\theta)O_k]$ 、 ρ_i : 初期状態、 O_k : オブザーバブル
- コスト関数 $\mathcal{L}(\theta) = f(\{\ell_{i,k}(\theta)\}_{i,k})$ の最適化
- 量子化学計算、組み合わせ最適化、機械学習などに応用



ここでは、変分量子回路を用いた機械学習を量子機械学習 (Quantum Machine Learning) と呼ぶ。教師あり量子機械学習ではデータセット $\{x_i, y_i\}_i$ の入力回路が必要となる。

教師あり学習のコスト関数 (量子回路学習モデル [Mitarai2018])

試行関数

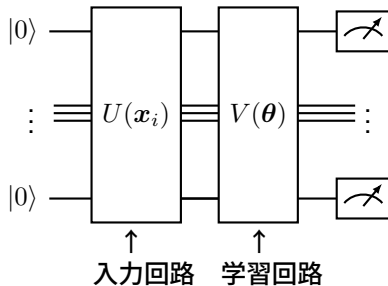
$$|\phi(x_i, \theta)\rangle := V(\theta)U(x_i)|0\rangle^{\otimes n}$$

予測ラベル

$$\ell_i(\theta) := \langle \phi(x_i, \theta) | O | \phi(x_i, \theta) \rangle$$

コスト関数

$$\mathcal{L}(\theta) := \frac{1}{N} \sum_{i=1}^N f(y_i, \ell_i(\theta))$$



バレンプラトー (Barren Plateau, BP)

7/27

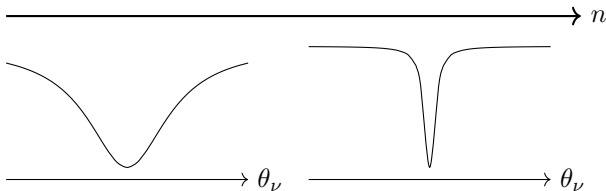
バレンプラトー [Mcclean2018]

定義

$\mathcal{L}(\theta)$: コスト関数、 $V(\theta)$: 学習回路、 n : 量子ビット数

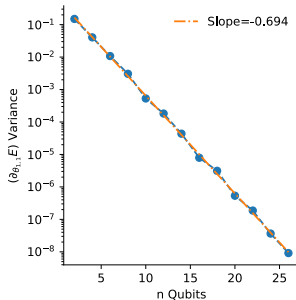
$$\mathbb{E}_{V(\theta)} \left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta_\nu} \right] = 0, \quad \text{Var}_{V(\theta)} \left[\frac{\partial \mathcal{L}(\theta)}{\partial \theta_\nu} \right] \in \mathcal{O}(2^{-\alpha n}), \quad \alpha > 0$$

勾配消失 → コスト関数の測定回数が指数関数的に増加
勾配の分散のスケールが計算量に影響



原因

- 学習回路の深さ
- オブザーバブルの局所性
- ノイズ
- データの入力



コスト関数の勾配の分散

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンプラトール (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンプラトール (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

研究背景

- 変分量子アルゴリズムによって機械学習を効率化したい
- しかし、バレンプラトール（勾配消失）が起きうる
- また、データの入力を与える影響は詳しく調べられていない

研究目的と方針

目的：データ入力によってバレンプラトールが生じないようにする

方針：① データ入力とコスト関数の勾配の分散の関係を解析する（新規性）
具体的には、コスト関数の勾配の分散の上界と下界を導出する。

② 勾配の分散のスケーリングが誤差関数の形に依存しないことを数値検証

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンプラトール (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

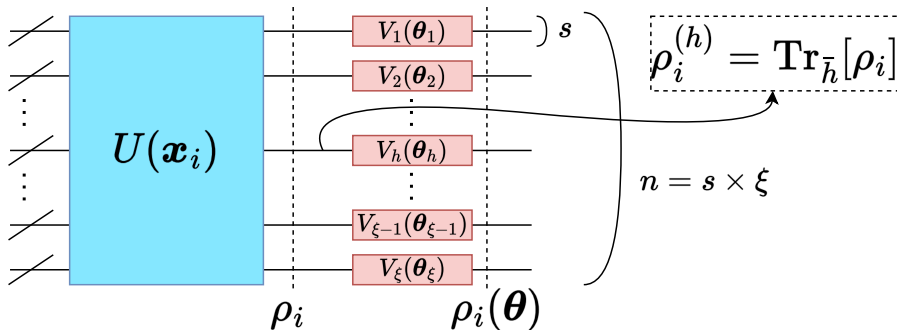
誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンプラトール (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

コスト関数の勾配の分散の上界：設定

11/27



$$\ell_i(\theta) = \text{Tr}[\rho_i(\theta) O_L] \in [0, 1], \quad O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$$

データ入力以外の要因でバレンブラトーが起きないように学習回路や O_L を設定。

コスト関数の勾配の分散の上界：結果

12/27

$$y_i \in \{0, 1\}, \quad \ell_i(\boldsymbol{\theta}) := \text{Tr}[\rho_i(\boldsymbol{\theta}) O_L] \in [0, 1], \quad \mathcal{L}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{i=1}^N f(y_i, \ell_i(\boldsymbol{\theta}))$$

先行研究 [Thanasilp2021] を基に、新たな上界を与えた。

定理

量子機械学習において、前述の量子回路を用い、パラメータをランダムにサンプリングしたとき ($V_j(\theta_j)$ がユニタリ 2-デザイン)、コスト関数の勾配の分散の上界は次のように与えられる。ただし、 $\mathbb{U} := \{U(x) | x \in \mathcal{X}\}$

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_{\theta_\nu} \mathcal{L}(\boldsymbol{\theta})] \leq A_f \times r_{n,s} \times \overline{D}_{\text{HS}}^s \leq A_f \times r_{n,s} \times \left(\frac{2^s - 2^{-s}}{2^n + 1} + \epsilon_{\mathbb{U}} \right)$$

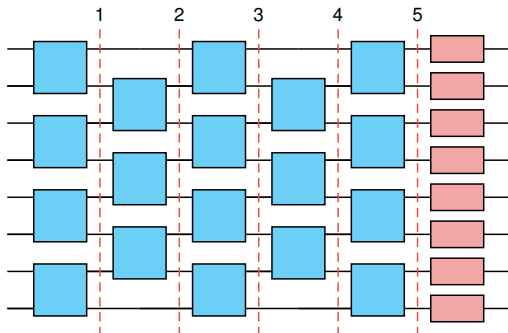
- A_f は二乗誤差などの誤差関数 f に依る項
- $r_{n,s}$ は測定するオブザーバブルに依る項
- $\overline{D}_{\text{HS}}^s := \int_{\mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s)$ はデータ入力に依る項
- $\epsilon_{\mathbb{U}}$ は入力回路の表現能力の指標であり、表現能力が高いほどこの値は小さい。

$\epsilon_{\mathbb{U}}$ が 0 にならない仮定の下、データ入力に依る項 $\overline{D}_{\text{HS}}^s$ が指数的に落ちない条件を調べた。

コスト関数の勾配の分散の上界：具体的な入力回路

13/27

入力回路 $U(x)$ に次の Alternating Layered Ansatz (ALT) を仮定して、 $\overline{D}_{\text{HS}}^{s=1}$ を解析計算 *



青は入力回路、赤は学習回路

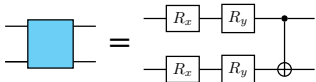
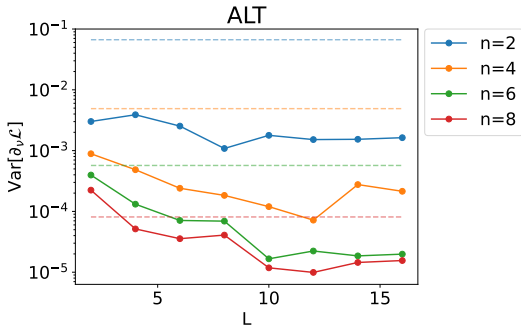
学習回路としての ALT の BP への影響は先行研究 [Cerezo2021] で調べられた。

(*青いブロックがユニタリ 2-デザインとして、Random Tensor Network Integrator (RTNI) [Fukuda2019] を用いて計算)

コスト関数の勾配の分散とその上界の比較

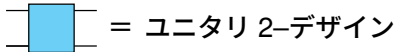
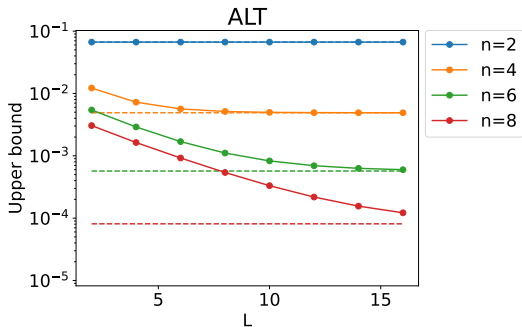
14/27

$$\text{Var}_V(\theta)[\partial_{\theta_\nu} \mathcal{L}(\theta)]$$



(アヤメのデータセットを用いてコスト関数を定義)

$$\text{上界} (A_f \times r_{n,s} \times \overline{D}_{\text{HS}}^{s=1})$$



確かに上界の役割を果たしている

層数の必要条件

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンブラトー (BP)

研究概要

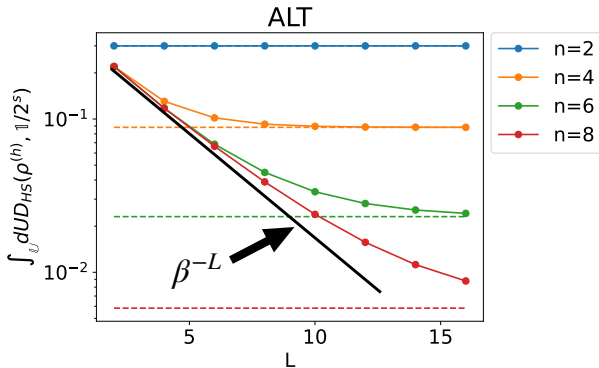
コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

上界が量子ビット数に対して指数関数的に落ちないための層数の設定



($0 < \alpha$, n : 量子ビット数)

$$\overline{D}_{HS}^s \propto e^{-\alpha n} \rightarrow \text{BP}$$

BP が起きない必要条件:

($0 < \gamma$, $1 < \beta$)

$$n^{-\gamma} \leq \beta^{-L} \leq \overline{D}_{HS}^s$$

$$\Rightarrow L \leq \frac{\gamma}{\log \beta} \log n$$

(BP: バレンブラトー)

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンブラトー (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンブラトー (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

コスト関数の勾配の分散の下界：設定

17/27

(解析のため単純な入力回路を仮定)

量子回路：

$n = s \times \xi$ 量子ビット

R_y ゲートからなる入力回路

コスト関数 (2 値分類)：

$$\mathcal{L}_{\text{MAE}}(\theta) = \frac{1}{N} \sum_{i=1}^N |\ell_i(\theta) - y_i|$$

入力データ：

$\mathcal{X} = \{x\}$ (ラベル 0)

$\mathcal{Z} = \{z\}$ (ラベル 1)

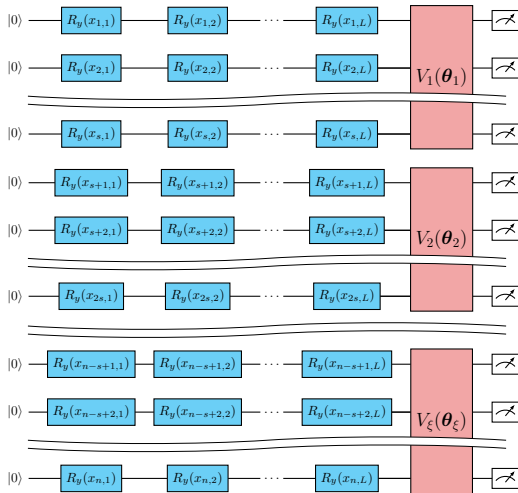
$|\mathcal{X}| : |\mathcal{Z}| = p : q$ ($p + q = 1$)

ガウス分布：

$$x_{j,d} \sim \mathcal{N}(\mu_{x|j,d}, \sigma_{x|j,d}^2)$$

$$z_{j,d} \sim \mathcal{N}(\mu_{z|j,d}, \sigma_{z|j,d}^2)$$

分散： $\sigma_{x|j,d}, \sigma_{z|j,d} \leq \sigma_{\max}$



L 層の R_y ゲートからなる入力回路と
 ξ 個の s 量子ビットユニタリからなる $V_x(\theta_x)$ 学習回路

コスト関数の勾配の分散の下界：結果

18/27

定理

絶対誤差のコスト関数の勾配の分散は次のように下から抑えられる。

$$\frac{r_{n,s}}{2^s} \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \leq \text{Var}_{V(\theta)} [\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)]$$

ただし、 $r_{n,s} := \frac{s 2^{3(s-1)}}{n^2 (2^{2s}-1)^2}$, $\Sigma_{x|j} = \sum_{d=1}^L \sigma_{x|j,d}^2$, $\Sigma_{z|j} = \sum_{d=1}^L \sigma_{z|j,d}^2$

$|\mathcal{X}| : |\mathcal{Z}| = 1 : 1 \iff p = q = 1/2$ のとき、分散の下界は

$$\frac{r_{n,s}}{2^{s+2}} \sum_{j=1}^s \left(e^{-\Sigma_{x|j}/2} - e^{-\Sigma_{z|j}/2} \right)^2 \leq \text{Var}_{V(\theta)} [\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)]$$

- 下界において分散に依存する $e^{-\Sigma_{x|j}/2}$, $e^{-\Sigma_{z|j}/2}$ の差が重要
- しかし、 $e^{-\Sigma_{x|j}/2}$, $e^{-\Sigma_{z|j}/2}$ は層数に関して指数関数的に小さくなる
- すべての j に対し $e^{-\Sigma_{x|j}/2} = e^{-\Sigma_{z|j}/2}$ である場合、下界は 0 となる

コスト関数の勾配の分散の下界：結果

19/27

定理

絶対誤差のコスト関数の勾配の分散は次のように下から抑えられる。

$$\frac{r_{n,s}}{2^s} \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \leq \text{Var}_{V(\theta)} [\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)]$$

ただし、 $r_{n,s} := \frac{s 2^{3(s-1)}}{n^2 (2^{2s} - 1)^2}$, $\Sigma_{x|j} = \sum_{d=1}^L \sigma_{x|j,d}^2$, $\Sigma_{z|j} = \sum_{d=1}^L \sigma_{z|j,d}^2$

$|\mathcal{X}| : |\mathcal{Z}| = 1 : 0 \iff p = 1, q = 0$ のとき、分散の下界は

$$\frac{r_{n,s}}{2^s} e^{-L\sigma_{\max}^2} \leq \frac{r_{n,s}}{2^s} \sum_{j=1}^s e^{-\Sigma_{x|j}} \leq \text{Var}_{V(\theta)} [\partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\theta)]$$

- $L\sigma_{\max}^2$ が小さいほど、分散の下界は大きくなる
- $s, L\sigma_{\max}^2 \in \mathcal{O}(\log n)$ である場合、分散の下界は $\mathcal{O}(1/\text{poly}(n))$ となる
→ バレンブラトー回避の十分条件

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンプラトール (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンプラトール (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

誤差関数の形とコスト関数の勾配の分散

21/27

誤差関数の形はコスト関数の勾配の分散のスケーリングにどのように影響するか？

$$\mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N |\ell_i(\boldsymbol{\theta}) - y_i|$$

$$\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N (\ell_i(\boldsymbol{\theta}) - y_i)^2$$

$$\mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N [-y_i \log \ell_i(\boldsymbol{\theta}) - (1 - y_i) \log (1 - \ell_i(\boldsymbol{\theta}))]$$

$$\Rightarrow \partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_{\theta_\nu} \ell_i(\boldsymbol{\theta})$$

$$\partial_{\theta_\nu} \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N 2|\ell_i(\boldsymbol{\theta}) - y_i| \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_{\theta_\nu} \ell_i(\boldsymbol{\theta})$$

$$\partial_{\theta_\nu} \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \frac{1}{1 - |\ell_i(\boldsymbol{\theta}) - y_i|} \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_{\theta_\nu} \ell_i(\boldsymbol{\theta})$$

ただし、 $y_i \in \{0, 1\}$, $\ell_i(\boldsymbol{\theta}) = \text{Tr}[\rho_i(\boldsymbol{\theta}) O_L] \in [0, 1]$, $(O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbf{1}_{\bar{j}})$

誤差関数の形とコスト関数の勾配の分散

22/27

学習回路がユニタリ 2-デザインであると仮定した場合の $\ell_i(\boldsymbol{\theta})$ の平均と分散：

$$\begin{aligned} \mathbb{E}_{\mathcal{U}(d)}[\ell_i(\boldsymbol{\theta})] &= \frac{1}{2} \\ \text{Var}_{\mathcal{U}(d)}[\ell_i(\boldsymbol{\theta})] &= \frac{1}{4n(2^n + 1)} \end{aligned}$$

そこで、すべての i について $\ell_i(\boldsymbol{\theta}) \sim \frac{1}{2} \iff |\ell_i(\boldsymbol{\theta}) - y_i| \sim \frac{1}{2}$ と仮定すると、

$$\partial_{\theta_\nu} \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) \sim \partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}), \quad \partial_{\theta_\nu} \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) \sim 2 \partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})$$

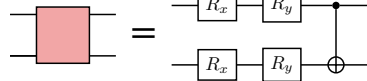
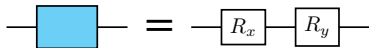
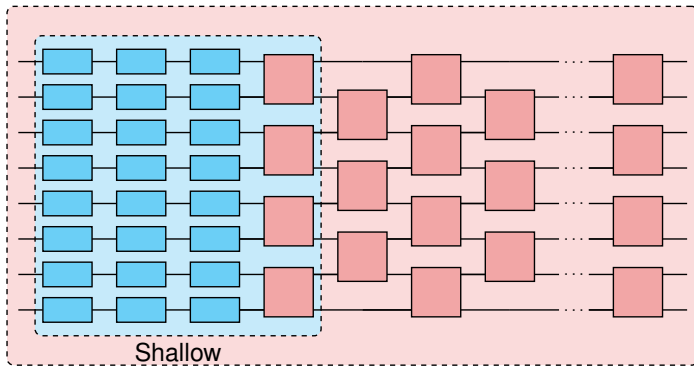
となる。したがって、二乗誤差と絶対誤差、交差エントロピー誤差と絶対誤差の勾配の分散の比は次のように近似される。

$$\begin{aligned} \implies \text{Var}_{\mathcal{U}(d)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})] / \text{Var}_{\mathcal{U}(d)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] &\sim 1 \\ \text{Var}_{\mathcal{U}(d)}[\partial_{\theta_\nu} \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta})] / \text{Var}_{\mathcal{U}(d)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] &\sim 4 \end{aligned}$$

学習回路が深い場合、誤差関数によらず勾配の分散のスケーリングは同程度と推測される。

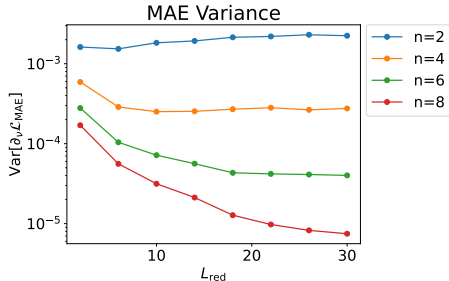
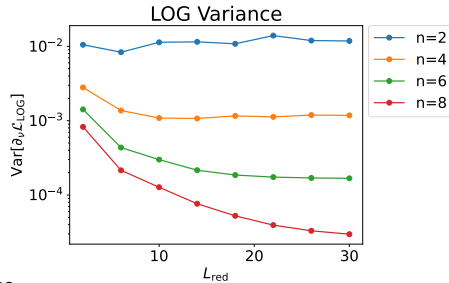
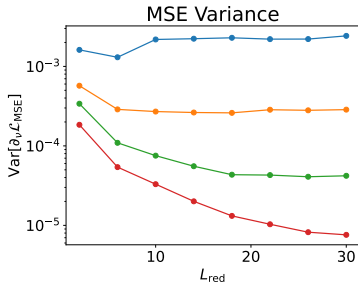
誤差関数の形とコスト関数の勾配の分散：具体的な回路 23/27

学習回路に ALT を仮定して、その層数 L_{red} を変化させたときのコスト関数の勾配の分散の比を調べる。(青は入力回路、赤は学習回路)



誤差関数の形とコスト関数の勾配の分散

24/27

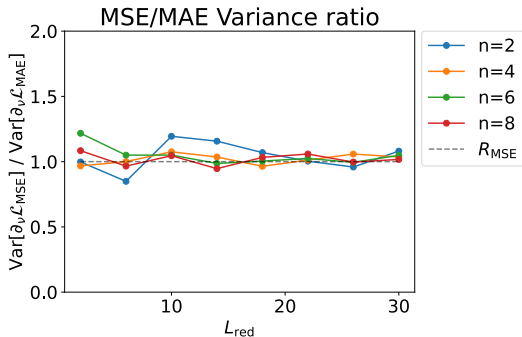
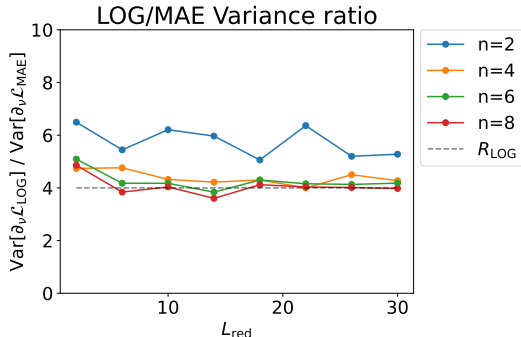


誤差関数の形とコスト関数の勾配の分散

25/27

$$\text{Var}_{V(\theta)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MSE}}(\theta)] / \text{Var}_{V(\theta)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\theta)]$$

$$\text{Var}_{V(\theta)}[\partial_{\theta_\nu} \mathcal{L}_{\text{LOG}}(\theta)] / \text{Var}_{V(\theta)}[\partial_{\theta_\nu} \mathcal{L}_{\text{MAE}}(\theta)]$$

破線は近似比 $R_{\text{MSE}} = 1$ を表す破線は近似比 $R_{\text{LOG}} = 4$ を表す

数値計算では、学習回路が浅くても誤差関数によらず勾配の分散のスケージングはほぼ同じ
(アヤメのデータセットを用いてコスト関数を定義した。)

目次

前提知識

量子コンピューター

変分量子アルゴリズム

量子機械学習

バレンプラトール (BP)

研究概要

コスト関数の勾配の分散の上界

コスト関数の勾配の分散の下界

誤差関数の形とコスト関数の勾配の分散

まとめ

- 1 前提知識
 - 量子コンピューター
 - 変分量子アルゴリズム
 - 量子機械学習
 - バレンプラトール (BP)
- 2 研究概要
- 3 コスト関数の勾配の分散の上界
- 4 コスト関数の勾配の分散の下界
- 5 誤差関数の形とコスト関数の勾配の分散
- 6 まとめ

バレンプラトールに対するデータ入力の影響の研究

- コスト関数の勾配の分散の上界をデータ入力の観点から導出し、数値的にも検証
- データ入力後の状態のエンタングルメントや入力回路の表現能力が大きい
→ 上界が小さくなり、バレンプラトールにつながる
- データ入力の層数が $O(\log n)$ なら、上界のスケーリングは指数関数的ではない
- 入力データがガウス分布に従う場合、データの分散が勾配の分散の下界において重要
- 勾配の分散のスケーリングは、誤差関数の形にほぼ依存しないことを数値的に確認

今後の課題

- より一般的な入力回路において、コスト関数の勾配の分散の下界を調べる。
- 汎化性能の観点からも入力回路について考慮する。(今回は訓練可能性のみ)
- 古典計算が計算量的に難しい入力回路になっているか

ローカルオブザーバブル

- $Z_1 := Z \otimes \mathbb{1} \otimes \mathbb{1} \otimes \cdots \otimes \mathbb{1}$
- $|0\rangle\langle 0|_1 = (Z_1 + \mathbb{1}_1)/2$
- $O_L := \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$ (ローカルオブザーバブルの線形和)
- 学習回路の深さが $\mathcal{O}(\log n)$ 程度ならバレンプラトーは起きない。

グローバルオブザーバブル

- $Z^{\otimes n} := Z \otimes Z \otimes \cdots \otimes Z$
- 学習回路の深さによらずバレンプラトーを引き起こす。

定義

$P_{t,t}(U)$ は、ユニタリ U と U^\dagger のそれぞれの成分に関して最大次数 t の斉次多項式とする。 K 個のユニタリの集合 $\{U_k\}$ が以下の条件を満たすとき、この集合はユニタリ t -デザインを成すという。(例： $P_{t,t}(U) = U^\dagger A U B U^\dagger C U$)

$$\frac{1}{K} \sum_{k=1}^K P_t(U_k) = \int_{\mathcal{U}(d)} P_t(U) d\mu(U)$$

- パウリ群： $\mathcal{P}(n) = \{e^{i\frac{k\pi}{2}} P_{j_1} \otimes \cdots \otimes P_{j_n} | k, j_l = 0, 1, 2, 3\}$ はユニタリ 1 デザイン
- クリフォード群： $\mathcal{C}(n) = \{U \in \mathcal{U}(2^n) | U P U^\dagger = \mathcal{P}(n)\}$ はユニタリ 3 デザイン
- もし $\{U_k\}$ がユニタリ t -デザインならば、ユニタリ $(t-1)$ -デザインでもある。

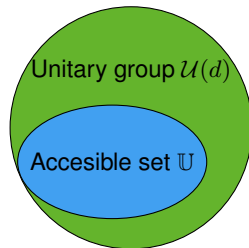
- (データ入力なしの場合) ローカルオブザーバブルを用いることより、学習回路の深さが $\mathcal{O}(\log n)$ であればバレンプラトーは起きないことが示されている。
- 学習回路の各層ごとにパラメーターを最適化する。これにより、実質的な学習回路の深さを減らすことができる。
- 一部の学習パラメーターの初期値を残りの回路と打ち消し合うように設定する。これにより、実質的な学習回路の深さを減らすことができる。
- 学習パラメーター間に相関を持たせる。これにより、学習回路の表現能力を抑えることができる。
- 学習パラメーターの初期値を、一様分布ではなく、正規分布からサンプリングする。

量子回路の表現力とは次のように定義される。

$$\epsilon_{\mathbb{U}}^{(t,p)}(X) := \left\| A_{\mathbb{U}}^{(t)}(X) \right\|_p,$$

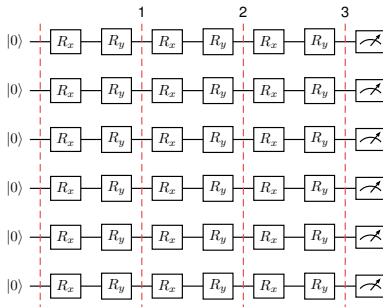
$$A_{\mathbb{U}}^{(t)}(X) := \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V) V^{\otimes t} X^{\otimes t} (V^{\otimes t})^\dagger - \int_{\mathbb{U}} dU U^{\otimes t} X^{\otimes t} (U^{\otimes t})^\dagger.$$

- $\|\cdot\|_p$ ：シャッテン p -ノルム
- X ：初期状態 $|0\rangle\langle 0|^{\otimes n}$
- $\mathcal{U}(d)$ ：次元 d のユニタリ群
- $t = 2$ の場合を考える

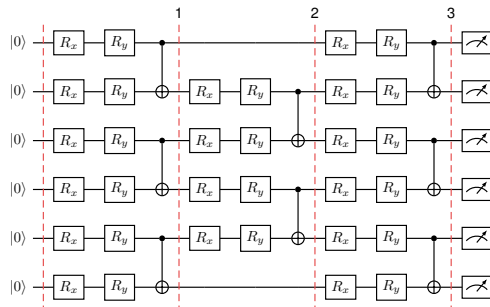


量子回路の表現領域

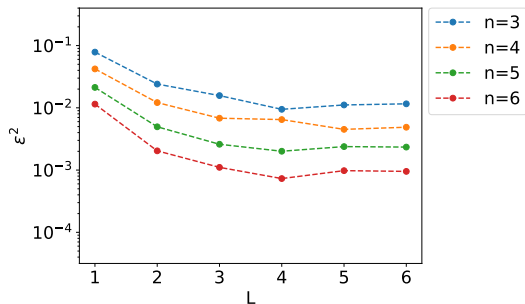
Tensor Product Ansatz の表現力



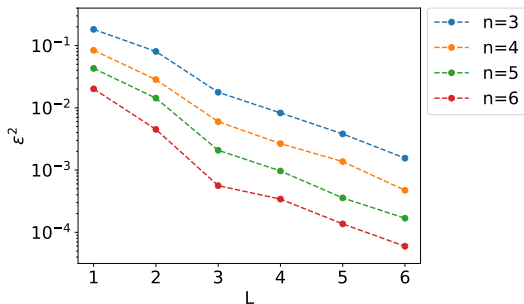
Alternating Layered Ansatz の表現力



Tensor Product Ansatz の表現力



Alternating Layered Ansatz の表現力



$$y_i \in \{0, 1\}, \quad \ell_i(\boldsymbol{\theta}) := \text{Tr}[\rho_i(\boldsymbol{\theta}) O_L] \in [0, 1], \quad \mathcal{L}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{i=1}^N f(y_i, \ell_i(\boldsymbol{\theta}))$$

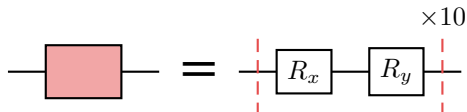
定理

量子機械学習におけるコスト関数の勾配の分散の上界

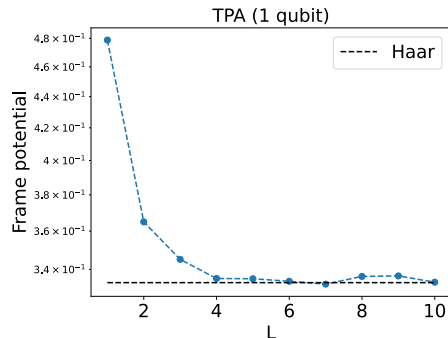
$$\begin{aligned} & \text{Var}_{\boldsymbol{\theta}}[\partial_{\theta_\nu} \mathcal{L}(\boldsymbol{\theta})] \\ & \leq A_f \times r_{n,s} \times \overline{D}_{\text{HS}} \\ & = 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i(\boldsymbol{\theta})} f)^2] \times \frac{2^{2s-1}}{(2^{2s} - 1)^2} \left(\text{Tr}[(O_L^h)^2] - \frac{\text{Tr}[O_L^h]^2}{2^s} \right) \times \int_{\mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{I}/2^s) \end{aligned}$$

ただし、 $O_L^h = \text{Tr}_{\overline{h}}[O_L]$

1 量子ビットのユニタリ 2-デザイン



TPA のゲートブロックとして用いた構造



1 量子ビットの Tensor Product Ansatz のフレームポテンシャル。黒の破線は 1 量子ビットがユニタリ 2-デザインを成す場合のフレームポテンシャル ($= 1/3$) である。

極端な例として、ラベル 0 に属する入力データ $\mathcal{X} = \{x\}$ と ラベル 1 に属する入力データ $\mathcal{Z} = \{z\}$ が $\mathcal{X} = \mathcal{Z}$ である場合を考える。このとき

$$\begin{aligned}\mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} |\ell_i(\boldsymbol{\theta}) - 0| + \sum_{z \in \mathcal{Z}} |\ell_i(\boldsymbol{\theta}) - 1| \right) \\ &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} \ell_i(\boldsymbol{\theta}) + \sum_{x \in \mathcal{X}} (1 - \ell_i(\boldsymbol{\theta})) \right) = \frac{1}{2}\end{aligned}$$

ゆえに、コスト関数の勾配は 0 となり、分散も 0 となる。

→ 入力回路の構造が極めて単純であっても、異なるラベル間の入力データが近いほど、コスト関数の勾配の分散は小さくなると考えられる。

極端な例として、ラベル 0 に属する入力データ $\mathcal{X} = \{x\}$ と ラベル 1 に属する入力データ $\mathcal{Z} = \{z\}$ が $\mathcal{X} = \mathcal{Z}$ である場合を考える。このとき

$$\begin{aligned}\mathcal{L}_{\text{MSE}}(\theta) &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} (\ell_i - 0)^2 + \sum_{x \in \mathcal{X}} (\ell_i - 1)^2 \right) \\ &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} \ell_i^2 + \sum_{x \in \mathcal{X}} (\ell_i - 1)^2 \right) \\ &= \frac{1}{2} + \frac{1}{N} \sum_{x \in \mathcal{X}} 2 \left(\ell_i - \frac{1}{2} \right)^2\end{aligned}$$

このとき、 ℓ_i が $\frac{1}{2}$ のときにコスト関数 $\mathcal{L}_{\text{MSE}}(\theta)$ は最小となるため、教師ラベル $y_i = \{0, 1\}$ に近づくことはない。

極端な例として、ラベル 0 に属する入力データ $\mathcal{X} = \{x\}$ と ラベル 1 に属する入力データ $\mathcal{Z} = \{z\}$ が $\mathcal{X} = \mathcal{Z}$ である場合を考える。このとき

$$\begin{aligned}\mathcal{L}_{\text{LOG}}(\theta) &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} -\log(1 - \ell_i) + \sum_{x \in \mathcal{X}} -\log(\ell_i) \right) \\ &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} -\log \ell_i(1 - \ell_i) \right) \\ &= \frac{1}{N} \left(\sum_{x \in \mathcal{X}} -\log \left[-\left(\ell_i - \frac{1}{2} \right)^2 + \frac{1}{4} \right] \right)\end{aligned}$$

このとき、 ℓ_i が $\frac{1}{2}$ のときにコスト関数 $\mathcal{L}_{\text{LOG}}(\theta)$ は最小となるため、教師ラベル $y_i = \{0, 1\}$ に近づくことはない。