

修士論文

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

(Analysis of Data-encoding Induced Barren Plateau in Quantum Machine Learning)

東京大学大学院 理学系研究科 物理学専攻 寺師研究室

上曾山 健介

令和6年1月

概要

現在、機械学習はあらゆる分野で応用されている。しかし、機械学習の計算コストは膨大であり、その計算コストを削減することは重要な課題である。そこで、機械学習の計算を量子コンピューター上でを行い、計算を効率化することが期待されている。この研究分野を量子機械学習と呼ぶ。

現在開発されている量子コンピューターは、数百量子ビット程度で、ノイズが大きく、量子ビット同士の相互作用が近接のものだけである。そのような量子コンピューターを Noisy Intermediate-Scale Quantum Computer (NISQ) と呼ぶ。NISQ はノイズの影響を受けるため、使用できる量子ゲート数が制限される。そのような制約の中でも動作する量子アルゴリズムとして、変分量子アルゴリズムが注目されている。変分量子アルゴリズムは、古典コンピューターと組み合わせる最適化問題を解くアルゴリズムである。特に、変分量子アルゴリズムの枠組みで、量子機械学習の効率化も研究されている。しかし、変分量子アルゴリズムにおいては、コスト関数の勾配消失問題が最適化における大きなボトルネックとなっている。

本研究は、量子機械学習における識別モデルのコスト関数の勾配消失問題について、訓練データの入力が与える影響を解析した。その結果、訓練データ入力後の量子状態のエンタングルメントの大きさや訓練データの入力を行うための量子回路の表現能力の高さが最適化を困難にしていることが分かった。そして、その解析結果が数値計算と矛盾しないことを確認した。また、訓練データがガウス分布に従う場合について、その訓練データの分散がコスト関数の勾配のスケーリングにおいて重要な役割を果たすことを示した。

目次

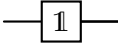
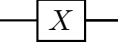
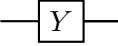
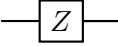
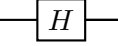

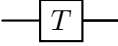
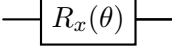
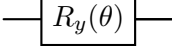
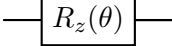
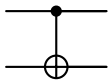

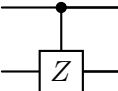
記号リスト	3
量子ゲート	4
第1章 序論	5
1.1 研究背景	5
1.2 本研究の目的と結果	6
1.3 本論文の構成	7
第2章 量子回路	8
2.1 量子ビット	8
2.2 純粋状態と混合状態	10
2.3 量子ゲート	11
2.3.1 1量子ビットゲート	11
2.3.2 制御ゲート	12
2.4 測定	15
2.5 エンタングルメント	15
2.6 量子状態の縮約	15
2.7 Schatten p -ノルム	16
2.8 Hilbert-Schmidt 距離	17
2.9 量子ノイズ	17
第3章 変分量子アルゴリズム	20
3.1 概要	20
3.1.1 コスト関数	21
3.1.2 変分量子回路	21
3.1.3 最適化	22
3.2 量子機械学習	24
3.2.1 量子回路学習	24
3.2.2 Data reuploading	25
3.2.3 量子カーネル法	25
3.3 バレンプラトナー	26
3.3.1 概要	26
3.3.2 原因	28

3.3.3	回避方法	30
第 4 章	ハール分布と量子回路の表現力	31
4.1	ハール分布	31
4.2	ユニタリ t -デザイン	33
4.3	表現力	35
4.3.1	$p = 1$	36
4.3.2	$p = 2$	36
第 5 章	量子機械学習におけるコスト関数の勾配の分散の上界	40
5.1	量子機械学習における勾配の分散の一般的な上界	41
5.2	グローバルユニタリ 2-デザインの入力状態とバレンプラトー	44
5.3	入力回路の表現力による上界	46
5.4	ローカルユニタリ 2-デザインの入力状態とバレンプラトー	48
5.5	ノイズの影響	51
第 6 章	量子機械学習におけるコスト関数の勾配の分散の下界	53
6.1	絶対誤差のコスト関数の勾配の分散	53
6.1.1	下界のスケーリング (1)	54
6.1.2	下界のスケーリング (2)	57
6.2	他のコスト関数の勾配の分散に関する考察	58
第 7 章	結論	62
	謝辞	63
付録 A	第 3 章の補足	70
A.1	機械学習の概要	70
A.1.1	教師あり学習のモデル	70
A.1.2	教師なし学習のモデル	71
A.1.3	強化学習のモデル	71
A.2	ユニタリ 2-デザインにおけるバレンプラトーの定理と証明	72
付録 B	第 5 章の補足	75
B.1	ローカルユニタリ 2-デザイン: ALT の計算例	75
付録 C	第 6 章の補足	78
C.1	定理 6.1.1 の略証	78
C.2	定理 6.1.2 の証明のための補題	78
C.3	定理 6.1.2 の証明	79
C.4	定理 6.1.4 の略証	82

記号リスト

\mathbb{N}	自然数
\mathbb{R}	実数
\mathbb{C}	複素数
δ_{ij}	Kronecker のデルタ
$\mathbf{x}, \boldsymbol{\theta}$	実数ベクトル
n	量子ビット数
L	量子回路の層数
$d, 2^n$	n 量子ビットの系の次元
$\mathbb{1}$	単位行列
U, V, W	ユニタリ
$\mathcal{U}(d)$	$d \times d$ 次元のユニタリ群
\mathbb{U}	$\mathcal{U}(d)$ の部分集合
\mathcal{H}	Hilbert 空間
ρ, σ	密度演算子
$\rho^{(\cdot)}$	縮約密度演算子
$\mathcal{S}(\mathcal{H})$	\mathcal{H} 上の密度演算子の集合
$\mathcal{L}(\mathcal{H})$	\mathcal{H} 上のベクトルに作用する線形演算子の集合
$\mathbb{1}, X, Y, Z; P_0, P_1, P_2, P_3$	Pauli 行列
\mathcal{P}	Pauli 群
\mathcal{C}	Clifford 群
\mathcal{E}	ユニタリチャネル
\mathcal{N}	ノイズチャネル
\mathcal{L}	量子機械学習におけるコスト関数
$\kappa(\cdot, \cdot)$	カーネル
\mathcal{O}	オブザーバブル
\mathcal{O}	Big-O 記法
i, j, k, l, p, q	添字
$\mathbf{i}, \mathbf{j}, \mathbf{p}, \mathbf{q}$	ベクトル添字
\mathcal{F}	フレームポテンシャル
$\epsilon_{\mathbb{U}}$	量子回路の表現力
μ_{Haar}	Haar 測度
$\ \cdot\ _p$	Schatten p -ノルム
$D_{\text{HS}}(\cdot, \cdot)$	Hilbert-Schmidt 距離

量子ゲート

Name	Circuit	Matrix
Identity		$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Pauli-X		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Pauli-Y		$\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$
Pauli-Z		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Hadamard		$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$
Phase		$\begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$
T		$\begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix}$
x -rotation		$\begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$
y -rotation		$\begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$
z -rotation		$\begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}$
CX, CNOT		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$
SWAP		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
CZ		$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

第1章 序論

1.1 研究背景

量子コンピューターの研究は、1980年代に Paul Anthony Benioff、David Deutsch、Richard Feynman などらが量子力学に基づいたコンピューターによる計算を提案したところから始まった [1-3]。量子コンピューターは、量子力学の性質を帯びる系（2準位系）を情報の最小単位として用いるコンピューターである。その最小単位を量子ビットと呼ぶ。量子ビットは、重ね合わせやエンタングルメントといった量子力学的性質を持つ。これらの性質を利用することで、量子コンピューターは従来のコンピューターよりもある種の計算を効率的に行うことができるのではないかと期待されている。以降では、従来のコンピューターのことを古典コンピューターと呼ぶ。

1990年代に入るとさまざまな量子アルゴリズムが提案されるようになり、とりわけ、Peter Shor が、素因数分解問題を多項式時間で解く量子アルゴリズム [4]（Shor のアルゴリズム）を発表したことで、量子コンピューターの研究が注目されるようになった。というのも、古典コンピューターでは素因数分解問題を解くための計算量は指数関数的に増大するが、その計算量的な難しさこそが、RSA 暗号の安全性を保証しているからである。

それ以来、より一層研究が進められ、量子計算理論の構築・量子アルゴリズム開発・量子コンピューターのソフトウェア/ハードウェア開発など、多方面の知識と技術が向上していった。そして、2010年代には D-wave の量子アニーリングマシン、Google と IBM の超伝導量子コンピューターが登場した。最近では、Google や IBM が量子超越に関する論文 [5, 6] を発表したことで、より量子コンピューターの研究が注目されている。

Shor のアルゴリズム以外にも、非構造的なデータベースから欲しいデータを見つける量子アルゴリズム [7]（1995年：Grover のアルゴリズム）や、線型方程式を解く量子アルゴリズム [8]（2009年：HHL アルゴリズム）などの効率的な量子アルゴリズムが提案された。これらのアルゴリズムは、量子並列性・エンタングルメントといった量子力学的性質をうまく利用することで、古典計算より効率的であると期待されている。

これらのアルゴリズムを実装するには、大規模で誤り訂正可能な量子コンピューターが必要である。しかしながら、現在開発されている量子コンピューターは、中規模（数百量子ビット）であり、ノイズが大きく、量子ビット同士の相互作用は近接のものだけである。そのような量子コンピューターを NISQ (Noisy Intermediate-Scale Quantum Computer) と呼ぶ [9]。NISQ は誤り訂正機能を持たず、ノイズの影響を受けやすいため、使用する量子ゲートの数をなるべく少なくする必要がある。

そのような量子ビット数と量子ゲート数の制約の中でも動作する量子アルゴリズムとして、変分量子アルゴリズム (Variational Quantum Algorithm, VQA) [10] が注目されている。なぜなら、このアルゴリズムはノイズに強く、少ない量子ゲートで実装できるという特徴があるからである。変分量子アルゴリズムは、古典コンピューターと組み合わせて最適化問題を解くアルゴリズムである。最適化問題とは、

ある関数の値を最小化（あるいは最大化）するようなパラメーターの値を求める問題である。変分量子アルゴリズムにおいては、パラメーター化された量子状態を試行関数として用いる。そして、その試行関数のパラメーターを古典コンピューター上で更新することで、コスト関数と呼ばれる関数の値を最小化するように学習する。初めは、量子化学計算における変分量子固有値ソルバー（VQE） [11] や、組み合わせ最適化問題における変分量子近似最適化（QAOA） [12] のアルゴリズムとして提案された。しかし、変分量子アルゴリズムは、問題をコスト関数に書き換えることができれば、どのような問題にも適用できる。そのため、応用範囲は極めて広く、量子化学計算 [11, 13, 14]、量子シミュレーション [15, 16]、量子コンパイル [17, 18]、機械学習 [19–22] といった様々な分野で研究されている。このように、変分量子アルゴリズムは NISQ 時代における有望なアルゴリズムとして注目されている。

そんな中、変分量子アルゴリズムにおいて、バレンプラトーと呼ばれるコスト関数の勾配消失が起こりうることが判明した [23]。バレンプラトーとは、コスト関数の勾配の平均が 0 で、勾配の分散が量子ビット数に関して指数関数的に小さくなる現象のことである。よって、バレンプラトーが起きると、コスト関数の勾配を正確に推定するために指数関数的に多くの測定が必要となり、効率的な学習ができなくなる [24]。バレンプラトーを引き起こす原因は 1 つではなく、量子変分回路の構造 [23, 25, 26] や、測定するオブザーバブル [27]、ノイズ [28]、訓練データの入力 [29, 30] など複数の原因が先行研究で指摘されている。

1.2 本研究の目的と結果

近年、機械学習はさまざまな分野で著しい成果を上げているが、その一方で、高い計算コストが問題となっている。例えば、2022 年に公開された ChatGPT-3 は最大 1750 億個のパラメーターを持つ大規模言語モデルである。そのような巨大なモデルを学習するためには、多くの計算機と数週間から数ヶ月に及ぶ学習が必要とされる。このように、機械学習の計算コストは膨大であり、その計算コストを削減することは重要な課題である。そこで、機械学習の計算を量子コンピューター上で行い、計算を効率化することが期待されている。この研究分野を量子機械学習 [31] と呼ぶ。

量子計算の機械学習への応用は、変分量子アルゴリズムの登場以前からあり、HHL アルゴリズムを応用したアルゴリズムが提案されていた。しかし、NISQ 上での実装が難しいため、最近では変分量子アルゴリズムを応用した機械学習アルゴリズムが次々と提案されるようになった。量子ニューラルネットワーク（QNN） [19, 20]、量子畳み込みニューラルネットワーク（QCNN） [32]、量子カーネル法 [21, 22] などがその例である。

本研究は、変分量子アルゴリズムの文脈における教師あり量子機械学習の効率化に焦点を当てた。量子機械学習アルゴリズムの効率化においては、訓練データとその入力方法（以降、二つを合わせて「データ入力」と呼ぶ）が重要であることが指摘されており [33, 34]、先ほど述べたバレンプラトーとの関連も見出された [29, 30]。しかしながら、多くのバレンプラトー研究は主に学習回路の構造に着目してきたため、データ入力がバレンプラトーに与える影響は十分には理解されていない。そこで、特に、量子機械学習に特有のデータ入力がバレンプラトーにどのような影響を与えるのか評価した。バレンプラトーは複数の原因によって生じるため、データ入力によるバレンプラトーの評価においては、他の原因を排除した量子回路の設定の下で解析を行った。その結果、データ入力後の量子状態のエンタングルメントが多い、あるいは、訓練データを入力するための量子回路（以降、入力回路と呼ぶ）の表現能力が高い

ほど、コスト関数の勾配の分散の上界が小さくなり、バレンプラトーにつながるものが明らかになった。そして、具体的な入力回路を用いたシミュレーションの結果と解析計算の結果を比較することで、解析計算の妥当性を確認した。また、量子回路に入力する訓練データにガウス分布を仮定し、絶対誤差のコスト関数の勾配の分散の下界についても解析を行った。その結果、絶対誤差のコスト関数の勾配の分散の下界において、訓練データの分散が大きな役割を果たすことが分かった。

1.3 本論文の構成

本論文の構成は以下の通りである。

- 第 2 章では、量子回路の要素と量子計算の基礎について述べる。
- 第 3 章では、変分量子アルゴリズムとバレンプラトーについて述べる。
- 第 4 章では、バレンプラトーの解析に必要となるハール分布とユニタリ t -デザインについて述べる。
- 第 5 章では、量子機械学習におけるバレンプラトーの解析として、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する。
- 第 6 章では、量子機械学習におけるバレンプラトーの解析として、コスト関数の勾配の分散の下界をデータ入力の観点から導出する。
- 付録には、本論文で用いる定理・補題の証明、その他の補足を記す。

第2章 量子回路

ここでは、量子計算を記述するための量子回路モデルについて説明する。量子回路モデルとは、量子計算を記述するための最も一般的なモデルである。量子回路は次の3つの要素からなる。量子ビットと呼ばれる状態を保持する量子系、量子ゲートと呼ばれる量子系の時間発展を表す操作、そして測定と呼ばれる量子系の状態から情報を読み出す操作である。この章では、これらの要素と後の章で必要となる指標について説明する。

2.1 量子ビット

1 量子ビット

古典コンピューターにおいて、情報を表現する最小単位はビットである。ビットは2つの状態を持ち、これらを0と1で表す。物理的には、電圧の高低や、磁化の向きなどの2つの状態によって実現される。量子コンピューターにおいては、量子ビットがこれに対応する。量子ビットは2準位系と呼ばれる2つのエネルギー準位を持つ量子系によって実現される。その2つのエネルギー準位にある状態を $|0\rangle$, $|1\rangle$ と表す。量子系の状態は複素 Hilbert 空間の単位ベクトルとして記述されるが、特に、 $|0\rangle$, $|1\rangle$ は1量子ビットの状態空間 \mathbb{C}^2 において正規直交基底を成し、計算基底と呼ばれる。よって、

$$|0\rangle := \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |1\rangle := \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2.1.1)$$

とする。1量子ビットの任意の（純粋¹）状態 $|\psi\rangle$ は、次のように計算基底の重ね合わせによって表すことができる。

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad (2.1.2)$$

ただし、 α, β は複素数で、 $|\alpha|^2 + |\beta|^2 = 1$ を満たす。

1量子ビットの状態は、より幾何的に、図 2.1.1 にあるような単位球面上の1点として表現することができる²。この単位球は Bloch 球と呼ばれる。Bloch 球面上の点は、

$$|\psi\rangle = e^{i\delta} \left[\cos\left(\frac{\theta}{2}\right) |0\rangle + e^{i\varphi} \sin\left(\frac{\theta}{2}\right) |1\rangle \right] \quad (2.1.3)$$

と表すことができる。 δ, θ, φ は実数で、 $e^{i\delta}$ を絶対位相、 $e^{i\varphi}$ を相対位相と呼ぶ。絶対位相は測定によっては観測できず、絶対位相のみ異なる2つの状態を区別することはできないため、一般には無視することができる。

¹2.2にて説明する。

²計算基底の内積は0であるが、Bloch 球面上では平行になっていることに注意。

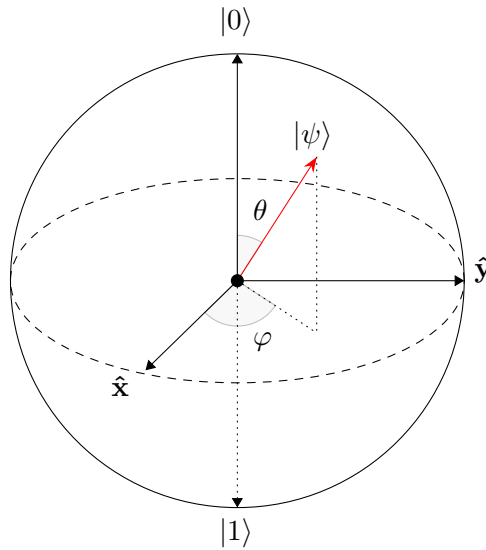
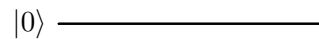


図 2.1.1: 1 量子ビットの状態に対応する Bloch 球面上の点

量子回路において、1 量子ビットは 1 本の配線のように描かれる。



左端の $|0\rangle$ は初期状態を表す。この配線上に、量子ゲートや測定などの操作を表す記号を並べることで量子回路を表現する。この配線は物理的実態があるわけではなく、量子ビットの状態にどのような操作を加えるかを、時系列に沿って左から右に並べて表現するためのものである。

多量子ビット

n 量子ビットの状態空間は、1 量子ビットの状態空間のテンソル積空間 $(\mathbb{C}^2)^{\otimes n}$ で表される。正規直交基底は、各量子ビットの計算基底のテンソル積のすべての組み合わせによって与えられる。つまり、

$$\{|i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle \mid i_1, i_2, \dots, i_n \in \{0, 1\}\} \quad (2.1.4)$$

である。簡単のため、 $|i_1 i_2 \cdots i_n\rangle := |i_1\rangle \otimes |i_2\rangle \otimes \cdots \otimes |i_n\rangle$ と書くこともある。左から第 k 番目の量子ビットを、第 k 量子ビットと呼ぶ³。

テンソル積は、次のように計算を行う。

$$|\psi\rangle \otimes |\varphi\rangle = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \otimes \begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} \alpha \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \\ \beta \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \alpha\gamma \\ \alpha\delta \\ \beta\gamma \\ \beta\delta \end{bmatrix} \quad (2.1.5)$$

³書籍によって右から数えることもある。

例えば、2 量子ビットの計算基底は次で与えられる。

$$|00\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad |01\rangle = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad |10\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad |11\rangle = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (2.1.6)$$

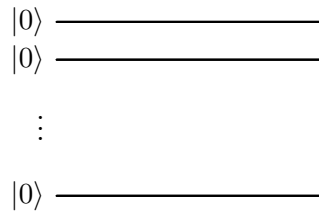
3 量子ビット以上の場合も同様に計算する。このことから、 n 量子ビットの状態は、 2^n 次元の複素ベクトルで表現されることがわかる。

n 量子ビットの任意の（純粋）状態 $|\psi\rangle$ は、次のように計算基底の重ね合わせによって表すことができる。

$$|\psi\rangle = \sum_{i_1, i_2, \dots, i_n \in \{0, 1\}} c_{i_1 i_2 \dots i_n} |i_1 i_2 \dots i_n\rangle \quad (2.1.7)$$

ただし、 $c_{i_1 i_2 \dots i_n}$ は複素数で、 $\sum_{i_1, i_2, \dots, i_n \in \{0, 1\}} |c_{i_1 i_2 \dots i_n}|^2 = 1$ を満たす。

量子回路において、 n 量子ビットは、以下のように n 本の配線を縦に並べることで表現される。



上から第 k 番目の量子ビットを、第 k 量子ビットと呼び、テンソル積の表現における第 k 量子ビットと対応づける。

2.2 純粋状態と混合状態

計算基底の重ね合わせによって表現できる状態を純粋状態と呼ぶ。しかし、実際の量子コンピューター上の量子状態は外部ノイズの影響を受けて純粋状態にいつまでも留まることはできず、様々な状態が混ざった状態になる（重ね合わせとは異なる）。例えば、古典的な確率 p_i で $|\psi_i\rangle$ にあるような状態である（ただし、 $\sum_i p_i = 1$ ）。このような状態を混合状態と呼ぶ。混合状態は、次のように定義される密度演算子 ρ として表される。

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \quad (2.2.1)$$

一方、純粋状態は $\rho = |\psi\rangle\langle\psi|$ と表すことができる。純粋状態であっても、混合状態であっても、常に $\text{Tr}[\rho] = 1$ が成り立つ。

純粋状態と混合状態の違いの指標として、純粋度がある。純粋度は、量子状態がどれくらい純粋状態に近いかを表す尺度であり、 $\text{Tr}[\rho^2]$ で定義される。純粋状態の場合、 $\text{Tr}[\rho^2] = \text{Tr}[\rho] = 1$ である。一方、混合状態の場合、 $\text{Tr}[\rho^2] < \text{Tr}[\rho] = 1$ である。混合状態の中でも、 $\rho = \mathbb{1}/d$ と表されるものを最大混合

状態と呼ぶ。ただし、量子ビット数を n とし、 $d = 2^n$ と定めた。その純粋度は、 $\text{Tr}[\rho^2] = 1/d$ である。一般に、純粋度は、 $1/d \leq \text{Tr}[\rho^2] \leq 1$ の範囲をとる。

状態 (2.1.3) の密度演算子は単位ベクトル $\mathbf{r} = (r_x, r_y, r_z) = (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta)$ とパウリ行列 $\boldsymbol{\sigma} := (X, Y, Z)$ を用いて次のように表すことができる。

$$\begin{aligned}
 |\varphi\rangle\langle\varphi| &= \begin{bmatrix} \cos^2(\frac{\theta}{2}) & e^{-i\varphi} \sin(\frac{\theta}{2}) \cos(\frac{\theta}{2}) \\ e^{i\varphi} \sin(\frac{\theta}{2}) \cos(\frac{\theta}{2}) & \sin^2(\frac{\theta}{2}) \end{bmatrix} \\
 &= \frac{1}{2} \begin{bmatrix} 1 + \cos \theta & e^{-i\varphi} \sin \theta \\ e^{i\varphi} \sin \theta & 1 - \cos \theta \end{bmatrix} \\
 &= \frac{1}{2} (\mathbb{1} + (\cos \varphi \sin \theta)X + (\sin \varphi \sin \theta)Y + (\cos \theta)Z) \\
 &= \frac{1}{2} (\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\sigma}) \tag{2.2.2}
 \end{aligned}$$

2.3 量子ゲート

2.3.1 1量子ビットゲート

量子ゲートとは量子状態に作用させるユニタリ演算子のことである。量子回路において、初期状態 $|\psi\rangle$ にユニタリ演算子 U を作用させ、 $U|\psi\rangle$ という状態に変換する過程を以下のように表す。

$$|\psi\rangle \text{ --- } \boxed{U} \text{ --- } U|\psi\rangle$$

以下に代表的な 1 量子ビットゲートを示す。計算基底によってユニタリ演算子を行列表示する。

パウリゲート

$$\mathbb{1} = |0\rangle\langle 0| + |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.3.1}$$

$$X = |1\rangle\langle 0| + |0\rangle\langle 1| = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2.3.2}$$

$$Y = i|1\rangle\langle 0| - i|0\rangle\langle 1| = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{2.3.3}$$

$$Z = |0\rangle\langle 0| - |1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{2.3.4}$$

Hadamard ゲート H 、位相ゲート S 、 T ゲート T

$$H = \frac{1}{\sqrt{2}}(|+\rangle\langle 0| + |-\rangle\langle 1|) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (2.3.5)$$

$$S = |0\rangle\langle 0| + i|1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix} \quad (2.3.6)$$

$$T = |0\rangle\langle 0| + e^{i\pi/4}|1\rangle\langle 1| = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{bmatrix} \quad (2.3.7)$$

ただし、 $|\pm\rangle := \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ である。

回転ゲート R_x, R_y, R_z

\mathbf{n} を 3次元の単位ベクトル、 $\boldsymbol{\sigma} = (X, Y, Z)$ をパウリ行列のベクトルとする。任意の正方行列 A に対して、 $\exp(iA\theta) = \cos(\theta)\mathbb{1} + i\sin(\theta)A$ であることを用いると、 $R_n := \exp(-i\theta \mathbf{n} \cdot \boldsymbol{\sigma}/2)$ を三角関数によって次のように表すことができる。

$$R_n := \exp(-i\theta \mathbf{n} \cdot \boldsymbol{\sigma}/2) = \cos\left(\frac{\theta}{2}\right)\mathbb{1} - i\sin\left(\frac{\theta}{2}\right)(\mathbf{n} \cdot \boldsymbol{\sigma}) \quad (2.3.8)$$

特に、 $\mathbf{n} = (1, 0, 0), (0, 1, 0), (0, 0, 1)$ のとき、 R_n は Bloch 球面上で x, y, z 軸周りの回転を表す。よって、これらの軸周りの回転を表すゲートを R_x, R_y, R_z と定義する。

$$R_x(\theta) := e^{-i\theta X/2} = \cos\left(\frac{\theta}{2}\right)\mathbb{1} - i\sin\left(\frac{\theta}{2}\right)X = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \quad (2.3.9)$$

$$R_y(\theta) := e^{-i\theta Y/2} = \cos\left(\frac{\theta}{2}\right)\mathbb{1} - i\sin\left(\frac{\theta}{2}\right)Y = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} \quad (2.3.10)$$

$$R_z(\theta) := e^{-i\theta Z/2} = \cos\left(\frac{\theta}{2}\right)\mathbb{1} - i\sin\left(\frac{\theta}{2}\right)Z = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix} \quad (2.3.11)$$

ブロッホ球面上の点は、直交する 2つの軸周りの回転の合成で任意の点に移すことができる。よって、任意の 1量子ビットゲートは、 R_x, R_y, R_z のいずれか 2つの軸周りの回転の合成で表すことができる。

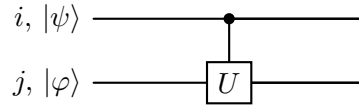
2.3.2 制御ゲート

制御ゲートとは、制御ビットが $|1\rangle$ のときのみ、標的ビットにユニタリ演算子 U を作用させるゲートである。制御ビットが $|0\rangle$ のときは、標的ビットに何も作用させない。この制御ゲートを用いることで、複数量子ビット間にエンタングルメント (2.5 節) という量子力学的な相関を作り出すことができる。一般に、第 i 量子ビットを制御ビット、第 j 量子ビットを標的ビットとする制御ゲートを $C_j^i[U]$ と表す。

行列表示すると、

$$C_j^i[U] = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{11} & U_{12} \\ 0 & 0 & U_{21} & U_{22} \end{bmatrix} \quad (2.3.12)$$

量子回路において、第 i 量子ビットを制御ビット、第 j 量子ビットを標的ビットとする制御ゲートを次のように表す。



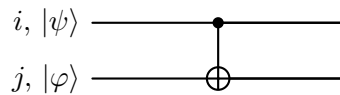
制御ビットと標的ビットは、複数の場合もある。例えば、 n 個の量子ビットのうち、 i_1, i_2, \dots, i_k 番目の量子ビットを制御ビット、 j_1, j_2, \dots, j_l 番目の量子ビットを標的ビットとする制御ゲートを $C_{j_1 j_2 \dots j_l}^{i_1 i_2 \dots i_k}[U]$ と表す。

CNOT ゲート

$U = X$ のときの制御ゲートを CNOT ゲートと呼ぶ。制御ビットが $|1\rangle$ のときのみ、標的ビットに X ゲートを作用させるゲートである。制御ビットと標的ビットを明示する場合は、 $C_j^i[X]$ と表す。行列表示すると、

$$C_j^i[X] = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.3.13)$$

量子回路において、第 i 量子ビットを制御ビット、第 j 量子ビットを標的ビットとする CNOT ゲートを次のように表す。

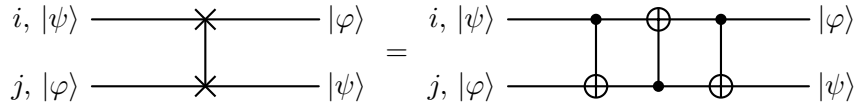


SWAP ゲート

2つの量子ビットの状態を入れ替えるゲートを SWAP ゲートと呼ぶ。つまり、 $\text{SWAP} |\psi\rangle \otimes |\varphi\rangle = |\varphi\rangle \otimes |\psi\rangle$ と作用する。SWAP ゲートは、CNOT ゲートを3つ組み合わせて作ることができる。行列表示すると、

$$\text{SWAP} = C_j^i[X] C_i^j[X] C_j^i[X] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.3.14)$$

量子回路において、第 i 量子ビットと第 j 量子ビットの状態を入れ替える SWAP ゲートを次のように表す。



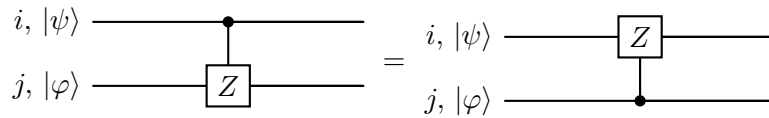
CZ ゲート

$U = Z$ のときの制御ゲートを CZ ゲートと呼ぶ。制御ビットが $|1\rangle$ のときのみ、標的ビットに Z ゲートを作用させるゲートである。行列表示すると、

$$CZ = |0\rangle\langle 0| \otimes \mathbf{1} + |1\rangle\langle 1| \otimes Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (2.3.15)$$

CZ ゲートは、計算基底 $|11\rangle$ のみに -1 をかけるゲートである。よって、制御ビットと標的ビットを入れ替えても CZ ゲートの作用は変わらないため、 $C_j^i[Z] = C_i^j[Z]$ である。

量子回路において、第 i 量子ビットを制御ビット、第 j 量子ビットを標的ビットとする CZ ゲートを次のように表す。



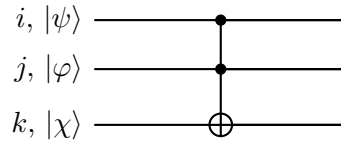
Toffoli ゲート

Toffoli ゲートは、2つの制御ビットと1つの標的ビットを持つ3量子ビットゲートである。2つの制御ビットがともに $|1\rangle$ のときのみ、標的ビットに X ゲートを作用させるゲートである。CCX ゲートとも呼ばれる。制御ビットと標的ビットを明示する場合は、 $C_k^{i,j}[X]$ と表す。行列表示すると、

$$C_k^{i,j}[X] = |00\rangle\langle 00| \otimes \mathbf{1} + |01\rangle\langle 01| \otimes \mathbf{1} + |10\rangle\langle 10| \otimes \mathbf{1} + |11\rangle\langle 11| \otimes X$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.3.16)$$

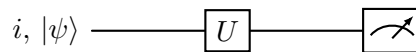
量子回路において、第 i 量子ビットと第 j 量子ビットを制御ビット、第 k 量子ビットを標的ビットとする Toffoli ゲートを次のように表す。



2.4 測定

量子コンピューターにおいて、測定は計算基底によって行われる。1 量子ビットのある状態を測定して得られる結果は、計算基底のどちらかである。それぞれの計算基底が得られる確率は、その状態の計算基底の振幅の絶対値の 2 乗である。例えば、 $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ である状態を測定すると、0 が出る確率は $|\alpha|^2$ 、1 が出る確率は $|\beta|^2$ である。この確率分布を得るためには、多数回の測定を行う必要がある。また、測定後の状態は、測定された計算基底に対応する状態となる。例えば、 $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ である状態を測定して 0 が出たとすると、測定後の状態は $|0\rangle$ になる。 n 量子ビットの場合も同様である。

量子回路において、第 i 量子ビットを測定するときは、以下のように表す。



2.5 エンタングルメント

エンタングルメントとは、複数の量子ビットの状態が純粋状態であるにもかかわらず、それぞれの量子ビットの状態に分離して記述することができないような状態の性質である。例えば、次の 2 量子ビットの状態はエンタングルメントを持つ。

$$|\psi_{\text{ent}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle) \quad (2.5.1)$$

$|\psi\rangle$ は、 $|0\rangle \otimes |0\rangle$ と $|1\rangle \otimes |1\rangle$ の重ね合わせであるが、それぞれの量子ビットの状態に分離した形、すなわち、 $|\phi\rangle \otimes |\varphi\rangle$ の形で表すことができない。

量子計算は行列計算によって記述されるので、原理的には古典コンピューターでシミュレートすることができる。エンタングルメントのない分離可能な量子状態は、各量子ビットの状態を個別に記述できるので、古典コンピューターによる計算量は量子ビット数に関して線形に増える。しかし、エンタングルメントを持つ量子状態の場合、各量子ビットの状態を個別に記述できないので、古典コンピューターによる計算量は量子ビット数に関して指数関数的に増える。このように、エンタングルメントの存在が量子計算と古典計算の計算量の差を生み出す。

2.6 量子状態の縮約

量子状態の縮約とは、全体の系の量子状態からある部分系の量子状態を求める操作のことである。例えば、2 量子ビットの状態 ρ において第 1 量子ビットの状態 ρ_1 を求めるには、次のように ρ の第 2 量

子ビットの部分トレースを取ることで求めることができる。

$$\rho_1 = \text{Tr}_2[\rho] \quad (2.6.1)$$

例として、2 量子ビットの系を考える。全体の系の状態 ρ が各量子ビットの状態のテンソル積で表される、つまり、 $\rho = (|\psi_1\rangle \otimes |\psi_2\rangle)(\langle\psi_1| \otimes \langle\psi_2|) = |\psi_1\rangle\langle\psi_1| \otimes |\psi_2\rangle\langle\psi_2|$ とする。このとき、第 1 量子ビットの状態は、

$$\rho_1 = \text{Tr}_2[|\psi_1\rangle\langle\psi_1| \otimes |\psi_2\rangle\langle\psi_2|] = |\psi_1\rangle\langle\psi_1| \text{Tr}[|\psi_2\rangle\langle\psi_2|] = |\psi_1\rangle\langle\psi_1|$$

である。これは純粋状態である。

一方、全体の系の状態 ρ がエンタングルメントを持ち、次の状態で表されるとする。

$$\begin{aligned} \rho &= |\psi_{\text{ent}}\rangle\langle\psi_{\text{ent}}| \\ &= \frac{1}{2}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle)(\langle 0| \otimes \langle 0| + \langle 1| \otimes \langle 1|) \\ &= \frac{1}{2}(|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| + |1\rangle\langle 1| \otimes |1\rangle\langle 1|) \end{aligned}$$

このとき、第 1 量子ビットの状態は、

$$\begin{aligned} \rho_1 &= \text{Tr}_2[|\psi_{\text{ent}}\rangle\langle\psi_{\text{ent}}|] \\ &= \frac{1}{2}(|0\rangle\langle 0| \text{Tr}[|0\rangle\langle 0|] + |0\rangle\langle 1| \text{Tr}[|0\rangle\langle 1|] + |1\rangle\langle 0| \text{Tr}[|1\rangle\langle 0|] + |1\rangle\langle 1| \text{Tr}[|1\rangle\langle 1|]) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \\ &= \frac{\mathbb{1}}{2} \end{aligned}$$

である。これは最大混合状態である。このように、エンタングルメントを持つ量子状態の部分系の状態は、混合状態になることがある。

2.7 Schatten p -ノルム

定義 2.7.1. $A \in \mathcal{L}(\mathcal{H})$ の Schatten p -ノルムは次のように定義される。

$$\|A\|_p = \begin{cases} (\text{Tr}[|A|^p])^{\frac{1}{p}} & (1 \leq p < \infty) \\ \lambda_{\max}(|A|) & (p = \infty) \end{cases} \quad (2.7.1)$$

ここで $|A| = \sqrt{A^\dagger A}$ であり、 $\lambda_{\max}(|A|)$ は $|A|$ の最大固有値である。

$A, B \in \mathcal{L}(\mathcal{H})$ とする。Schatten p -ノルムは以下の性質を持つ。

1. $1 \leq \forall p \leq \forall q \leq \infty$ に対して $\|A\|_q \leq \|A\|_p$
2. $\forall p \in [1, \infty]$, $\forall U, V \in \mathcal{U}(d)$ に対して $\|UAV^\dagger\|_p = \|A\|_p$
3. $\|A\|_p = \|A^\top\|_p = \|A^*\|_p = \|A^\dagger\|_p$

4. $1 \leq p, q, r$ and $1/p + 1/q \leq 1/r$ に対して $\|AB\|_r \leq \|A\|_p \|B\|_q$ (Hölder の不等式)

- 例 1 $\|AB\|_1 \leq \|A\|_1 \|B\|_\infty$.
- 例 2 $\|AB\|_1 \leq \|A\|_2 \|B\|_2$.

2.8 Hilbert–Schmidt 距離

$A, B \in \mathcal{L}(\mathcal{H})$ の Hilbert–Schmidt 距離を次のように定義する。

$$D_{\text{HS}}(A, B) := \|A - B\|_{\text{HS}}^2 = \text{Tr}[(A - B)^\dagger(A - B)] \quad (2.8.1)$$

X がエルミート演算子のとき、 $X^\dagger = X$ であるから、密度演算子 ρ, σ の Hilbert–Schmidt 距離は

$$D_{\text{HS}}(\rho, \sigma) = \|\rho - \sigma\|_{\text{HS}}^2 = \text{Tr}[(\rho - \sigma)^2] \in [0, 2] \quad (2.8.2)$$

となる。例えば、 $\rho = \sigma$ のとき $D_{\text{HS}}(\rho, \sigma) = 0$ であり、 $\rho = |0\rangle\langle 0|^{\otimes n}$, $\sigma = |1\rangle\langle 1|^{\otimes n}$ のとき $D_{\text{HS}}(\rho, \sigma) = 2$ である。また、 ρ と最大混合状態 $\mathbb{1}/d$ ($d = 2^n$) の Hilbert–Schmidt 距離は、

$$\begin{aligned} D_{\text{HS}}(\rho, \mathbb{1}/d) &= \text{Tr}[(\rho - \mathbb{1}/d)^2] \\ &= \text{Tr}[\rho^2 - 2\rho \cdot \mathbb{1}/d + \mathbb{1}/d^2] \\ &= \text{Tr}[\rho^2] - \frac{1}{d} \in [0, 1 - 1/d] \end{aligned} \quad (2.8.3)$$

となる。特に、 $\rho = |\psi\rangle\langle\psi|$ のとき $D_{\text{HS}}(|\psi\rangle\langle\psi|, \mathbb{1}/d) = 1 - 1/d$ である。

2.9 量子ノイズ

量子ノイズには、コヒーレントなノイズとデコヒーレントなノイズの 2 種類がある。コヒーレントなノイズは、量子状態の純粋度を変化させないが、位相を変化させるノイズである。デコヒーレントなノイズは、量子状態の純粋度を変化させるノイズである。ここでは、後者の例を紹介する。

ビット反転チャンネル

ビット反転チャンネルとは、確率 p で量子ビットに X ゲートを作用させるチャンネルであり、1 量子ビットに対して次のように定義される。

$$\mathcal{N}_p(\rho) := (1 - p)\rho + pX\rho X \quad (2.9.1)$$

式 (2.2.2) にて確認したように、 $\rho = \frac{1}{2}(\mathbb{1} + r_x X + r_y Y + r_z Z)$ と表せることから、 $X\rho X = \frac{1}{2}(\mathbb{1} + r_x X - r_y Y - r_z Z)$ となる。これは、ブロッホ球上の点 (r_x, r_y, r_z) を x 軸まわりに π 回転させることに相当する。これと同様に、 y 軸まわりに π 回転させるチャンネルと z 軸まわりに π 回転させるチャンネルを定義することができ、それぞれ、ビット位相反転チャンネル、位相反転チャンネルと呼ばれる。

Depolarizing チャンネル

Depolarizing チャンネルは、ビット反転チャンネル、位相反転チャンネル、ビット位相反転チャンネルを含むチャンネルであり、1 量子ビットに対して次のように定義される。

$$\mathcal{N}_p(\rho) := \left(1 - \frac{3}{4}p\right)\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z) \quad (2.9.2)$$

ここで、次の等式が成り立つ。

$$\begin{aligned} X\rho X &= \frac{1}{2}(\mathbb{1} + r_x X - r_y Y - r_z Z) \\ Y\rho Y &= \frac{1}{2}(\mathbb{1} - r_x X + r_y Y - r_z Z) \\ Z\rho Z &= \frac{1}{2}(\mathbb{1} - r_x X - r_y Y + r_z Z) \end{aligned} \quad (2.9.3)$$

これより、

$$\frac{\mathbb{1}}{2} = \frac{\rho + X\rho X + Y\rho Y + Z\rho Z}{4} \quad (2.9.4)$$

であるから、Depolarizing チャンネルは次のように書き換えることができる。

$$\mathcal{N}_p(\rho) = (1 - p)\rho + p \frac{\mathbb{1}}{2} \quad (2.9.5)$$

例えば、 $p = \frac{1}{2}$ のとき、Depolarizing チャンネルは図 2.9.1 のように、ブロッホ球面上の点 (r_x, r_y, r_z) を収縮するように変化させるチャンネルである。

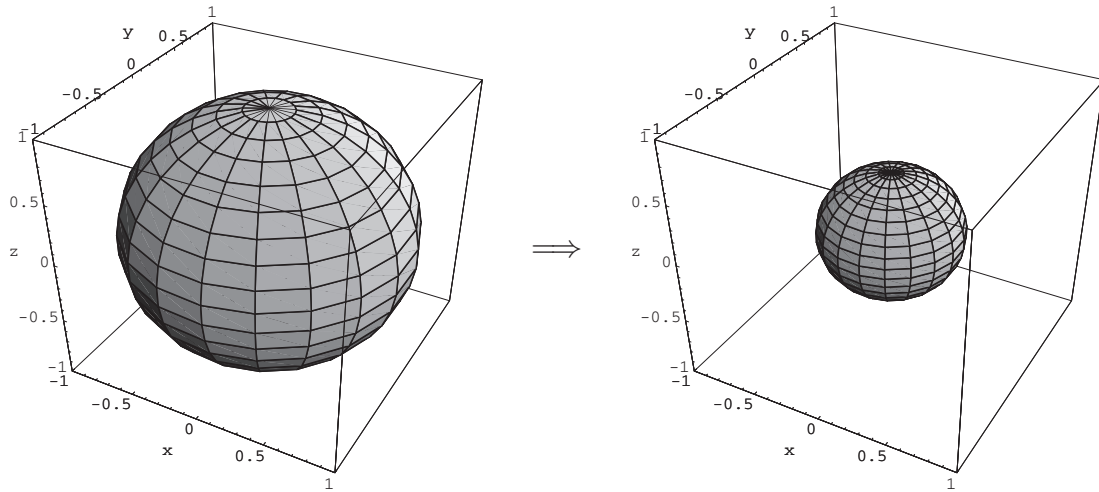


図 2.9.1: Depolarizing channel [35]

n 量子ビットに対する Depolarizing チャンネルは、次のように定義される。

$$\mathcal{N}_p(\rho) = (1 - p)\rho + p \frac{\mathbb{1}}{2^n} \quad (2.9.6)$$

Depolarizing チャンネルによる量子ノイズモデル

量子コンピューター上で、 n 量子ビットの初期状態 ρ に対して、 L 個のユニタリ演算子を作用させることを考える。このとき、各ユニタリ演算子に付随して、Depolarizing チャンネル \mathcal{N}_{p_i} が作用すると仮定する。作用させるユニタリチャンネルを $\mathcal{E}_i(\rho) = U_i \rho U_i^\dagger$, $i = 1, 2, \dots, L$ とすると、全体のチャンネル \mathcal{N} は次のように表される。

$$\mathcal{N}(\rho) = \left(\bigcirc_{i=1}^L \mathcal{N}_{p_i} \circ \mathcal{E}_i \right) (\rho) \quad (2.9.7)$$

ここで、次のように $\mathcal{N}_{p_i} \circ \mathcal{E}_i = \mathcal{E}_i \circ \mathcal{N}_{p_i}$ が成り立つ。

$$\mathcal{N}_{p_i} \circ \mathcal{E}_i(\rho) = (1 - p_i)(U_i \rho U_i^\dagger) + p_i \frac{\mathbb{1}}{2^n} = U_i \left((1 - p_i)\rho + p_i \frac{\mathbb{1}}{2^n} \right) U_i^\dagger = \mathcal{E}_i \circ \mathcal{N}_{p_i}(\rho) \quad (2.9.8)$$

よって、 $\mathcal{E} := \bigcirc_{i=1}^L \mathcal{E}_i$, $\mathcal{N}_{\mathbf{p}}^L := \bigcirc_{i=1}^L \mathcal{N}_{p_i}$, $\mathbf{p} = (p_1, p_2, \dots, p_L)$ とすると、

$$\mathcal{N} = \bigcirc_{i=1}^L \mathcal{N}_{p_i} \circ \mathcal{E}_i = \left(\bigcirc_{i=1}^L \mathcal{N}_{p_i} \right) \circ \left(\bigcirc_{i=1}^L \mathcal{E}_i \right) = \mathcal{N}_{\mathbf{p}}^L \circ \mathcal{E} \quad (2.9.9)$$

となる。つまり、 \mathcal{N} は、 \mathcal{E} の後に $\mathcal{N}_{\mathbf{p}}^L$ が作用することと同じである。簡単のため、 $p_i = p$ とすると、 $\mathcal{N}_{\mathbf{p}}^L(\rho)$ は次のように計算される。

$$\mathcal{N}_{\mathbf{p}}^L(\rho) = (1 - p)^L \rho + \{1 - (1 - p)^L\} \frac{\mathbb{1}}{2^n} \quad (2.9.10)$$

このとき、次の \mathcal{N} を得る。

$$\mathcal{N}(\rho) = \mathcal{N}_{\mathbf{p}}^L \circ \mathcal{E}(\rho) = (1 - p)^L \mathcal{E}(\rho) + \{1 - (1 - p)^L\} \frac{\mathbb{1}}{2^n} \quad (2.9.11)$$

第3章 変分量子アルゴリズム

この章では、NISQ 上でも実行可能なアルゴリズムとして現在盛んに研究されている変分量子アルゴリズムについて説明する。3.1 節では、変分量子アルゴリズムをコスト関数・変分量子回路・最適化の3つ部分に分けて説明する。3.2 節では、量子機械学習について説明する。3.3 節では、バレンプラトーについて説明する。

3.1 概要

現在実現されている量子コンピューターは、中規模（数百量子ビット）であり、ノイズが大きいため、Noisy Intermediate Scale Quantum computers (NISQ) と呼ばれている [9]。このため、使用できる量子ビットと量子ゲートの数に制約がある。そのような制約のもと、古典計算よりも効率的に計算を行うことを期待されているのが、変分量子アルゴリズムである。変分量子アルゴリズムは、古典コンピューターと組み合わせて最適化問題を解くアルゴリズムである。最適化問題とは、ある関数の値を最小化するようなパラメーターの値を求める問題である。変分量子アルゴリズムでは、パラメーターに依存した量子状態によってオブザーバブルの期待値を計算し、それを用いて定義されたコスト関数 $C(\theta)$ を最適化することによって問題を解く。パラメーターの更新を古典コンピューターに任せることで、量子回路を浅く保つことができる。図 3.1.1 のように、量子状態の生成からパラメーターの更新までのサイクルを繰り返すことで、最終的に最適なパラメーター $\theta_* = \arg \min_{\theta} C(\theta)$ を得る。試行関数を生成するパラメーター付き量子回路のことを、変分量子回路、アンザッツ (Ansatz)、あるいは、量子ニューラルネットワーク (Quantum Neural Network, QNN) と呼ぶ。

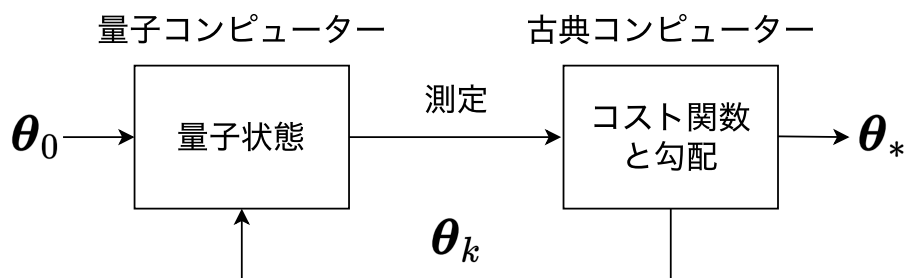


図 3.1.1: 変分量子アルゴリズムの概要

ある問題を解くための変分量子アルゴリズムを構成するには、以下の3つの要素が必要である。

1. 解きたい問題をコスト関数の最適化問題として定式化すること
2. コスト関数を最適化するための変分量子回路を構成すること
3. パラメーターを最適化するための最適化手法を選択すること

これらについて、順に説明する。

3.1.1 コスト関数

変分量子アルゴリズムにおけるコスト関数は、初期状態の集合 $\{\rho_i\}$ 、変分量子回路 $V(\boldsymbol{\theta})$ 、オブザーバブルの集合 $\{O_k\}$ を用いて定義される。変分量子回路を作用させた状態 $V(\boldsymbol{\theta})\rho_iV^\dagger(\boldsymbol{\theta})$ によるオブザーバブル O_k の期待値は、

$$\text{Tr}\left[V(\boldsymbol{\theta})\rho_iV^\dagger(\boldsymbol{\theta})O_k\right] \quad (3.1.1)$$

である。コスト関数は、この期待値を用いて定義される。ほとんどのコスト関数は、 f を任意の関数として次のような形で与えられる。

$$C(\boldsymbol{\theta}) = f\left(\left\{\text{Tr}\left[V(\boldsymbol{\theta})\rho_iV^\dagger(\boldsymbol{\theta})O_k\right]\right\}_{i,k}\right) \quad (3.1.2)$$

コスト関数は、その最小点が求める解に対応するように設計する。例えば、変分量子固有値ソルバー (Variational Quantum Eigensolver, VQE) では、ハミルトニアンが H で与えられる系の最小エネルギーを求めるために、 $\rho_i = |\psi_0\rangle\langle\psi_0|$ 、 $O_k = H$ とする。このとき、コスト関数は以下のように定義される。

$$C(\boldsymbol{\theta}) = \langle\psi_0|V^\dagger(\boldsymbol{\theta})HV(\boldsymbol{\theta})|\psi_0\rangle \quad (3.1.3)$$

また、変分量子回路を用いた機械学習 (量子機械学習) では、データセット $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ が与えられたとき、入力状態を $\rho_i = U(\mathbf{x}_i)|0\rangle\langle 0|U^\dagger(\mathbf{x}_i)$ 、オブザーバブルを $O_k = O$ として、各データに依存するオブザーバブルの期待値 $\text{Tr}[V(\boldsymbol{\theta})\rho_iV^\dagger(\boldsymbol{\theta})O]$ が対応するラベル y_i になるように、以下のようなコスト関数を定義することができる。

$$C(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \left(\text{Tr}\left[V(\boldsymbol{\theta})\rho_iV^\dagger(\boldsymbol{\theta})O\right] - y_i\right)^2 \quad (3.1.4)$$

3.1.2 変分量子回路

一般に、変分量子回路 $V(\boldsymbol{\theta})$ は、回転ゲートなどのパラメーター付き量子ゲートと CNOT ゲートなどのパラメーターなし量子ゲートの積で表すことができる。 $U_j(\boldsymbol{\theta}_j)$ を任意の数のパラメーター付き量子ゲートからなるユニタリ、 W_j を任意の数のパラメーターなし量子ゲートからなるユニタリであるとすると、 $V(\boldsymbol{\theta})$ は次のように表わすことができる。

$$V(\boldsymbol{\theta}) = \prod_{j=1}^L U_j(\boldsymbol{\theta}_j)W_j = U_L(\boldsymbol{\theta}_L)W_L \cdots U_1(\boldsymbol{\theta}_1)W_1 \quad (3.1.5)$$

変分量子回路においては、類似した構造を複数回繰り返すことがあり、その構造単位の繰り返しの数を層数と呼び、この数が多いほど深い回路であるという。図 3.1.2 は L 層の変分量子回路を表している。

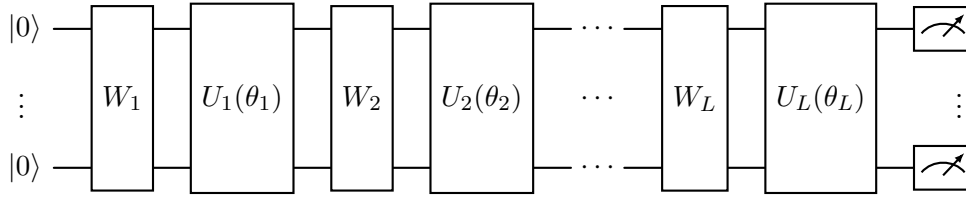


図 3.1.2: L 層の変分量子回路

アンザツの構造は、問題の性質に応じて様々なものが考えられる。問題の対称性や制約を構造に取り入れたアンザツを Problem-inspired アンザツと呼ぶ。例えば、Quantum Alternating Operator Ansatz (QAOA) は組合せ最適化問題に対して提案された Problem-inspired アンザツである。一方、問題の性質に依存しないアンザツを Problem-agnostic アンザツと呼ぶ。

Hardware Efficient Ansatz は量子コンピューターの実機の構造を考慮し、実装が効率的にできるよう設計された Problem-agnostic アンザツである。例えば、量子ビット q_1, q_2, q_3, q_4 が隣り合うインデックスのみ実機上でつながっているときに、図 3.1.3 のような回路の構造を持つアンザツである。これにより、実機上で隣り合わない量子ビットを結ぶための量子ゲートを実装する必要がなくなり、使用する SWAP ゲートの数が減る。SWAP ゲートは CNOT ゲート 3 つから成るが、CNOT ゲートは量子ゲートの中でも特にノイズが多い。そのため、SWAP ゲートを減らすことはノイズの対策として有用である。

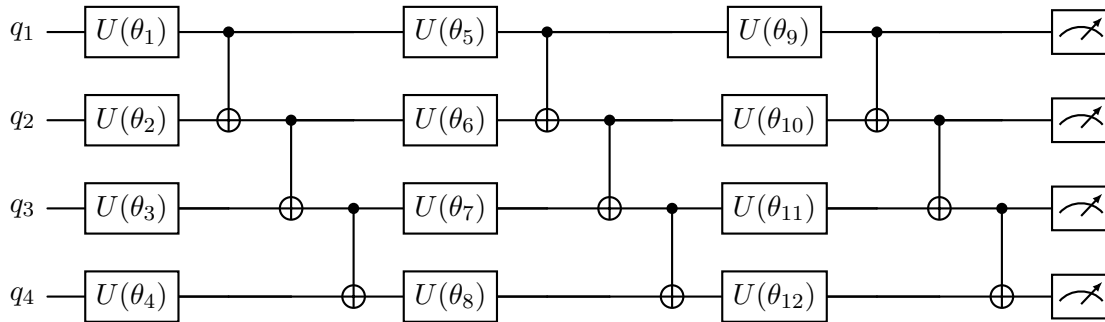


図 3.1.3: Hardware Efficient Ansatz

3.1.3 最適化

最適化とは、コスト関数を最小化するパラメーターを求めることであり、変分量子アルゴリズムでは古典コンピューターによって行われる。最適化のアルゴリズムのことを、オプティマイザーと呼ぶ。変分量子アルゴリズムにおいては、従来の古典計算で用いられるオプティマイザーをそのまま用いることができるが、量子計算に合わせて改良されたオプティマイザーも提案されている。オプティマイザーには、ノイズや測定誤差に強いものを選択する必要がある。

オプティマイザーには、コスト関数の勾配を用いるものと、用いないものに大別される。勾配を用いるオプティマイザーには、勾配降下法、ニュートン法、量子自然勾配法 [36]、確率的勾配降下法 [37]、Adam[38]、などの方法がある。勾配自体は、有限差分法やパラメーターシフトルール [19, 39] によって

求めることができる。例えば、勾配降下法では、パラメーター θ_j を以下のように更新する。

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial C(\boldsymbol{\theta})}{\partial \theta_j} \quad (3.1.6)$$

ここで、 α は 1 回の更新でのパラメーターの変化量を決める、学習率と呼ばれるハイパーパラメーターである。通常、学習率は小さい正の値に設定されるが、適切な値は扱う問題によって異なる。

勾配を用いない方法では、コスト関数の値のみを用いてパラメーターを更新する。例えば、Nelder-Mead[40]、Powell[41]、COBYLA (Constrained Optimization By Linear Approximation optimizer)[42]、SPSA (Simultaneous Perturbation Stochastic Approximation)[43]、逐次最小化問題最適化法 [44–47] などが挙げられる。

パラメーターシフトルール

変分量子アルゴリズムにおいては、すべてのパラメーターが独立であれば、パラメーターシフトルールと呼ばれる手法によって、厳密に勾配を求めることができる [19, 39] (測定誤差は除く)。パラメーターシフトルールは、パラメーター付き量子ゲート $V(\boldsymbol{\theta})$ によるオブザーバブルの期待値のパラメーター θ_j に関する勾配を、パラメーター θ_j を一定値ずらして得た期待値の差として求める方法である。

定理 3.1.1. (パラメーターシフトルール [19, 39]) 量子回路 $V(\boldsymbol{\theta})$ はパラメーター付き量子ゲート $U_j(\theta_j) := \exp(-i\theta_j P_j/2)$ とパラメーターなし量子ゲート W_j によって $V(\boldsymbol{\theta}) := U_L(\theta_L)W_L \cdots U_1(\theta_1)W_1$ と定義される。ここでは、すべてのパラメーターは独立であるとする。このとき、あるオブザーバブル O の期待値 $\langle O(\boldsymbol{\theta}) \rangle = \text{Tr}[V(\boldsymbol{\theta})|0\rangle\langle 0|^{\otimes n} V^\dagger(\boldsymbol{\theta})O]$ の θ_j に関する勾配は

$$\frac{\partial \langle O(\boldsymbol{\theta}) \rangle}{\partial \theta_j} = \frac{1}{2} \left[\left\langle O\left(\boldsymbol{\theta} + \frac{\pi}{2} \mathbf{e}_j\right) \right\rangle - \left\langle O\left(\boldsymbol{\theta} - \frac{\pi}{2} \mathbf{e}_j\right) \right\rangle \right] \quad (3.1.7)$$

で与えられる。ここで \mathbf{e}_j は j 番目の方向の単位ベクトルである。

より高次の勾配も同様に計算することができる。例えば、 $V(\boldsymbol{\theta})$ の θ_j と θ_k に関する勾配は、

$$\frac{\partial^2 \langle O(\boldsymbol{\theta}) \rangle}{\partial \theta_j \partial \theta_k} = \frac{1}{2} \left[\left. \frac{\partial \langle O(\boldsymbol{\theta}) \rangle}{\partial \theta_j} \right|_{\theta_k = \theta_k + \frac{\pi}{2}} - \left. \frac{\partial \langle O(\boldsymbol{\theta}) \rangle}{\partial \theta_j} \right|_{\theta_k = \theta_k - \frac{\pi}{2}} \right] \quad (3.1.8)$$

で与えられる。

一般的なコスト関数 (3.1.2) の勾配については、連鎖律を用いると、

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} f \left(\left\{ \text{Tr}[V(\boldsymbol{\theta})\rho_i V^\dagger(\boldsymbol{\theta})O_k] \right\}_{i,k} \right) = \sum_{i,k} \frac{\partial f}{\partial \ell_{i,\boldsymbol{\theta},k}} \frac{\partial \ell_{i,\boldsymbol{\theta},k}}{\partial \theta_j} \quad (3.1.9)$$

となるので、 $\ell_{i,\boldsymbol{\theta},k}$ にパラメーターシフトルールを用いて勾配を計算することができる。ここで、 $\ell_{i,\boldsymbol{\theta},k} := \text{Tr}[V(\boldsymbol{\theta})\rho_i V^\dagger(\boldsymbol{\theta})O_k]$ とした。

3.2 量子機械学習

近年、機械学習¹はさまざまな分野で著しい成果を上げているが、その一方で、高い計算コストが問題となっている。そこで、機械学習の計算を量子コンピューター上で行い、計算を効率化できないか研究されている。そのような研究分野を量子機械学習 (Quantum Machine Learning, QML) と呼ぶ。量子機械学習におけるアルゴリズムの研究アプローチは次の 3 つがあげられる [48]。

1. 古典的な学習方法から量子的なものへの再構築
2. 古典的な学習方法に勝る新たな量子機械学習の方法の探索
3. 近い将来に向けた、NISQ に特化した学習方法の探索

2009 年、Harrow, Hassidim, Lloyd は線形方程式を解く量子アルゴリズム (HHL アルゴリズム) を提案した [8]。このアルゴリズムは、古典的なアルゴリズムよりも効率的に線形方程式を解くことが期待される量子アルゴリズムである。量子機械学習の初期の研究は、機械学習における線形代数の問題を HHL アルゴリズムによって量子コンピューター上で解くことに焦点を当てていた (1 のアプローチ)。しかし、HHL アルゴリズムは量子位相推定を用いており、量子ビットと量子ゲートが数多く必要であるため、NISQ で実現できるアルゴリズムではなかった。変分量子アルゴリズムが登場したことにより、その文脈で量子機械学習の研究が盛んになった (3 のアプローチ)。この節では、変分量子アルゴリズムとして研究されている 3 つの教師あり量子機械学習モデルについて説明する。量子回路学習 (Quantum Circuit Learning) [19, 20]、Data reuploading[49–51]、量子カーネル法 (Quantum Kernel Methods) [21, 22, 52] である。いずれのモデルも、図 3.1.1 にある流れに従って学習を行うが、量子回路で行う計算が異なる。

3.2.1 量子回路学習

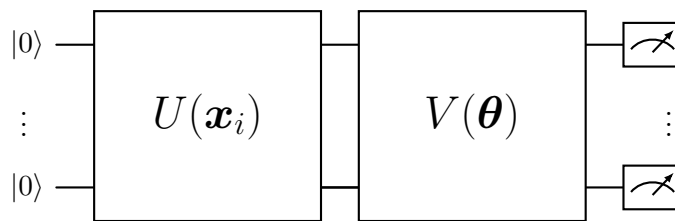


図 3.2.1: 量子回路学習における量子回路の構造

図 3.2.1 は量子回路学習における量子回路の構造を表している。まず、初期状態 $|0\rangle^{\otimes n}$ にデータを入力するためのユニタリ $U(\mathbf{x}_i)$ を作用させ、続いて、学習パラメータを入力するためのユニタリ $V(\boldsymbol{\theta})$ を作用させる。前者を入力回路、後者を学習回路と呼ぶことにする。これにより、状態 $V(\boldsymbol{\theta})U(\mathbf{x}_i)|0\rangle^{\otimes n}$ が生成される。そして、この状態によって測定したオブザーバブルの期待値 $\{\text{Tr}[V(\boldsymbol{\theta})U(\mathbf{x}_i)|0\rangle\langle 0|^{\otimes n}U^\dagger(\mathbf{x}_i)V^\dagger(\boldsymbol{\theta})O_k]\}_k$ を用いて、3.1.4 のようなコスト関数が定義される。古典コンピューター上では、コスト関数とその勾配を計算し、学習パラメータを更新する。このプロセスを繰り返すことで、最適なパラメータ $\boldsymbol{\theta}_*$ を得る。後の第 5,6 章において扱う量子機械学習のモデルも、この量子回路学習に基づいて解析を行う。

¹機械学習の概要については付録の A.1 節に置いた。

3.2.2 Data reuploading

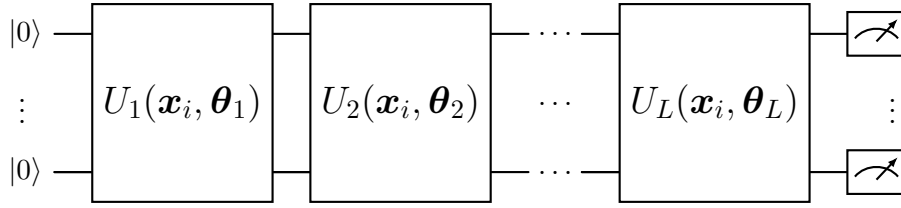


図 3.2.2: Data reuploading における量子回路の構造

量子回路学習とは異なり、いくつかの研究においては、データの入力と学習パラメータの入力を交互に繰り返す量子回路 [50, 51] や、一つのゲートにデータと学習パラメータの線形和を入力する量子回路 [49] が用いられている。このような量子回路の構造は、Data reuploading と呼ばれている (図 3.2.2)。コスト関数の定式化は、量子回路学習と同様である。

3.2.3 量子カーネル法

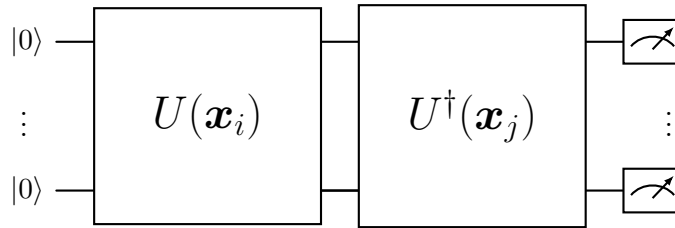


図 3.2.3: 量子カーネル法における量子回路の構造

カーネル法は、入力データを高次元の特徴空間に写像し、その空間での内積を用いることで非線形な分類や回帰を行う機械学習である。カーネル法のコスト関数は凸であり、局所的な解に陥ることがないので、量子回路学習と Data reuploading による機械学習よりも、学習が安定すると期待されている [52]。量子カーネル法では、カーネルの計算のみを量子コンピューター上で行う [21, 22]。例えば、データ $\mathbf{x}_i, \mathbf{x}_j$ に依存する量子状態

$$\begin{aligned} \rho(\mathbf{x}_i) &:= U(\mathbf{x}_i) |0\rangle\langle 0|^{\otimes n} U^\dagger(\mathbf{x}_i), \\ \rho(\mathbf{x}_j) &:= U(\mathbf{x}_j) |0\rangle\langle 0|^{\otimes n} U^\dagger(\mathbf{x}_j) \end{aligned}$$

を用いて定義される量子カーネル (Fidelity Quantum Kernel) $\kappa_{FQ}(\mathbf{x}_i, \mathbf{x}_j) := \text{Tr}[\rho(\mathbf{x}_i)\rho(\mathbf{x}_j)]$ を計算するための量子回路は図 3.2.3 で与えられる。初期状態 $|0\rangle^{\otimes n}$ に対してユニタリ $U(\mathbf{x}_i), U^\dagger(\mathbf{x}_j)$ を順に作用させて、状態 $U^\dagger(\mathbf{x}_j)U(\mathbf{x}_i) |0\rangle^{\otimes n}$ を作り、 $|0\rangle\langle 0|^{\otimes n}$ によって測定を行う。これによって得られる期待値は、確かに $\text{Tr}[\rho(\mathbf{x}_i)\rho(\mathbf{x}_j)]$ に等しい。

$$\text{Tr}\left[U^\dagger(\mathbf{x}_j)U(\mathbf{x}_i) |0\rangle\langle 0|^{\otimes n} (U^\dagger(\mathbf{x}_j)U(\mathbf{x}_i))^\dagger |0\rangle\langle 0|^{\otimes n}\right] = \text{Tr}[\rho(\mathbf{x}_i)\rho(\mathbf{x}_j)] \quad (3.2.1)$$

しかしながら、量子カーネル $\kappa_{FQ}(\mathbf{x}_i, \mathbf{x}_j)$ は、訓練可能性²、汎化性能が悪いことが知られている [30, 33, 53]。そのため、Bandwidth λ を導入した量子カーネル $\kappa_{FQ}^\lambda(\mathbf{x}_i, \mathbf{x}_j) := \kappa_{FQ}(\lambda\mathbf{x}_i, \lambda\mathbf{x}_j)$ を用いること [54, 55] や、次式で定義されるカーネル (Projected Quantum Kernel) $\kappa_{PQ}(\mathbf{x}_i, \mathbf{x}_j)$ が用いることが提案されている [33]。

$$\kappa_{PQ}(\mathbf{x}_i, \mathbf{x}_j) := \exp\left(-\gamma \sum_{i=1}^n \left\| \rho^{(k)}(\mathbf{x}_i) - \rho^{(k)}(\mathbf{x}_j) \right\|_2^2\right) \quad (3.2.2)$$

ただし、 $\rho^{(k)}(\mathbf{x}_i)$ は、 k 番目の量子ビットにおける縮約密度演算子である。

3.3 バレンプラトー

3.3.1 概要

変分量子アルゴリズムにおいては、いくつかの原因によってバレンプラトーと呼ばれるコスト関数の勾配消失 (平坦化) が起こりうる [23]。バレンプラトーに陥ると、量子回路のパラメーターの更新に量子ビット数に関して指数関数的な計算量が必要となるため、古典計算以上の効率化は不可能である。この節では、バレンプラトーの定義、原因、提案されている回避方法について述べる。まず初めに、バレンプラトーの定義を与える。

定義 3.3.1. バレンプラトーとは、コスト関数 $C(\theta)$ のあるパラメーター θ_ν に関する勾配が次の性質を持つことである [23]。

$$\mathbb{E}_{V(\theta)} \left[\frac{\partial C(\theta)}{\partial \theta_\nu} \right] = 0, \quad \text{Var}_{V(\theta)} \left[\frac{\partial C(\theta)}{\partial \theta_\nu} \right] \in \mathcal{O}(2^{-\alpha n}), \quad 0 < \exists \alpha \quad (3.3.1)$$

このように、バレンプラトーとは、コスト関数の勾配の平均が 0 になり、勾配の分散が量子ビット数に関して指数関数的に小さくなる現象を指す。 X を確率変数とすると、 $\forall \delta > 0$ に対して次のチェビシェフの不等式が成り立つ。

$$\Pr[|X - \mathbb{E}[X]| \geq \delta] \leq \frac{\text{Var}[X]}{\delta^2} \quad (3.3.2)$$

コスト関数の勾配を確率変数とみなすと、チェビシェフの不等式より、

$$\Pr \left[\left| \frac{\partial C(\theta)}{\partial \theta_\nu} \right| \geq \delta \right] \leq \frac{1}{\delta^2} \text{Var}_{V(\theta)} \left[\frac{\partial C(\theta)}{\partial \theta_\nu} \right] \in \mathcal{O}(2^{-\alpha n}) \quad (3.3.3)$$

となる。よって、量子ビット数を増やすにつれて、勾配が 0 から離れる確率が指数関数的に小さくなるのがわかる。変分量子アルゴリズムにおいて、パラメーターの初期値として適当なものがわからない場合、しばしば初期値はランダムに選ばれる。バレンプラトーが起きる場合、ほとんどの初期値は勾配が 0 の平坦な領域にあるため、最適化は難しい。

²訓練可能性 (Trainability) とは、最適化ができるかどうかということである。量子カーネル $\kappa_{FQ}(\mathbf{x}_i, \mathbf{x}_j)$ は定義からグローバルオプザーバブル $|0\rangle\langle 0|^{\otimes n}$ を用いるため、勾配消失 (バレンプラトー) が生じ、最適化が困難になる。詳しくは次の節を参照。

これを視覚的に示したものが図 3.3.1 である。バレンプラトーが起きるとき、あるパラメーターに関するコスト関数の断面は、量子ビット数が多いほど平坦な領域（勾配がほぼ 0）が増え、狭い谷（narrow gorge）を持つ [56]。

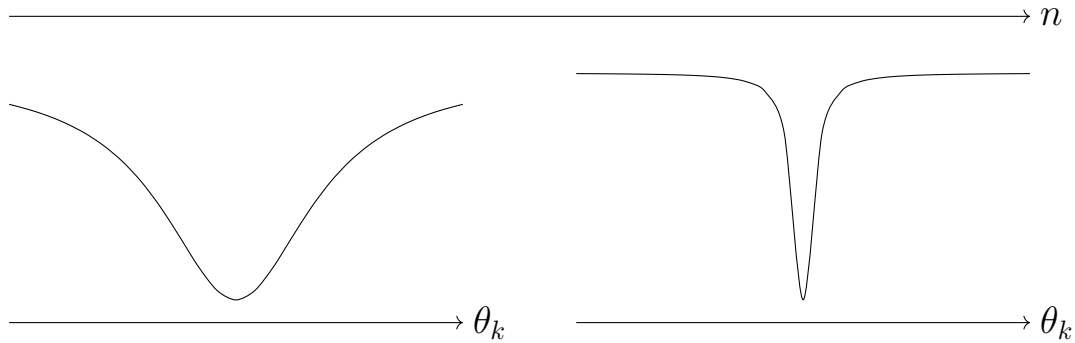


図 3.3.1: Noise-free バレンプラトーにおけるコスト関数の断面の変化のイメージ

定義 3.3.1 によるバレンプラトーのことを Noise-free バレンプラトー、あるいは確率論的なバレンプラトーと呼ぶ。一方で、論文 [23] の後に示されたような、ノイズが原因のバレンプラトーのことを Noise-induced バレンプラトー、あるいは決定論的なバレンプラトーと呼ぶ。この場合、コスト関数の勾配の大きさが、層数 L に関して指数関数的に小さくなる [28]。特に、 $L \in \mathcal{O}(n)$ のとき、次のようになる。

$$\left| \frac{\partial C(\boldsymbol{\theta})}{\partial \theta_\nu} \right| \in \mathcal{O}(2^{-\alpha n}), \quad 0 < \exists \alpha \tag{3.3.4}$$

これを視覚的に示したものが図 3.3.2 である。コスト関数全体が収縮しているため、さらに最適化が難しくなる。

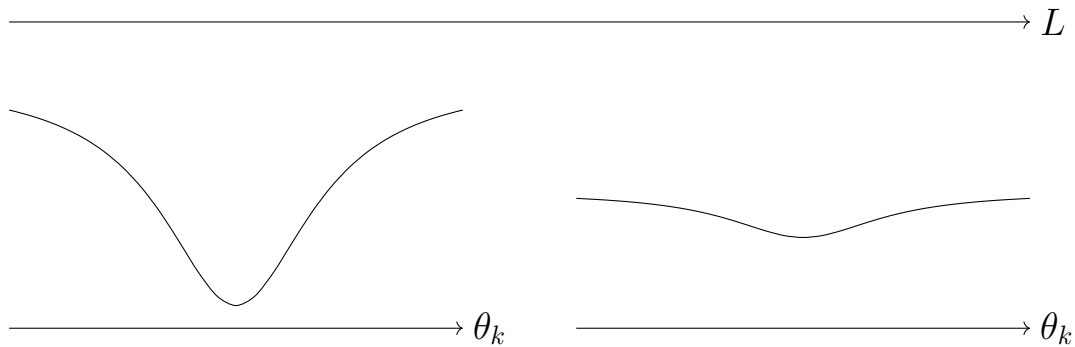


図 3.3.2: Noise-induced バレンプラトーにおけるコスト関数の断面の変化のイメージ

そもそも、コスト関数やその勾配は測定によってのみ得られるため、測定誤差が必ず存在する。もし、その測定誤差によってコスト関数の勾配の推定値と実際の勾配の正負が逆になると、最適化が進まない。測定回数が N_{shot} の場合、測定誤差は $\mathcal{O}(1/\sqrt{N_{\text{shot}}})$ である。バレンプラトーが起きる場合、コスト関数の勾配の分散は指数関数的に小さくなるため、その値の正負を判別するには、測定誤差も指数関数的に小さくする必要がある。これは、測定回数を指数関数的に増やす必要があることを意味する。

変分量子アルゴリズムの目的は、コスト関数を最小化するパラメーターを、量子ビット数の高々多項式の計算量で求めることである。バレンプラトーが起きる場合、コスト関数の最小点を量子ビット数の

多項式の計算量で求めることはできない。高次の勾配に関しても、同様の現象が見られることが示されている [57]。また、勾配を用いないオプティマイザーを用いる場合でも、必要な測定回数は量子ビット数に関して指数関数的に増えることが知られている [24]。これを表しているのが、図 3.3.3 である。したがって、オプティマイザーを変えることでバレンプラトーを回避することはできない。

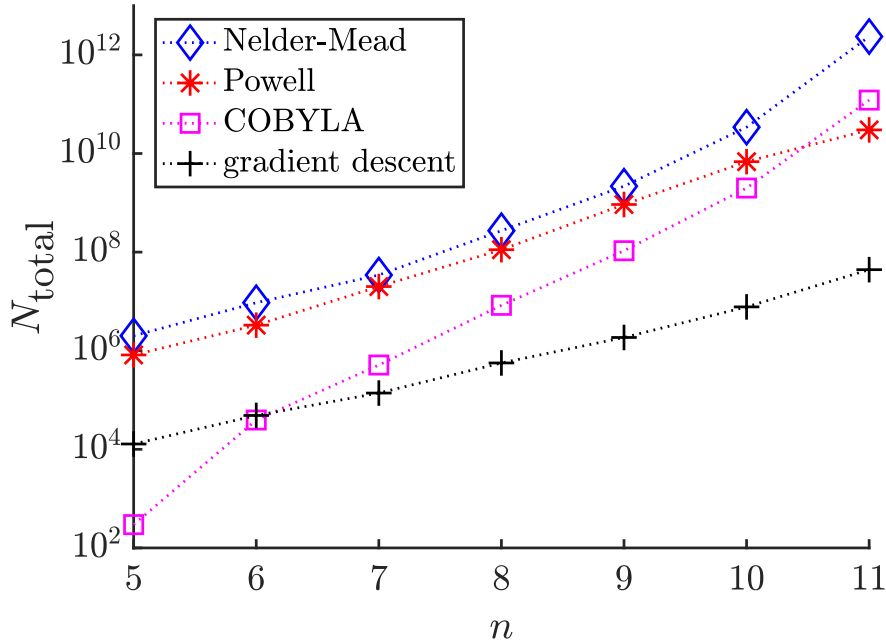


図 3.3.3: バレンプラトーにおいて必要なオブザーバブルの測定回数 N_{total} の量子ビット数 n に対する変化 [24]。Nelder-Mead・Powell・COBYLA は勾配を用いないオプティマイザーである。

3.3.2 原因

バレンプラトーの原因として既に知られているものを以下に挙げる。

- 学習回路の深さ・学習回路の表現能力・学習回路の生成するエンタングルメント [23, 25–27, 58]
- オブザーバブルの局所性 [27, 59]
- ノイズ [28]
- データの入力 [29, 30, 58]

バレンプラトーは、学習回路がユニタリ 2-デザイン (4.2 節) という条件を満たすときに起きることが最初に示された [23]。具体的な定理と証明については、付録の A.2 節に記す。学習回路がユニタリ 2-デザインを満たすことは、学習回路が深く、高い表現能力を持ち、多くのエンタングルメントを生成することを意味する。

オブザーバブルの局所性とは、測定するオブザーバブルが全量子ビットのうちいくつかの量子ビットによって測定されるかを指す。グローバルオブザーバブルとは、全量子ビットの状態によって測定されるオブザーバブルのことであり、例えば、 $Z^{\otimes n}$ である。ローカルオブザーバブルとは、全量子ビット数 n に依存しない m 個の量子ビットによって測定されるオブザーバブルのことである。例えば、 $Z^{\otimes m}$ である。また、ローカルオブザーバブルの線形和もローカルオブザーバブルである。

ここで、量子回路の構造・量子ビット数 n ・量子回路の層数 L ・オブザーバブルの局所性がコスト関数の勾配の分散に与える影響を再現したものを示す。量子回路の構造として図 3.3.4 にあるように、Tensor Product Ansatz (TPA) と Hardware Efficient Ansatz (HEA) を用いた。

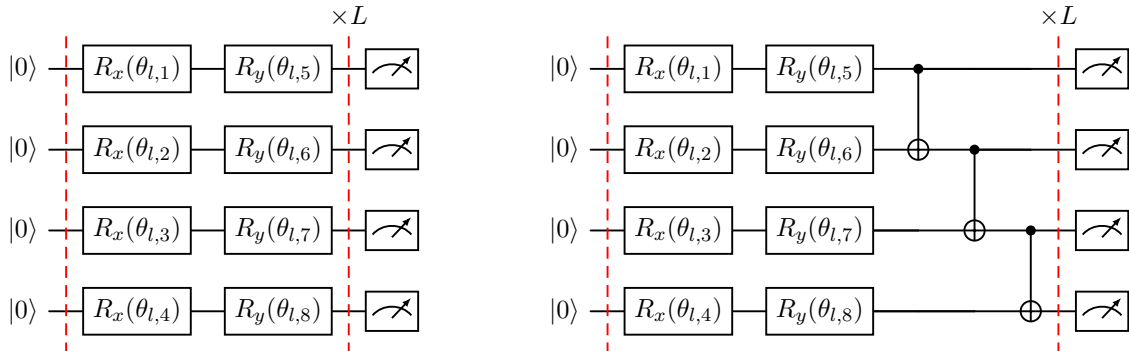


図 3.3.4: TPA (左) と HEA (右) の量子回路

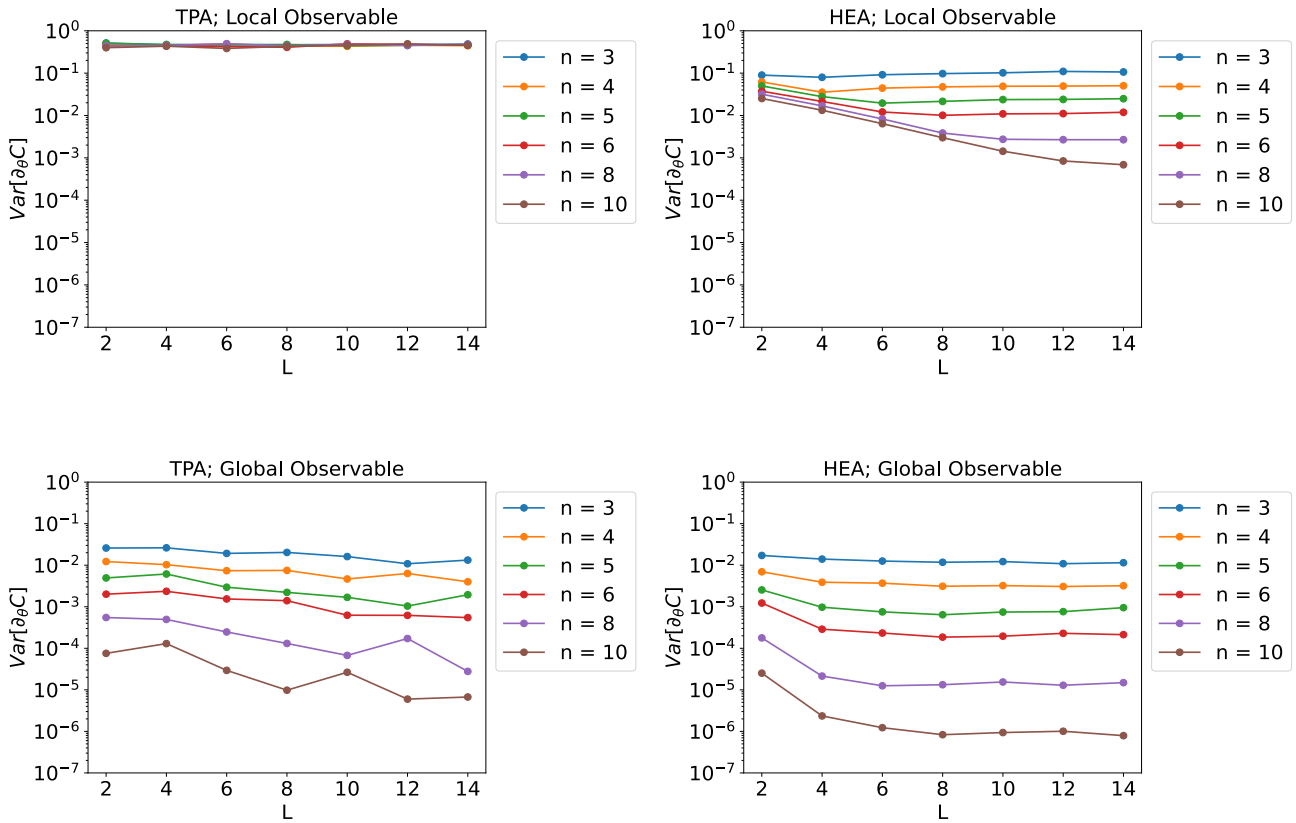


図 3.3.5: それぞれのプロットの横軸は量子回路の各層数 L 、縦軸はコスト関数の勾配の分散を表す。上段は、TPA (左) と HEA (右) に対してローカルオブザーバブル $Z \otimes I \otimes \dots \otimes I$ を用いた場合の勾配の分散の変化を示している。下段は、グローバルオブザーバブル $|0\rangle\langle 0|^{\otimes n}$ を用いた場合のそれぞれの量子回路における勾配の分散の変化を示している。

図 3.3.5 において、上段は、Tensor Product Ansatz (TPA) (左) と Hardware Efficient Ansatz (HEA) (右) に対してローカルオブザーバブル $Z \otimes I \otimes \dots \otimes I$ を用いた場合の、各量子ビット数 n と量子回路の各層数 L に対する勾配の分散の変化を示している。下段は、グローバルオブザーバブル $|0\rangle\langle 0|^{\otimes n}$ を

用いた場合の勾配の分散の変化を示している。これらの図から、エンタングルメントを多く生成すること、層数 L を増やすこと、グローバルオブザーバブルを用いることはいずれも、勾配の分散のスケールをより小さくすることがわかる。

3.3.3 回避方法

豊かな表現能力を持ち、エンタングルメントを多く生成するような深い量子回路を用いることと、グローバルオブザーバブルを用いることは、パラメーターの探索すべき領域を指数関数的に増大させるため、バレンプラトールを引き起こすと考えられている。よって、バレンプラトールを回避するためには、学習回路の深さ・表現能力・エンタングルメントを抑えて、ローカルオブザーバブルを用い、かつ、学習回路の表現領域にコスト関数の最適解があるようにその構造を設計する必要がある。これらの原因を考慮して、バレンプラトールを回避するための方法がいくつか提案されている。

- ローカルオブザーバブルを用いる [27]。これにより、データ入力がない場合は、学習回路の深さが $O(\log n)$ であればバレンプラトール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 [60]。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 [61]。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 [26]。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 [62]。

また、QAOA では、最適なパラメーターがあるパターンを持つことが先行研究 [63] で示された。このパターンに近いパラメーターを初期値として用いることで、ランダムにパラメーターを選んだ場合よりも学習が早くなることも示されている。また、Quantum Tree Tensor Network (QTTN) や Quantum Convolutional Neural Network (QCNN) はその構造上、バレンプラトールが起きないことが示されている [32, 64]。

しかしながら、バレンプラトールを避けることだけであれば良いわけではない。というのも、例えば、バレンプラトールを避けるためにエンタングルメントが生じない量子回路を用いたとすると、その量子回路は古典コンピューターによって効率的にシミュレーションすることができる。よって、変分量子アルゴリズムにおいて古典計算よりも効率的な計算を行うためには、バレンプラトールを避ける一方で、古典コンピューターによるシミュレーションが困難な量子回路を用いる必要がある。

第4章 ハール分布と量子回路の表現力

変分量子アルゴリズムにおいては、しばしば、パラメーターの初期値がランダムに選ばれ、それがバレンプラトールと関係することを前章で述べた。ユニタリ群における一様なランダムネスを考える場合、ハール分布やユニタリデザインという概念が重要になる。後の5, 6章の解析においても、これらの概念に基づいて解析を行う。よって、この章においては、ハール分布やユニタリデザインの定義から始め、それらの性質や関連する定理、更に量子回路の表現力と呼ばれる量について説明する。4.1節では、ハール分布とそれに関する積分の公式について説明する。4.2節でユニタリ t -デザインというユニタリの集合に関する性質について説明する。4.3節では、量子回路の表現力について説明する。

4.1 ハール分布

ハール分布（測度）とはコンパクト群上の一様分布であり、ユニタリ群上のハール分布 μ_{Haar} は、ユニタリ群 $U(d)$ 上のユニタリ不変な測度であって、以下の性質を満たすものである。

$$\mu_{\text{Haar}}(U(d)) = \text{const.}, \quad \mu_{\text{Haar}}(gU) = \mu_{\text{Haar}}(Ug) = \mu_{\text{Haar}}(U), \quad \forall g \in U(d), \quad \forall U \subset U(d) \quad (4.1.1)$$

ここでは確率分布を考えるため $\mu_{\text{Haar}}(U(d)) = 1$ とする。ハールランダムなユニタリ U とは、ハール分布 μ_{Haar} からサンプリングされたユニタリ U のことを指す。

2×2 次元のハールランダムなユニタリをサンプリングして $|0\rangle$ に作用させると、それらの状態は、図 4.1.1 のようにブロッホ球面上で一様に分布する。

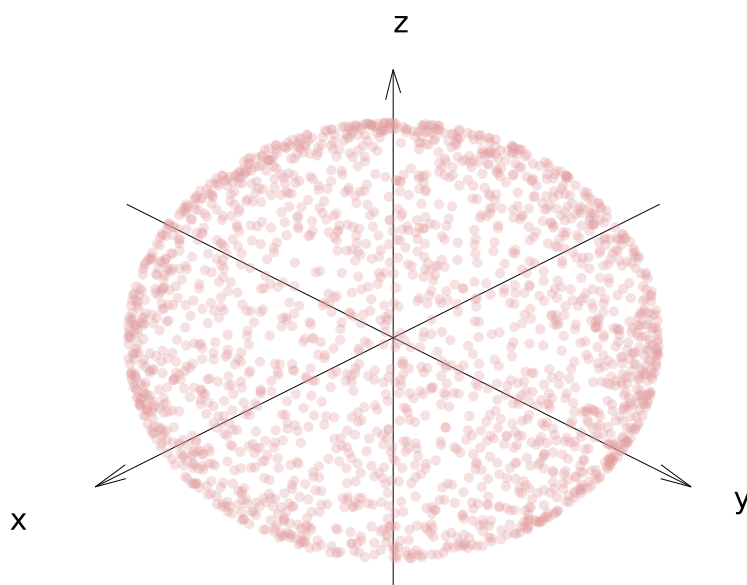


図 4.1.1: ブロッホ球面上でのハール分布

ユニタリ群上のハール分布においては、次のような積分公式が成り立つ。

定理 4.1.1. ハール分布に従うユニタリ W に関する積分は以下のように計算できる。

$$\int_{U(d)} d\mu_{\text{Haar}}(W) w_{i,j} w_{p,q}^* = \frac{\delta_{i,p} \delta_{j,q}}{d} \quad (4.1.2)$$

$$\begin{aligned} \int_{U(d)} d\mu_{\text{Haar}}(W) w_{i_1,j_1} w_{i_2,j_2} w_{i'_1,j'_1}^* w_{i'_2,j'_2}^* &= \frac{1}{d^2-1} (\delta_{i_1,i'_1} \delta_{i_2,i'_2} \delta_{j_1,j'_1} \delta_{j_2,j'_2} + \delta_{i_1,i'_2} \delta_{i_2,i'_1} \delta_{j_1,j'_2} \delta_{j_2,j'_1}) \\ &\quad - \frac{1}{d(d^2-1)} (\delta_{i_1,i'_1} \delta_{i_2,i'_2} \delta_{j_1,j'_2} \delta_{j_2,j'_1} + \delta_{i_1,i'_2} \delta_{i_2,i'_1} \delta_{j_1,j'_1} \delta_{j_2,j'_2}) \end{aligned} \quad (4.1.3)$$

たとえば、公式 (4.1.2) を用いると、次のような積分を計算することができる。

$$\int_{U(d)} d\mu_{\text{Haar}}(W) |\langle \mathbf{0} | W | \mathbf{0} \rangle|^2 = \int_{U(d)} d\mu_{\text{Haar}}(W) w_{\mathbf{0},\mathbf{0}} w_{\mathbf{0},\mathbf{0}}^* = \frac{1}{d} \quad (4.1.4)$$

さらに、公式 (4.1.3) を用いることで、次の公式を導くことができる。

$$\begin{aligned} \int_{U(d)} d\mu_{\text{Haar}}(W) \text{Tr}[WAW^\dagger B] \text{Tr}[WCW^\dagger D] &= \frac{1}{d^2-1} (\text{Tr}[A] \text{Tr}[B] \text{Tr}[C] \text{Tr}[D] + \text{Tr}[AC] \text{Tr}[BD]) \\ &\quad - \frac{1}{d(d^2-1)} (\text{Tr}[AC] \text{Tr}[B] \text{Tr}[D] + \text{Tr}[A] \text{Tr}[C] \text{Tr}[BD]) \end{aligned} \quad (4.1.5)$$

$$\begin{aligned} \int_{U(d)} d\mu_{\text{Haar}}(W) \text{Tr}[AW^{\otimes 2} B (W^\dagger)^{\otimes 2}] &= \frac{1}{d^2-1} (\text{Tr}[A] \text{Tr}[B] + \text{Tr}[A \text{ SWAP}] \text{Tr}[B \text{ SWAP}]) \\ &\quad - \frac{1}{d(d^2-1)} (\text{Tr}[A \text{ SWAP}] \text{Tr}[B] + \text{Tr}[A] \text{Tr}[B \text{ SWAP}]) \end{aligned} \quad (4.1.6)$$

ただし、 $\text{SWAP} = \sum_{i,j=1}^d |ij\rangle \langle ji|$ である。証明は、論文 [27] の Supplementary Note 1 を参照。

これらの公式とテンソルネットワークによる記法を組み合わせることで、ハール分布に関する積分を行列の脚の繋ぎ替えたものの和として表現することができる。テンソルネットワークとは、大きな量子状態を効率的に表現するための方法である。以下に、必要なテンソルネットワークの記法を示す。

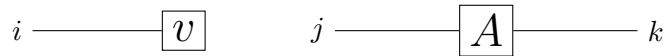


図 4.1.2: ベクトルと行列

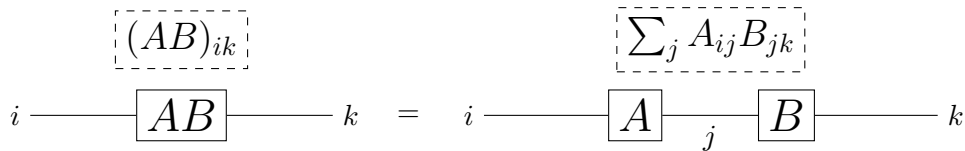


図 4.1.3: 行列の積



図 4.1.4: トレース

例えば、公式 (4.1.2) を用いて、次の積分公式を示せる。

$$\int_{U(d)} d\mu_{\text{Haar}}(U) AUBU^\dagger C = \frac{\text{Tr}[B]AC}{d} \tag{4.1.7}$$

テンソルネットワークによる記法を用いると、図 4.1.5 のように表現することができる。

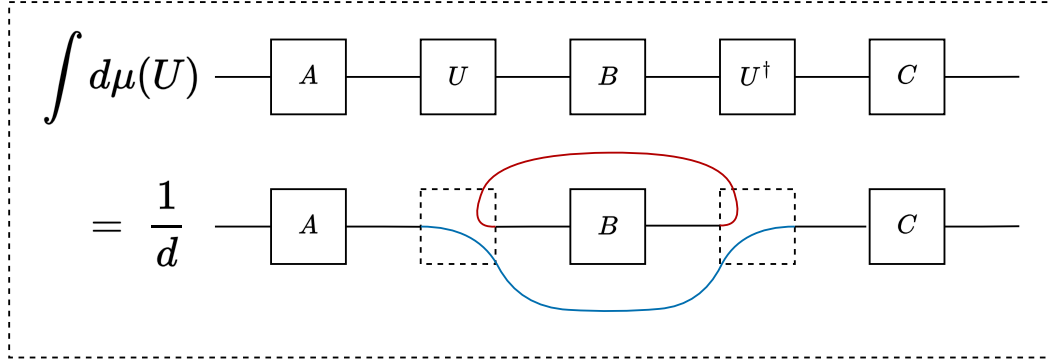


図 4.1.5: 公式 (4.1.7) に対応するテンソルネットワーク

また、公式 (4.1.3) を用いて、次の積分公式を示せる。

$$\begin{aligned} \int_{U(d)} d\mu_{\text{Haar}}(U) AUBU^\dagger CUDU^\dagger E &= \frac{1}{d^2 - 1} (\text{Tr}[B] \text{Tr}[D]ACE + \text{Tr}[BD] \text{Tr}[C]AE) \\ &\quad - \frac{1}{d(d^2 - 1)} (\text{Tr}[BD]ACE + \text{Tr}[B] \text{Tr}[C] \text{Tr}[D]AE) \end{aligned} \tag{4.1.8}$$

テンソルネットワークによる記法を用いると、図 4.1.6 のように表現することができる。

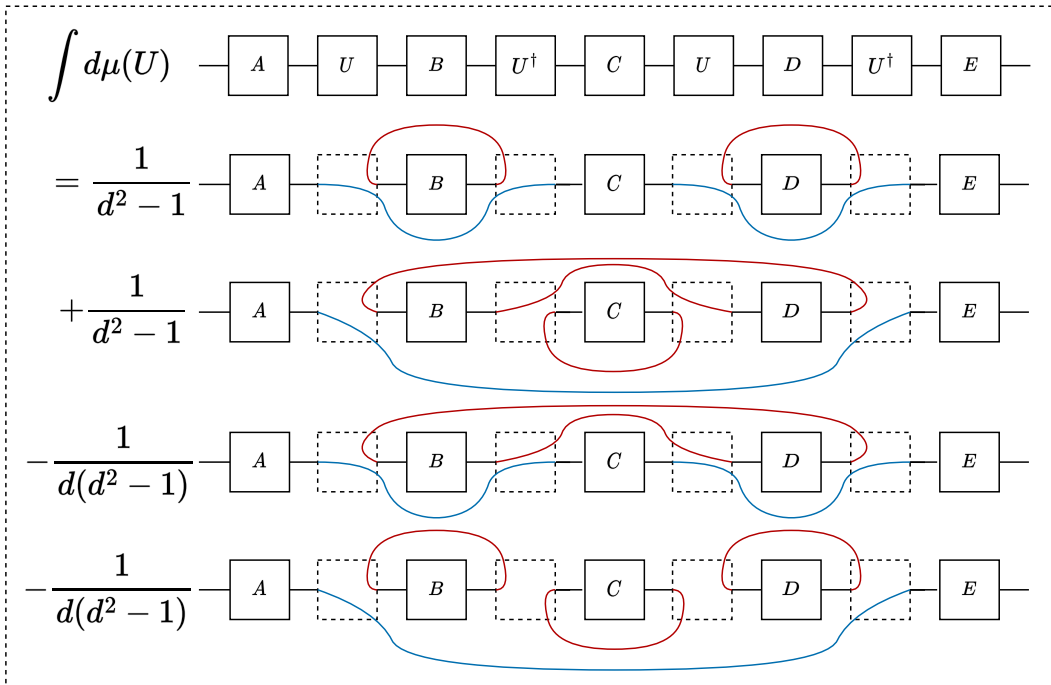


図 4.1.6: 公式 (4.1.8) に対応するテンソルネットワーク

4.2 ユニタリ t -デザイン

ユニタリ t -デザインとは、ユニタリの集合に関する性質の一つであり、次のように定義される。

定義 4.2.1. $P_{t,t}(U)$ 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。

$$\frac{1}{K} \sum_{k=1}^K P_{t,t}(U_k) = \int_{\mathcal{U}(d)} P_{t,t}(U) d\mu_{\text{Haar}}(U) \quad (4.2.1)$$

ここで、 $d\mu_{\text{Haar}}$ はハール分布である。

すなわち、ユニタリ t -デザインとは、ユニタリの集合に関する斉次多項式 $P_{t,t}(U)$ の平均値が、ユニタリ群上のハール分布に関する積分と厳密に等しいような性質のことである。ユニタリ t -デザインの性質を持つユニタリの集合は、ユニタリ群上において一様に分布した要素の集合である。このような性質を持つユニタリの集合を構成することができれば、ある斉次多項式 $P_{t,t}(U)$ のハール分布に関する積分を、有限のユニタリの集合による平均値として計算することができる。また、 $\{U_k\}$ がユニタリ t -デザインであれば、 $\{U_k\}$ はユニタリ $(t-1)$ -デザインでもあることが知られている。

例えば、ユニタリ t -デザインの集合として以下のものがある。

- Pauli 群 $\mathcal{P}(n) = \{e^{i\frac{k\pi}{2}} P_{j_1} \otimes P_{j_2} \otimes \cdots \otimes P_{j_n} \mid k, j_l = 0, 1, 2, 3\}$ はユニタリ 1-デザインである。
- Clifford 群 $\mathcal{C}(n) = \{U \in \mathcal{U}(2^n) \mid U\mathcal{P}(n)U^\dagger = \mathcal{P}(n)\}$ はユニタリ 3-デザインである。

図 4.2.1 は、 2×2 のユニタリ 1-デザインとユニタリ 2-デザインを成す集合を 3次元球内の点として表現したものである [65]。

次に言葉の定義を行う。 n 量子ビットの系を考えると、 $2^n \times 2^n$ のユニタリの集合がユニタリ t -デザインを成すとき、このことを、考えている系に対してグローバルユニタリ t -デザインを成すという。 n より小さい自然数 s に対して、 $2^s \times 2^s$ のユニタリの集合がユニタリ t -デザインを成すとき、このことを、考えている系に対してローカルユニタリ t -デザインを成すという。

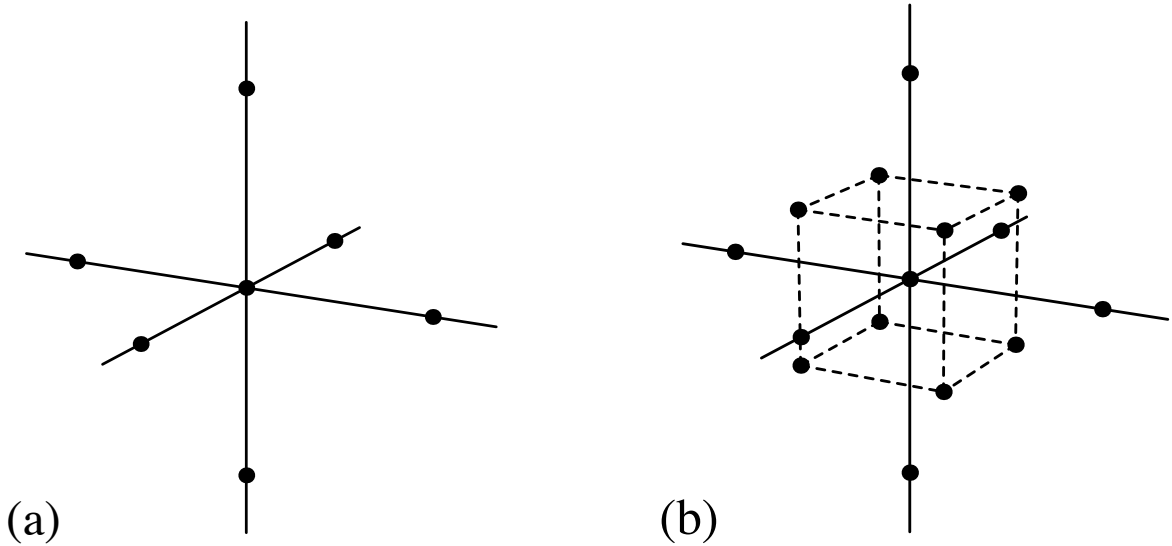


図 4.2.1: $SU(2) \sim SO(3)$ なので、 2×2 のユニタリは位相を除いて 3 次元回転に対応する。更に、単位ベクトル \hat{n} 周りの角度 $\varphi \in [0, \pi]$ の回転をベクトル $\varphi \hat{n} \in \mathbb{R}^3$ と対応させると、 $SO(3)$ は半径 π の球内の点として表現できる。左がユニタリ 1-デザイン、右がユニタリ 2-デザインの点の集合を表している。ただし、球面の反対側の点は同一視される。論文 [65] より引用

4.3 表現力

量子回路によって生成されるユニタリの集合を \mathbb{U} と表す。量子回路の表現力とは、量子回路がどれだけの多くのユニタリを表現できるかという能力（表現能力）を定量化したものであり、次のように定義される。

定義 4.3.1. 量子回路 \mathbb{U} の表現力は、次のように定義される。

$$\epsilon_{\mathbb{U}}^{(t,p)}(X) := \left\| A_{\mathbb{U}}^{(t)}(X) \right\|_p, \quad (4.3.1)$$

$$A_{\mathbb{U}}^{(t)}(X) := \int_{V \in U(d)} d\mu_{\text{Haar}}(V) (VXV^\dagger)^{\otimes t} - \int_{U \in \mathbb{U}} dU (UXU^\dagger)^{\otimes t}, \quad (4.3.2)$$

ここで、 $\|\cdot\|_p$ は Schatten p -ノルムであり、 X は入力状態またはオブザーバブルである。

量子回路の表現力は、量子回路の表現領域 \mathbb{U} と $U(d)$ のそれぞれに依存する量の差として定義される。そのため、量子回路の表現力という値は、多くのユニタリを表現できるほど（表現能力が高いほど）小さくなる。たとえば、量子回路がその次元の全てのユニタリを表現できる場合、表現力の値は 0 となる。一般的には、量子回路の層数が多いほど、すなわち、使用するパラメータ付きゲート数や制御ゲートが多いほど表現能力は高く、表現力は小さくなる。

バレンプラトーという現象は、量子回路がユニタリ 2-デザインを成すことが原因の一つであることをすでに述べた。よって、ここでは $t = 2$ の場合のみを考えることにする。

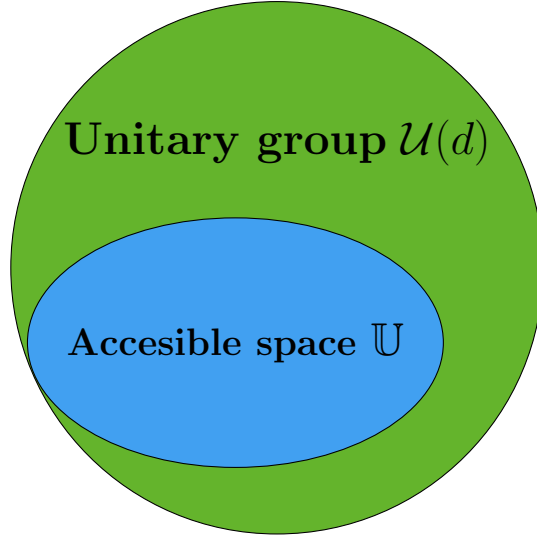


図 4.3.1: ユニタリ群 $\mathcal{U}(d)$ と量子回路の表現領域 \mathbb{U}

4.3.1 $p = 1$

$p = 1$ のとき、 X として $|\mathbf{0}\rangle\langle\mathbf{0}| := |\mathbf{0}\rangle\langle\mathbf{0}|^{\otimes n}$ を用いると、表現力は次のようになる。

$$\begin{aligned} \epsilon_{\mathbb{U}}^{(2,1)}(|\mathbf{0}\rangle\langle\mathbf{0}|) &:= \left\| A^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) \right\|_1 \\ &= \text{Tr} \left| \int_{V \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) (V |\mathbf{0}\rangle\langle\mathbf{0}| V^\dagger)^{\otimes 2} - \int_{U \in \mathbb{U}} dU (U |\mathbf{0}\rangle\langle\mathbf{0}| U^\dagger)^{\otimes 2} \right| \end{aligned} \quad (4.3.3)$$

トレースの第一項は以下のように計算される [66]。

$$\int_{V \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) (V |\mathbf{0}\rangle\langle\mathbf{0}| V^\dagger)^{\otimes 2} = \frac{\mathbf{1}_{d^2} + \text{SWAP}}{d(d+1)}, \quad (4.3.4)$$

ただし、 $\text{SWAP} := \sum_{i,j=1}^d |ij\rangle\langle ji|$ である。

4.3.2 $p = 2$

$p = 2$ のとき、次の定理が成り立つ。

定理 4.3.1. 表現力 $\epsilon_{\mathbb{U}}^{(t,2)}(X) = \|A^{(t)}(X)\|_2$ は以下のようにフレームポテンシャルと呼ばれる量 \mathcal{F} によって表すことができる [26]。

$$\epsilon_{\mathbb{U}}^{(t,2)}(X) := \sqrt{\mathcal{F}_{\mathbb{U}}^{(t)}(X) - \mathcal{F}_{\text{Haar}}^{(t)}(X)} \quad (4.3.5)$$

$$\mathcal{F}_{\mathbb{U}}^{(t)}(X) := \int_{W \in \mathbb{U}} \int_{V \in \mathbb{U}} dW dV \text{Tr} \left[W X W^\dagger (V X V^\dagger)^\dagger \right]^t \quad (4.3.6)$$

$$\begin{aligned} \mathcal{F}_{\text{Haar}}^{(t)}(X) &:= \int_{W \in \mathcal{U}(d)} \int_{V \in \mathcal{U}(d)} d\mu_{\text{Haar}}(W) d\mu_{\text{Haar}}(V) \text{Tr} \left[W X W^\dagger (V X V^\dagger)^\dagger \right]^t \\ &= \int_{V \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) \text{Tr} \left[V X^\dagger V^\dagger X \right]^t \end{aligned} \quad (4.3.7)$$

特に $t = 2$ のとき、 $\mathcal{F}_{\mathcal{U}(d)}^{(2)}(X)$ は、公式 (4.1.5) を用いて以下のように計算できる。

$$\mathcal{F}_{\text{Haar}}^{(2)}(X) = \frac{\text{Tr}[X^4] + \text{Tr}[X^2]^2}{2^{2n} - 1} - \frac{2 \text{Tr}[X^2] \text{Tr}[X]^2}{2^n(2^{2n} - 1)} \quad (4.3.8)$$

証明.

$$\left\{ \epsilon_{\mathbb{U}}^{(t,2)}(X) \right\}^2 := \|A^{(t)}(X)\|_2^2 \quad (4.3.9)$$

$$:= \text{Tr} \left[A^{(t)}(X)^\dagger A^{(t)}(X) \right] \quad (4.3.10)$$

$$\begin{aligned} &= \text{Tr} \left(\int_{V \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) (V X^\dagger V^\dagger)^{\otimes t} - \int_{U_1 \in \mathbb{U}} dU_1 (U_1 X^\dagger U_1^\dagger)^{\otimes t} \right) \\ &\quad \times \left(\int_{W \in \mathcal{U}(d)} d\mu_{\text{Haar}}(W) (W X W^\dagger)^{\otimes t} - \int_{U_2 \in \mathbb{U}} dU_2 (U_2 X U_2^\dagger)^{\otimes t} \right) \end{aligned} \quad (4.3.11)$$

$$\begin{aligned} &\stackrel{(1)}{=} \int_{V \in \mathcal{U}(d)} \int_{W \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) d\mu_{\text{Haar}}(W) \text{Tr} \left[V X^\dagger V^\dagger W X W^\dagger \right]^t \\ &\quad - \int_{V \in \mathcal{U}(d)} \int_{U_2 \in \mathbb{U}} d\mu_{\text{Haar}}(V) dU_2 \text{Tr} \left[V X^\dagger V^\dagger U_2 X U_2^\dagger \right]^t \\ &\quad - \int_{U_1 \in \mathbb{U}} \int_{W \in \mathcal{U}(d)} dU_1 d\mu_{\text{Haar}}(W) \text{Tr} \left[U_1 X^\dagger U_1^\dagger W X W^\dagger \right]^t \\ &\quad + \int_{U_1 \in \mathbb{U}} \int_{U_2 \in \mathbb{U}} dU_1 dU_2 \text{Tr} \left[U_1 X^\dagger U_1^\dagger U_2 X U_2^\dagger \right]^t \end{aligned} \quad (4.3.12)$$

$$\begin{aligned} &\stackrel{(2)}{=} \int_{V \in \mathcal{U}(d)} \int_{W \in \mathcal{U}(d)} d\mu_{\text{Haar}}(V) d\mu_{\text{Haar}}(W) \text{Tr} \left[V X^\dagger V^\dagger X \right]^t \\ &\quad - \int_{V \in \mathcal{U}(d)} \int_{U_2 \in \mathbb{U}} d\mu_{\text{Haar}}(V) dU_2 \text{Tr} \left[V X^\dagger V^\dagger X \right]^t \\ &\quad - \int_{U_1 \in \mathbb{U}} \int_{W \in \mathcal{U}(d)} dU_1 d\mu_{\text{Haar}}(W) \text{Tr} \left[X^\dagger W X W^\dagger \right]^t \\ &\quad + \int_{U_1 \in \mathbb{U}} \int_{U_2 \in \mathbb{U}} dU_1 dU_2 \text{Tr} \left[U_1 X^\dagger U_1^\dagger U_2 X U_2^\dagger \right]^t \end{aligned} \quad (4.3.13)$$

$$\begin{aligned} &\stackrel{(3)}{=} \int_{U_1 \in \mathbb{U}} \int_{U_2 \in \mathbb{U}} dU_1 dU_2 \text{Tr} \left[U_1 X^\dagger U_1^\dagger U_2 X U_2^\dagger \right]^t \\ &\quad - \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V) \text{Tr} \left[V X^\dagger V^\dagger X \right]^t \end{aligned} \quad (4.3.14)$$

$$= \mathcal{F}_{\mathbb{U}}^{(t)}(X) - \mathcal{F}_{\text{Haar}}^{(t)}(X) \quad (4.3.15)$$

等号 (1) では、 $A^{\otimes t} B^{\otimes t} = (AB)^{\otimes t}$ であることと、 $\text{Tr}[C^{\otimes t}] = \text{Tr}[C]^t$ であることを用いた。等式 (2) では、ハール分布の性質（両側不変性）を用いた。等式 (3) では、 $\int_{W \in \mathcal{U}(d)} d\mu_{\text{Haar}}(W) = \int_{U_{1,2} \in \mathcal{U}} dU_{1,2} = 1$ であることを用いた。 ■

系 4.3.2. $t = 2$, $X = |\mathbf{0}\rangle\langle\mathbf{0}|$ のとき、表現力は次のように計算される [26]。

$$(\epsilon_{\mathcal{U}}^{(2,2)}(|\mathbf{0}\rangle\langle\mathbf{0}|))^2 = \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} dW dV |\langle\mathbf{0}| W^\dagger V |\mathbf{0}\rangle|^4 - \frac{1}{2^{n-1}(2^n + 1)} \quad (4.3.16)$$

式 4.3.16 の右辺の第一項をみると、これは対象となる量子回路から生成される 2 つの状態ベクトル $V|\mathbf{0}\rangle$, $W|\mathbf{0}\rangle$ の内積の絶対値の 4 乗を積分（平均）したものであることがわかる。変分量子回路の表現力を数値計算で求める場合は、パラメーターをランダムにサンプリングして 2 つの状態ベクトルを生成し、それらの内積の絶対値の 4 乗を平均すればよい（モンテカルロ積分）。

証明. $X = |\mathbf{0}\rangle\langle\mathbf{0}|$ のとき、 $\mathcal{F}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|)$, $\mathcal{F}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|)$ は以下のように計算される。

$$\mathcal{F}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) := \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} dW dV \text{Tr} \left[W |\mathbf{0}\rangle\langle\mathbf{0}| W^\dagger (V |\mathbf{0}\rangle\langle\mathbf{0}| V^\dagger)^\dagger \right]^2 \quad (4.3.17)$$

$$= \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} dW dV |\langle\mathbf{0}| W^\dagger V |\mathbf{0}\rangle|^4 \quad (4.3.18)$$

$$\mathcal{F}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) := \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} d\mu_{\text{Haar}}(W) d\mu_{\text{Haar}}(V) \text{Tr} \left[W |\mathbf{0}\rangle\langle\mathbf{0}| W^\dagger (V |\mathbf{0}\rangle\langle\mathbf{0}| V^\dagger)^\dagger \right]^2 \quad (4.3.19)$$

$$= \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} d\mu_{\text{Haar}}(W) d\mu_{\text{Haar}}(V) \text{Tr} \left[(V^\dagger W) |\mathbf{0}\rangle\langle\mathbf{0}| (V^\dagger W)^\dagger |\mathbf{0}\rangle\langle\mathbf{0}| \right]^2 \quad (4.3.20)$$

$$\stackrel{(1)}{=} \int_{W \in \mathcal{U}} \int_{V \in \mathcal{U}} d\mu_{\text{Haar}}(W) d\mu_{\text{Haar}}(V) \text{Tr} \left[W |\mathbf{0}\rangle\langle\mathbf{0}| W^\dagger |\mathbf{0}\rangle\langle\mathbf{0}| \right]^2 \quad (4.3.21)$$

$$\stackrel{(2)}{=} \int_{W \in \mathcal{U}} d\mu_{\text{Haar}}(W) \text{Tr} \left[W |\mathbf{0}\rangle\langle\mathbf{0}| W^\dagger |\mathbf{0}\rangle\langle\mathbf{0}| \right]^2 \quad (4.3.22)$$

$$= \int_{W \in \mathcal{U}} d\mu_{\text{Haar}}(W) |\langle\mathbf{0}| W |\mathbf{0}\rangle|^4 \quad (4.3.23)$$

$$= \int_{W \in \mathcal{U}} d\mu_{\text{Haar}}(W) w_{\mathbf{0},\mathbf{0}} w_{\mathbf{0},\mathbf{0}} w_{\mathbf{0},\mathbf{0}}^* w_{\mathbf{0},\mathbf{0}}^* \quad (4.3.24)$$

$$\stackrel{(3)}{=} \frac{1}{2^{n-1}(2^n + 1)} \quad (4.3.25)$$

等号 (1) では、ハール分布の性質（左不変性）を用いた。等号 (2) では、 $\int_{V \in \mathcal{U}} d\mu_{\text{Haar}}(V) = 1$ を用いた。等号 (3) では、公式 (4.1.2) を用いた。 ■

実際に系 4.3.2 を用いて 2 つの具体的な量子回路に対して表現力 $(\epsilon_{\mathcal{U}}^{(2,2)}(|\mathbf{0}\rangle\langle\mathbf{0}|))^2$ を計算してみる。1 つは、Tensor Product Ansatz (TPA) と呼ばれる回路であり、図 4.3.2 の左のような、2 量子ビットゲートを持たない回路である。もう 1 つは、Alternating Layered Ansatz (ALT) と呼ばれる回路であり、図 4.3.2 の右のような、2 量子ビットゲートを交互にずらして配置する量子回路である。

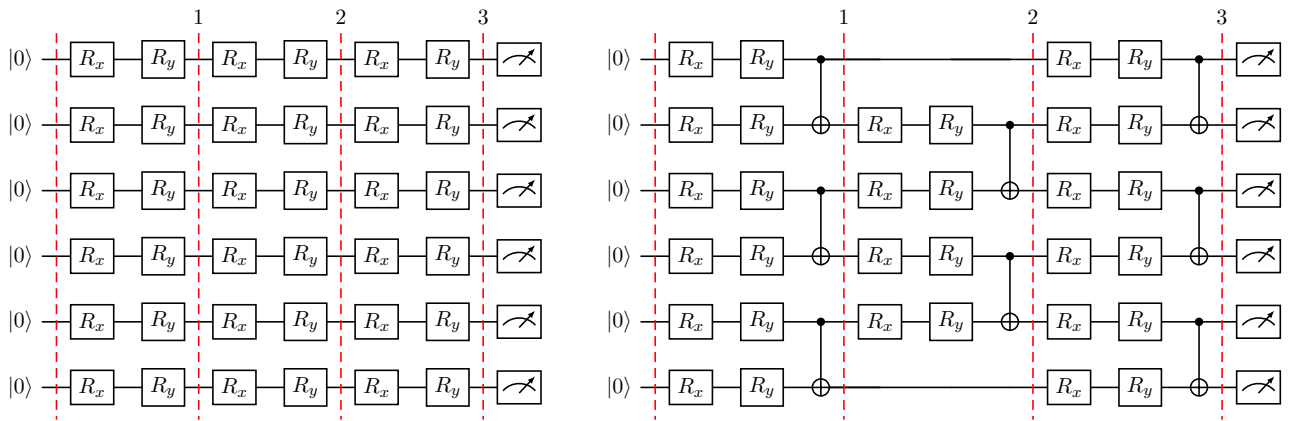


図 4.3.2: 6 量子ビット 3 層の TPA (左) と ALT (右)

TPA の表現力の 2 乗した値を各量子ビット数 n 、各層数 L について計算した結果を図 4.3.3 の左に示す。ALT の表現力の 2 乗した値を各量子ビット数 n 、各層数 L について計算した結果を図 4.3.3 の右に示す。表現力は、全てのパラメーターを 5000 回ランダムにサンプリングして計算した。

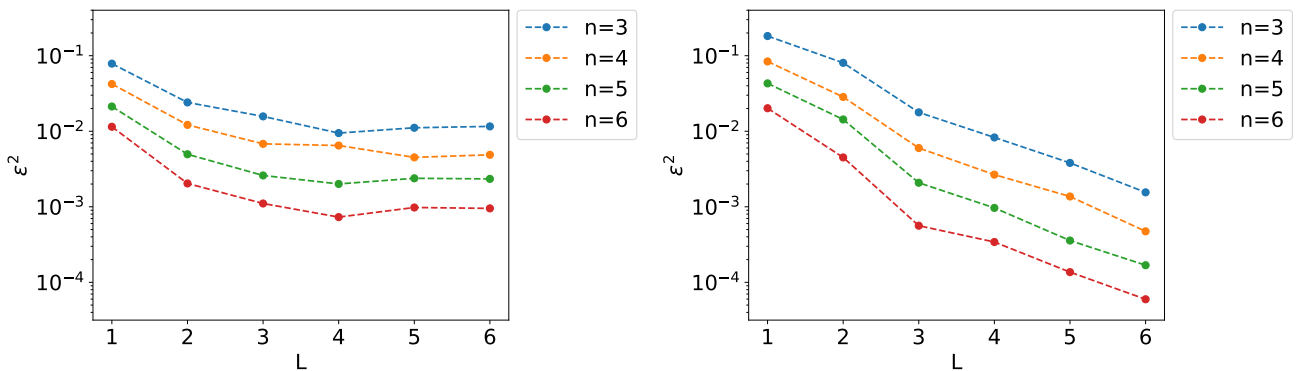


図 4.3.3: TPA (左) と ALT (右) の表現力

TPA はエンタングルメントを持たないため、一定以上の層数になると表現力がほとんど変化していない。一方、ALT の総数を増やすほどエンタングルメントが増え、表現力が小さくなっていることがわかる。

第5章 量子機械学習におけるコスト関数の勾配の分散の上界

量子機械学習におけるバレンプラトーの研究は、それ以外の変分量子アルゴリズムにおけるバレンプラトーの研究に比べて少ない [29, 30]。特に、量子機械学習のデータの inputs がバレンプラトーにどのように影響するかについては、詳しくわかっていない。そこで、この章では、量子機械学習におけるコスト関数の勾配の分散の上界をデータ input の観点から調べた結果を示す。まず、5.1 節では、勾配の分散の上界を、先行研究に基づき新たに導出する。5.2 節では、入力状態がグローバルユニタリ 2-デザインを成す場合の勾配の分散の上界を求める。5.3 節では、勾配の分散の上界を入力回路の表現力によって表す。5.4 節では、入力状態がローカルユニタリ 2-デザインを成すというグローバルユニタリ 2-デザインよりも弱い仮定の下で、勾配の分散の上界を計算する。最後に、5.5 節では、5.4 節の設定に加えて、Depolarizing ノイズがある場合の勾配の分散の上界を計算する。

本論文のバレンプラトーの解析においては、基本的に図 5.0.1 にあるような量子回路学習モデル (3.2.1 節) を用いる。ただし、定理 5.1.2 については Data reuploading モデル (3.2.2 節) についても成り立つ。量子カーネル法 (3.2.3 節) に関するバレンプラトーについては、先行研究 [30] において既に解析されているため、本論文では扱わない。

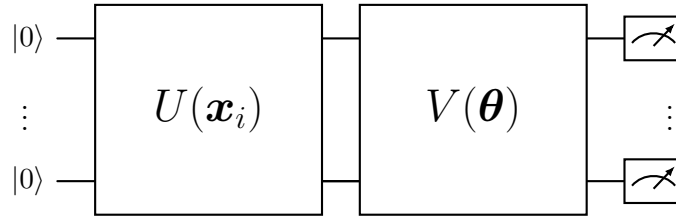


図 5.0.1: 量子回路学習における量子回路の構造

\mathbf{x}_i はデータセット $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ の内、 i 番目の組の input に対応するデータであり、 $\boldsymbol{\theta}$ は学習パラメータである。 $U(\mathbf{x}_i)$ はデータを input するためのユニタリであり、 $V(\boldsymbol{\theta})$ は学習パラメータを input するためのユニタリである。ただし、全てのパラメータ $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots)$ は独立であり、取りうる区間は $[0, 2\pi]$ であると仮定する。

データ input 後の量子状態の密度演算子を $\rho_i = U(\mathbf{x}_i) |0\rangle\langle 0| U^\dagger(\mathbf{x}_i)$ と表し、これを input 状態と呼ぶ。また、学習パラメータ input 後の量子状態の密度演算子を $\rho_i(\boldsymbol{\theta}) = V(\boldsymbol{\theta}) U(\mathbf{x}_i) |0\rangle\langle 0| U^\dagger(\mathbf{x}_i) V^\dagger(\boldsymbol{\theta})$ とする。

以降の解析においては、 $\mathbb{U} := \{U(\mathbf{x}) | \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^m, m \in \mathbb{N}\}$ とする。すなわち、 \mathbb{U} は input データ \mathbf{x} によって生成されるユニタリの集合である。関数 $g(U(\mathbf{x}))$ の input データ $\mathbf{x} \in \mathcal{X}$ に関する積分は、次のように $U \in \mathbb{U}$ に関する積分として表現する。

$$\int_{\mathbb{U}} dU g(U) = \int_{\mathcal{X}} d\mathbf{x} g(U(\mathbf{x})) \quad (5.0.1)$$

5.1 量子機械学習における勾配の分散の一般的な上界

バレンプラトーに関連するほとんどの論文では、コスト関数は式 (5.1.1) のように、あるオブザーバブルの期待値として与えられる (VQE や QAOA などの典型的な変分量子アルゴリズムにおけるコスト関数)。一方、教師ありの量子機械学習においては式 (5.1.2) のように、各訓練データに依存する出力 $l_i(\boldsymbol{\theta})$ それ自体、あるいは更に関数を作用させたものをモデルの予測としてコスト関数が定義される。このようなコスト関数に関するバレンプラトーの先行研究はまだ少ない。

$$l_i(\boldsymbol{\theta}) = \text{Tr}[\rho_i(\boldsymbol{\theta}) O], \quad (5.1.1)$$

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N f(l_i(\boldsymbol{\theta}), y_i) \quad (5.1.2)$$

ただし、 f は $f(x, y) := |x - y|$ や $f(x, y) := (x - y)^2$ 、 $f(x, y) := -y \log x - (1 - y) \log(1 - x)$ などの関数であり、 O は任意のオブザーバブルである。

$l_i(\boldsymbol{\theta})$ の勾配は、パラメーターシフトルール (3.1.3 節) によって求めることができる。一方、 $\mathcal{L}(\boldsymbol{\theta})$ の勾配は、 $\partial_\nu := \partial/\partial\theta_\nu$ として、次のように連鎖律によって求める。

$$\partial_\nu \mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^N \partial_{l_i(\boldsymbol{\theta})} f \cdot \partial_\nu l_i(\boldsymbol{\theta}). \quad (5.1.3)$$

この勾配の計算量が量子ビット数に関してどのようにスケールするかを調べることは、量子機械学習の計算量の評価において重要である。よって、バレンプラトーの定義にもあるように、この勾配の分散 $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})]$ を調べる。まず、分散の公式を用いて式 (5.1.3) を次のように変形してみる。

$$\begin{aligned} \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})] &= \text{Var}_{V(\boldsymbol{\theta})} \left[\frac{1}{N} \sum_{i=1}^N \partial_\nu f(l_i(\boldsymbol{\theta}), y_i) \right] \\ &= \frac{1}{N^2} \sum_{i,j=1}^N \text{Cov}_{V(\boldsymbol{\theta})}[\partial_\nu f(l_i(\boldsymbol{\theta}), y_i), \partial_\nu f(l_j(\boldsymbol{\theta}), y_j)] \end{aligned} \quad (5.1.4)$$

すると、 $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})]$ は $\partial_\nu f(l_i(\boldsymbol{\theta}), y_i)$ の共分散に分解できる。共分散は負の値を取りうるため、 $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})]$ のスケールングが $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})]$ のスケールングと同じとは限らないことがわかる。

$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})]$ の性質に関して、先行研究 [29] により次の結果が得られている。

定理 5.1.1. (論文 [29] の Theorem 1 の結果) 量子機械学習におけるコスト関数 $\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N f(l_i(\boldsymbol{\theta}; y_i), y_i)$ のパラメーター $\theta_\nu \in \boldsymbol{\theta}$ に関する勾配 $\partial_\nu \mathcal{L}(\boldsymbol{\theta}) \equiv \partial \mathcal{L}(\boldsymbol{\theta}) / \partial \theta_\nu$ の分散は次のように上から抑えられる。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})] \leq \left(\frac{1}{N} \sum_{i=1}^N g_i \sqrt{\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})] + (\text{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})])^2} \right)^2 \quad (5.1.5)$$

ここで、 $l_i(\boldsymbol{\theta}) \equiv l_i(\boldsymbol{\theta}; y_i)$ であり、平均と分散はパラメーター $\boldsymbol{\theta}$ に関して取られている。また、 g_i は以下のように定義される。

$$g_i = \sqrt{2} \max_{\ell_i} \left| \frac{\partial f}{\partial \ell_i} \right| \quad (5.1.6)$$

ここで、 $\max_{\ell_i} |\partial f / \partial \ell_i|$ は $f(\ell_i, y_i)$ の偏微分の絶対値の最大値である。

この定理は、 $l_i(\boldsymbol{\theta})$ の勾配の分散の解析を $\mathcal{L}(\boldsymbol{\theta})$ の勾配の分散の解析に適用できることを示した最初の結果であった。例えば、全ての i について $l_i(\boldsymbol{\theta})$ がバレンプラトーになるとすると、

$$\mathbb{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})] = 0, \quad \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})] \in \mathcal{O}(2^{-\alpha n}), \quad 0 < \alpha \text{ for all } i \quad (5.1.7)$$

が成り立つ。よって、全ての i について $g_i \in \mathcal{O}(\text{poly}(n))$ であるならば、上界 (5.1.5) は $\mathcal{O}(2^{-\alpha n})$ となる。したがって、 $\mathcal{L}(\boldsymbol{\theta})$ もバレンプラトーとなる。

本研究では、上界について次の定理 5.1.2 を導いた。上界 (5.1.8) は上界 (5.1.5) よりも緩いが、具体的な回路に対して上界を解析的に計算することができる。

定理 5.1.2. Theorem 5.1.1 と同じ仮定の下、次の不等式が成り立つ。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})] \leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{l_i(\boldsymbol{\theta})} f)^2] \times \int_{U \in \mathcal{U}} dU \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \ell(\boldsymbol{\theta})] \quad (5.1.8)$$

この定理の証明においては、次の補題を用いる。証明は論文 [30] の Lemma 1, 5 を参照されたい。

補題 5.1.3. 分散についての以下の不等式が成り立つ。

$$\text{Var} \left[\sum_i X_i \right] \leq N \sum_i \text{Var}[X_i], \quad (5.1.9)$$

$$\text{Var}[XY] \leq 2 \text{Var}[X] |Y^2|_{\max} + 2(\mathbb{E}[X])^2 \text{Var}[Y]. \quad (5.1.10)$$

証明. まず、 $\mathbb{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})] = 0$ を示す。ここでは、定理 A.2.1 と同じように、 $l_i(\boldsymbol{\theta}) := C(\boldsymbol{\theta}) = \text{Tr}[\rho_i(\boldsymbol{\theta})O]$, $V(\boldsymbol{\theta}) = V_+ V_-$, $V_+ := \prod_{l=k+1}^L U_l(\theta_l) W_l$, $V_- := \prod_{l=1}^k U_l(\theta_l) W_l$ とする。また、証明 A.2 で得られるように、次の等式が成り立つ。

$$\partial_\nu l_i(\boldsymbol{\theta}) = \frac{i}{2} \text{Tr} \left[V_- \rho V_-^\dagger [P_k, V_+^\dagger O V_+] \right] \quad (5.1.11)$$

簡単のため、 $c = \cos(\theta_\nu/2)$, $s = \sin(\theta_\nu/2)$ とする。このとき、 $\partial_\nu l_i(\boldsymbol{\theta})$ は次のように展開できる。

$$\partial_\nu l_i(\boldsymbol{\theta}) = \frac{i}{2} \text{Tr} \left[V_- \rho V_-^\dagger [P_k, V_+^\dagger O V_+] \right] \quad (5.1.12)$$

$$= \frac{i}{2} \text{Tr} \left[(c \mathbf{1} - i s P_k) W_k \left(\prod_{l=1}^{k-1} U_l W_l \right) \rho \left(\prod_{l=1}^{k-1} U_l W_l \right)^\dagger W_k^\dagger (c \mathbf{1} - i s P_k)^\dagger [P_k, V_+^\dagger O V_+] \right] \quad (5.1.13)$$

$$= \frac{i}{2} \left\{ (c^2 - s^2) \text{Tr} \left[[\rho', P_k] V_+^\dagger O V_+ \right] + i c s \text{Tr} \left[[\rho', P_k] \right] \right\} \quad (5.1.14)$$

ただし、 $\rho' = W_k \left(\prod_{l=1}^{k-1} U_l W_l \right) \rho \left(\prod_{l=1}^{k-1} U_l W_l \right)^\dagger W_k^\dagger$ とした。 $\theta_\nu \in [0, 2\pi]$ による平均は、

$$\frac{1}{2\pi} \int_0^{2\pi} d\theta_\nu (c^2 - s^2) = 0, \quad \frac{1}{2\pi} \int_0^{2\pi} d\theta_\nu c s = 0 \quad (5.1.15)$$

である。よって、 $\mathbb{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta})] = 0$ である。

次に $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})]$ の上界を導く。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})] = \text{Var}_{V(\boldsymbol{\theta})} \left[\frac{1}{N} \sum_{i=1}^N \partial_\nu f(l_i(\boldsymbol{\theta}), y_i) \right] \quad (5.1.16)$$

$$\stackrel{(1)}{\leq} \frac{1}{N} \sum_{i=1}^N \text{Var}_{V(\theta)}[\partial_\nu f(\ell_i(\theta), y_i)] \quad (5.1.17)$$

$$= \frac{1}{N} \sum_{i=1}^N \text{Var}_{V(\theta)}[\partial_{\ell_i(\theta)} f \cdot \partial_\nu \ell_i(\theta)] \quad (5.1.18)$$

$$\stackrel{(2)}{\leq} \frac{1}{N} \sum_{i=1}^N 2 \max_{\theta} [(\partial_{\ell_i(\theta)} f)^2] \cdot \text{Var}_{V(\theta)}[\partial_\nu \ell_i(\theta)] \quad (5.1.19)$$

$$\leq 2 \max_{i, \theta} [(\partial_{\ell_i(\theta)} f)^2] \times \frac{1}{N} \sum_{i=1}^N \text{Var}_{V(\theta)}[\partial_\nu \ell_i(\theta)] \quad (5.1.20)$$

$$=: 2 \max_{i, \theta} [(\partial_{\ell_i(\theta)} f)^2] \times \int_{U \in \mathcal{U}} dU \text{Var}_{V(\theta)}[\partial_\nu \ell(\theta)] \quad (5.1.21)$$

ただし、不等式(1)では、不等式(5.1.9)を用いた。不等式(2)では、不等式(5.1.10)と $E_{V(\theta)}[\partial_\nu \ell_i(\theta)] = 0$ を用いた。 ■

この結果から、コスト関数の勾配の分散の上界を求めるためには、 $\int_{U \in \mathcal{U}} dU \text{Var}_{V(\theta)}[\partial_\nu \ell(\theta)]$ を計算する必要があるとわかる。具体的なスケーリングを求めるため、以降の解析においては、図 5.1.1 の構造を仮定する。すなわち、初めは入力回路には条件を課さず、学習回路のみ ξ 個の s 量子ビットユニタリのテンソル積からなる回路であるとする。

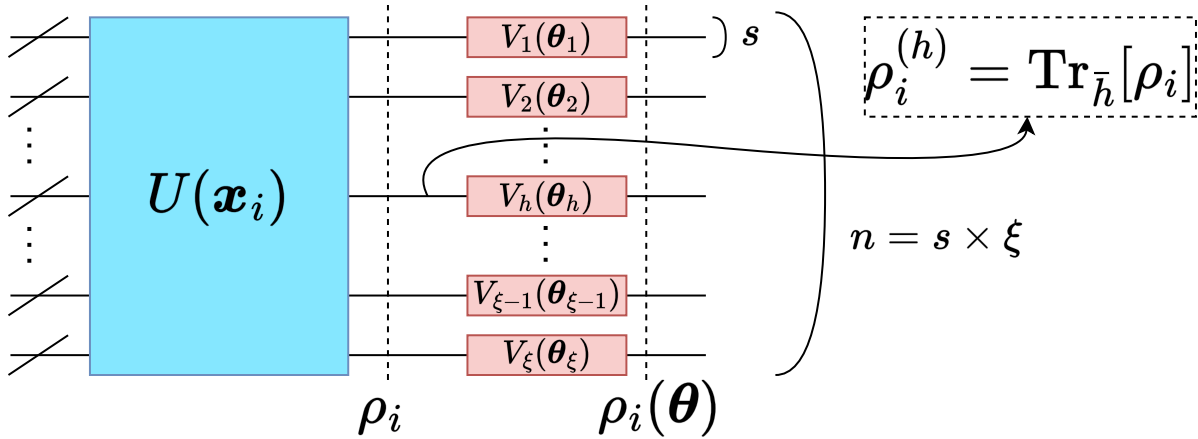


図 5.1.1: 解析する量子回路の構造：学習回路は s 量子ビットユニタリのテンソル積からなる

この学習回路の各 s 量子ビットユニタリがそれぞれユニタリ 2-デザインを成す、すなわち、ローカルユニタリ 2-デザインを成すと仮定する。このとき、次の命題が成り立つことが先行研究 [27] により示されている。

命題 5.1.4. (論文 [27] の Supplementary Note 5, B の結果) $V(\boldsymbol{\theta})$ が、 s 量子ビットユニタリのテンソル積からなり、それぞれがローカルユニタリ 2-デザインを成すと仮定する。コスト関数 $\ell_i(\boldsymbol{\theta}) = \text{Tr}[\rho_i(\boldsymbol{\theta})O_L]$, $O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbf{1}_{\bar{j}}$ について、 h 番目のゲートブロック中のパラメーター θ_ν に関する偏微分を考える。このとき、次の等式が成り立つ。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \ell_i(\boldsymbol{\theta})] = r_{n,s} D_{\text{HS}}(\rho_i^{(h)}, \mathbf{1}/2^s) \quad (5.1.22)$$

ここで、 $r_{n,s} := \frac{s \cdot 2^{3(s-1)}}{n^2(2^{2s}-1)^2} \in \Omega(1/\text{poly}(n))$ であり、 $D_{\text{HS}}(A, B) = \text{Tr}[(A - B)^2]$ は Hilbert-Schmidt 距離である。また、 $\mathbf{1}/2^s$ は s 量子ビット上の最大混合状態であり、 $\rho_i^{(h)} = \text{Tr}_{\bar{h}}[U(\mathbf{x}_i) |\mathbf{0}\rangle\langle \mathbf{0}| U^\dagger(\mathbf{x}_i)]$ は h 番目のユニタリが作用する s 量子ビット上の入力状態の縮約密度演算子である。

この命題において、学習回路とオブザーバブルの情報は $r_{n,s}$ に含まれるが、 s は n によらず、オブザーバブルは局所的なものであるため、これらはバレンプラトーの原因とならない。一方で、この命題は、入力状態の縮約密度演算子 $\rho_i^{(h)}$ が最大混合状態に近いほど、 $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \ell_i(\boldsymbol{\theta})]$ が小さくなることを意味する。つまり、式 (6.2.12) は、データセットと入力回路の構造の情報を含む ρ_i がバレンプラトーを引き起こす可能性があることを意味する。

式 (6.2.12) を式 (5.1.8) に代入すると、次の上界を得る。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}(\boldsymbol{\theta})] \leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i(\boldsymbol{\theta})} f)^2] \times r_{n,s} \times \int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s). \quad (5.1.23)$$

よって、 $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s)$ が上界のスケーリングを決めることがわかる。

5.2 グローバルユニタリ 2-デザインの入力状態とバレンプラトー

バレンプラトーの解析においては、しばしば、学習回路がグローバルユニタリ 2-デザインを成すと仮定する。そこで、 $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s)$ を計算する最初の簡単なモデルとして、入力状態の集合がグローバルユニタリ 2-デザインを成すと仮定する。 $D_{\text{HS}}(\rho, \mathbf{1}/2^s) = \text{Tr}[\rho^2] - 1/2^s$ であることから、

$$D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s) = \text{Tr}[(\text{Tr}_{\bar{h}}[U |\mathbf{0}\rangle\langle \mathbf{0}| U^\dagger])^2] - 1/2^s \quad (5.2.1)$$

であり、最初の項は U, U^\dagger の成分に関してそれぞれ 2 次の多項式である。よって、ユニタリ 2-デザインの定義から次のように $\mathbb{U} = \mathbb{U}(d)$, $dU = d\mu_{\text{Haar}}(U)$ として積分してよい。

$$\int_{U \in \mathbb{U}(d)} d\mu_{\text{Haar}}(U) D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s) = \int_{U \in \mathbb{U}(d)} d\mu_{\text{Haar}}(U) \text{Tr}[(\text{Tr}_{\bar{h}}[U |\mathbf{0}\rangle\langle \mathbf{0}| U^\dagger])^2] - 1/2^s \quad (5.2.2)$$

となる。右辺の第一項の積分計算は、次の定理により与えられる。

定理 5.2.1. \mathcal{H} を $\mathcal{H} = \mathcal{H}_h \otimes \mathcal{H}_{\bar{h}}$ のように二つの系からなる n 量子ビットの Hilbert 空間とする。ここで、 \mathcal{H}_h は s 量子ビットからなる Hilbert 空間であり、 $\mathcal{H}_{\bar{h}}$ は残りの $n - s$ 量子ビットからなる Hilbert 空間である。このとき、次の式が成り立つ。

$$\int_{U \in \mathbb{U}(d)} \text{Tr}_h[(\text{Tr}_{\bar{h}}[U |\mathbf{0}\rangle\langle \mathbf{0}| U^\dagger])^2] dU = \frac{2^{n-s} + 2^s}{2^n + 1}. \quad (5.2.3)$$

証明. $a_{ij} = U_{i1}U_{j1}^*$, $t = 2^{n-s}$ とすると、

$$\text{Tr}_{\bar{h}}[U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger] = \begin{bmatrix} a_{1,1} + a_{2,2} + \cdots + a_{t,t} & a_{1,t+1} + a_{2,t+2} + \cdots + a_{t,2t} & \cdots & a_{1,2^{n-t+1}} + a_{2,2^{n-t+2}} + \cdots + a_{t,2^n} \\ a_{t+1,1} + a_{t+2,2} + \cdots + a_{2t,t} & a_{t+1,t+1} + a_{t+2,t+2} + \cdots + a_{2t,2t} & \cdots & a_{t+1,2^{n-t+1}} + a_{t+2,2^{n-t+2}} + \cdots + a_{2t,2^n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{2^{n-t+1},1} + \cdots + a_{2^n,t} & a_{2^{n-t+1},t+1} + \cdots + a_{2^n,2t} & \cdots & a_{2^{n-t+1},2^{n-t+1}} + \cdots + a_{2^n,2^n} \end{bmatrix} \quad (5.2.4)$$

よって、上の行列の二乗 $(\text{Tr}_{\bar{h}}[U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger])^2$ の (1,1) 成分は以下ようになる。

$$(a_{1,1} + a_{2,2} + \cdots + a_{t,t})^2 + (a_{1,t+1} + a_{2,t+2} + \cdots + a_{t,2t})(a_{t+1,1} + a_{t+2,2} + \cdots + a_{2t,t}) \\ \cdots + (a_{1,2^{n-t+1}} + \cdots + a_{t,2^n})(a_{2^{n-t+1},1} + \cdots + a_{2^n,t}) \quad (5.2.5)$$

ユニタリ 2-デザイン の定義から、ハール分布の積分公式 (4.1.3) を用いることができ、

$$\int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) a_{ij}a_{i'j'} = \frac{1}{d(d+1)}(\delta_{ij}\delta_{i'j'} + \delta_{ij'}\delta_{i'j}) \quad (5.2.6)$$

が成り立つ。よって、 $a_{p,p}a_{p,p}$, $a_{p,p}a_{q,q}$, $a_{p,q}a_{q,p}$ ($p \neq q$) のみが生き残る。他の対角成分も同様であるから、求める式は以下ようになる。

$$\int_{U \in \mathcal{U}(d)} \text{Tr}[(\text{Tr}_{\bar{h}}[U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger])^2] d\mu_{\text{Haar}}(U) \\ = \int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) \left(\left[\sum_{i=1}^t a_{ii}^2 \right] + \left[2 \sum_{i < j} a_{ii}a_{jj} \right] + \underbrace{\left[\sum_{i=1}^t a_{t+i,i}a_{i,t+i} \right] + \cdots + \left[\sum_{i=1}^t a_{2^{n-t+i},i}a_{i,2^{n-t+i}} \right]}_{2^s - 1 \text{ 個}} \right) \times 2^s \quad (5.2.7)$$

$$= \int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) \left(\left[\sum_{i=1}^t a_{ii}^2 \right] + \left[2 \sum_{i < j} a_{ii}a_{jj} \right] + \left[\sum_{i=1}^t a_{t+i,i}a_{i,t+i} \right] \times (2^s - 1) \right) \times 2^s \quad (5.2.8)$$

$$= \left(t \times \frac{2}{d(d+1)} + 2 \times {}_t C_2 \times \frac{1}{d(d+1)} + t \times \frac{1}{d(d+1)} \times (2^s - 1) \right) \times 2^s \quad (5.2.9)$$

$$= \frac{2^{n-s} + 2^s}{2^n + 1} \quad (5.2.10)$$

■

この定理より、

$$\int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s) = \int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) \text{Tr}[(\text{Tr}_{\bar{h}}[U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger])^2] - 1/2^s \\ = \frac{2^{n-s} + 2^s}{2^n + 1} - 1/2^s \\ = \frac{2^s - 2^{-s}}{2^n + 1} \in [0, 1 - 1/2^s] \quad (5.2.11)$$

$$=: \mathcal{D}_{\text{Haar}} \quad (5.2.12)$$

したがって、入力状態の集合がグローバルユニタリ 2-デザインを成す場合の上界は次で与えられる。

$$\text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}(\theta)] \leq 2 \max_{i, \theta} [(\partial_{\ell_i}(\theta) f)^2] \times r_{n,s} \times \frac{2^s - 2^{-s}}{2^n + 1} \in \mathcal{O}\left(\frac{1}{2^n}\right) \quad (5.2.13)$$

以上より、量子回路の構造は図 5.1.1 で与えられ、学習回路やオブザーバブルがパレンプラトールを引き起こさないとしても、入力状態の集合がグローバルユニタリ 2-デザインを成す場合は、コスト関数はパレンプラトールとなることが分かった。

5.3 入力回路の表現力による上界

この節では、上界 (5.1.23) のうち $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s)$ の項が、前節で得た結果

$$\mathcal{D}_{\text{Haar}} = \frac{2^s - 2^{-s}}{2^n + 1}$$

と入力回路の表現力によって、次のように上から抑えられることを示す。

$$\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s) \leq \mathcal{D}_{\text{Haar}} + \epsilon_{\mathbb{U}}^{(2,1)}(|0\rangle\langle 0|)$$

この不等式により、入力回路の表現能力が高いほど、上界がきつくなることが定性的に理解される。

まず、その証明に必要な 2 つの補題を示す。

補題 5.3.1. 密度演算子 $\rho_A \in \mathcal{S}(\mathcal{H}_A)$, $\rho_B \in \mathcal{S}(\mathcal{H}_B)$ ($\dim \mathcal{H}_A = \dim \mathcal{H}_B$) に対して、次の等式が成り立つ。

$$\text{Tr}_A[\text{SWAP}(\rho_A \otimes \rho_B)] = \rho_A \rho_B \quad (5.3.1)$$

ただし、 $\text{SWAP} := \sum_{i,j=1}^d |i\rangle \otimes |j\rangle \langle j| \otimes \langle i|$, e.g. $\text{SWAP} |x_i\rangle \otimes |y_j\rangle = |y_j\rangle \otimes |x_i\rangle$ である。

証明. ρ_A と ρ_B は、それぞれ $\rho_A = \sum_{i=1}^d \lambda_i |x_i\rangle \langle x_i|$ と $\rho_B = \sum_{i=1}^d \mu_i |y_i\rangle \langle y_i|$ のように固有分解されると仮定する。これを $\text{Tr}_A[\text{SWAP}(\rho_A \otimes \rho_B)]$ に代入すると、

$$\begin{aligned} \text{Tr}_A[\text{SWAP}(\rho_A \otimes \rho_B)] &= \text{Tr}_A \left[\text{SWAP} \left(\sum_{i,j=1}^d \lambda_i \mu_j (|x_i\rangle \otimes |y_j\rangle) (\langle x_i| \otimes \langle y_j|) \right) \right] \\ &= \text{Tr}_A \left[\sum_{i,j=1}^d \lambda_i \mu_j (|y_j\rangle \otimes |x_i\rangle) (\langle x_i| \otimes \langle y_j|) \right] \\ &= \sum_{i,j=1}^d \lambda_i \mu_j \text{Tr}_A[|y_j\rangle \langle x_i| |x_i\rangle \langle y_j|] \\ &= \sum_{i,j=1}^d \lambda_i \mu_j \langle x_i | y_j \rangle |x_i\rangle \langle y_j| \\ &= \rho_A \rho_B \end{aligned}$$

■

補題 5.3.2. 密度演算子 $\rho_{C1} \in \mathcal{S}(\mathcal{H}_{C1} = \mathcal{H}_{A1} \otimes \mathcal{H}_{B1})$, $\rho_{C2} \in \mathcal{S}(\mathcal{H}_{C2} = \mathcal{H}_{A2} \otimes \mathcal{H}_{B2})$ ($\dim \mathcal{H}_{A1} = \dim \mathcal{H}_{A2}$, $\dim \mathcal{H}_{B1} = \dim \mathcal{H}_{B2}$) に対して、次の等式が成り立つ。

$$\mathrm{Tr}[\mathrm{Tr}_{B1} \rho_{C1} \times \mathrm{Tr}_{B2} \rho_{C2}] = \mathrm{Tr}[(\mathrm{SWAP}_{A1,A2} \otimes \mathbf{1}_{B1,B2})(\rho_{C1} \otimes \rho_{C2})] \quad (5.3.2)$$

証明.

$$\begin{aligned} \mathrm{Tr}_{B1} \rho_{C1} \times \mathrm{Tr}_{B2} \rho_{C2} &\stackrel{(1)}{=} \mathrm{Tr}_{A1}[\mathrm{SWAP}_{A1,A2} (\mathrm{Tr}_{B1} \rho_{C1} \otimes \mathrm{Tr}_{B2} \rho_{C2})] \\ &= \mathrm{Tr}_{A1}[\mathrm{SWAP}_{A1,A2} \mathrm{Tr}_{B1,B2}[\rho_{C1} \otimes \rho_{C2}]] \\ &\stackrel{(2)}{=} \mathrm{Tr}_{A1,B1,B2}[(\mathrm{SWAP}_{A1,A2} \otimes \mathbf{1}_{B1,B2})(\rho_{C1} \otimes \rho_{C2})] \\ &= \mathrm{Tr}_{C1,B2}[(\mathrm{SWAP}_{A1,A2} \otimes \mathbf{1}_{B1,B2})(\rho_{C1} \otimes \rho_{C2})]. \end{aligned}$$

等式 (1) では、補題 5.3.1 を用いた。等式 (2) では、任意の $V, W \in \mathcal{L}(\mathcal{H}_{AB})$ に対して、 $\mathrm{Tr}_A[V \mathrm{Tr}_B[W]] = \mathrm{Tr}_{AB}[(V \otimes \mathbf{1}_B)W]$ が成り立つことを用いた。最後に、残りの系 A_2 についてトレースを取ればよい。 ■

これら 2 つの補題を用いて、次の定理を示す。

定理 5.3.3. 図 5.1.1 の量子回路を用い、入力状態には何も仮定せず、学習回路の s 量子ビットユニタリはそれぞれローカルユニタリ 2-デザインを成すと仮定する。このとき、 $\int_{U \in \mathcal{U}} dU D_{\mathrm{HS}}(\rho^{(h)}, \mathbf{1}/2^s)$ は $\mathcal{D}_{\mathrm{Haar}} := \frac{2^s - 2^{-s}}{2^n + 1}$ と入力回路の表現力 $\epsilon_{\mathcal{U}}^{(t,p)}(\cdot)$ によって次のように上から抑えられる。

$$\int_{U \in \mathcal{U}} dU D_{\mathrm{HS}}(\rho^{(h)}, \mathbf{1}/2^s) \leq \mathcal{D}_{\mathrm{Haar}} + \epsilon_{\mathcal{U}}^{(2,1)}(|\mathbf{0}\rangle\langle\mathbf{0}|) \quad (5.3.3)$$

$$\leq \mathcal{D}_{\mathrm{Haar}} + 2^n \epsilon_{\mathcal{U}}^{(2,2)}(|\mathbf{0}\rangle\langle\mathbf{0}|). \quad (5.3.4)$$

証明. $\mathcal{A}_{\mathcal{U}}^{(t)}(X)$ の定義 4.3.1 において、 $t = 2$, $X = |\mathbf{0}\rangle\langle\mathbf{0}|$ とすると、

$$\int_{U \in \mathcal{U}} dU (U |\mathbf{0}\rangle\langle\mathbf{0}| U^\dagger)^{\otimes 2} = \int_{U \in \mathcal{U}(d)} d\mu_{\mathrm{Haar}}(U) (U |\mathbf{0}\rangle\langle\mathbf{0}| U^\dagger)^{\otimes 2} - \mathcal{A}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) \quad (5.3.5)$$

であるから、 $\rho = U |\mathbf{0}\rangle\langle\mathbf{0}| U^\dagger$ として、両辺に SWAP_{h_1, h_2} を作用させてトレースを取ると

$$\begin{aligned} \int_{U \in \mathcal{U}} dU \mathrm{Tr}[(\rho \otimes \rho) \mathrm{SWAP}_{h_1, h_2}] &= \int_{U \in \mathcal{U}(d)} d\mu_{\mathrm{Haar}}(U) \mathrm{Tr}[(\rho \otimes \rho) \mathrm{SWAP}_{h_1, h_2}] \\ &\quad - \mathrm{Tr}[\mathcal{A}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) \mathrm{SWAP}_{h_1, h_2}] \end{aligned} \quad (5.3.6)$$

補題 5.3.2 より等式 $\mathrm{Tr}[(\rho \otimes \rho) \mathrm{SWAP}_{h_1, h_2}] = \mathrm{Tr}[(\rho^{(h)})^2]$ が成り立つから

$$\int_{U \in \mathcal{U}} dU \mathrm{Tr}[(\rho^{(h)})^2] = \int_{U \in \mathcal{U}(d)} d\mu_{\mathrm{Haar}}(U) \mathrm{Tr}[(\rho^{(h)})^2] - \mathrm{Tr}[\mathcal{A}_{\mathcal{U}}^{(2)}(|\mathbf{0}\rangle\langle\mathbf{0}|) \mathrm{SWAP}_{h_1, h_2}] \quad (5.3.7)$$

したがって、

$$\left| \int_{U \in \mathcal{U}} dU \mathrm{Tr}[(\rho^{(h)})^2] - \frac{1}{2^s} \right| = \left| \int_{U \in \mathcal{U}(d)} d\mu_{\mathrm{Haar}}(U) \mathrm{Tr}[(\rho^{(h)})^2] - \frac{1}{2^s} - \mathrm{Tr}[\mathcal{A}_{\mathcal{U}}^{(2)}(\rho) \mathrm{SWAP}_{h_1, h_2}] \right| \quad (5.3.8)$$

$$\stackrel{(1)}{\leq} \left| \int_{U \in \mathcal{U}(d)} d\mu_{\text{Haar}}(U) \text{Tr} [(\rho^{(h)})^2] - \frac{1}{2^s} \right| + \left| \text{Tr} [\mathcal{A}_{\mathbb{U}}^{(2)}(\rho) \text{SWAP}_{h_1, h_2}] \right| \quad (5.3.9)$$

$$\stackrel{(2)}{\leq} \mathcal{D}_{\text{Haar}} + \|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_1 \|\text{SWAP}_{h_1, h_2}\|_{\infty} \quad (5.3.10)$$

$$\stackrel{(3)}{=} \mathcal{D}_{\text{Haar}} + \|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_1 \quad (5.3.11)$$

$$\stackrel{(4)}{\leq} \mathcal{D}_{\text{Haar}} + 2^n \|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_2 \quad (5.3.12)$$

不等式 (1) では、三角不等式を用いた。不等式 (2) では、Schatten ノルムの性質 $\|AB\|_1 \leq \|A\|_1 \|B\|_{\infty}$ を用いた (2.7 節)。不等式 (3) では、 $\|\text{SWAP}_{h_1, h_2}\|_{\infty} = 1$ であることを用いた。不等式 (4) では、Schatten ノルムの性質 $\|AB\|_1 \leq \|A\|_2 \|B\|_2$ から $\|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_1 \leq 2^n \|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_2$ であることを用いた。 ■

定理 5.3.3 より、一般的な入力回路に対する式 (5.1.23) の上界は、入力回路の表現力によって次のように上から抑えられる。

$$\text{Var}_{V(\boldsymbol{\theta})}[\partial_{\nu} L(\boldsymbol{\theta})] \leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i}(\boldsymbol{\theta}) f)^2] \times r_{n,s} \times \int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s) \quad (5.3.13)$$

$$\leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i}(\boldsymbol{\theta}) f)^2] \times r_{n,s} \times [\mathcal{D}_{\text{Haar}} + \epsilon_{\mathbb{U}}^{(2,1)}(|\mathbf{0}\rangle\langle\mathbf{0}|)] \quad (5.3.14)$$

$$\leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i}(\boldsymbol{\theta}) f)^2] \times r_{n,s} \times [\mathcal{D}_{\text{Haar}} + 2^n \epsilon_{\mathbb{U}}^{(2,2)}(|\mathbf{0}\rangle\langle\mathbf{0}|)] \quad (5.3.15)$$

不等式 (2) はタイトな上界であり、 \mathbb{U} がユニタリ 2-デザインを成すとき、 $\epsilon_{\mathbb{U}}^{(2,1)}(|\mathbf{0}\rangle\langle\mathbf{0}|) = \|\mathcal{A}_{\mathbb{U}}^{(2)}(\rho)\|_1 = 0$ であるから、等号 $\int_{U \in \mathbb{U}} dU \text{Tr} [(\rho^{(h)})^2] - 1/2^s = \mathcal{D}_{\text{Haar}}$ が成り立つ。すなわち、前節の結果 (5.2.13) と同じ上界が得られる。以上のことから、入力回路の表現能力が高いほど、コスト関数の勾配の分散の上界がきつくなることが解析的に示された。

ここでの結果は、先行研究の結果 [26] と類似している。しかし、この先行研究においては、学習回路の表現力とコスト関数の勾配の分散の関係を示しており、本研究はデータ入力について考慮している点で異なる。

5.4 ローカルユニタリ 2-デザインの入力状態とバレンプラトー

5.1 節では、コスト関数の勾配の分散の上界が、 $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$ を用いて上から抑えられることを見た。5.2 節では、入力状態がグローバルユニタリ 2-デザインを成すと仮定して、 $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$ を計算した。しかし、グローバルユニタリ 2-デザインの仮定は、入力回路が十分深い場合であり、実用的な場合では成り立たない可能性がある。そこでこの節では、入力回路の局所的な量子ゲートブロックがそれぞれユニタリ 2-デザインを成す、すなわち、ローカルユニタリ 2-デザインというグローバルユニタリ 2-デザインよりも弱い仮定の下で $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$ を計算する。具体的には、入力回路に図 5.4.1 のような Alternating Layered Ansatz (ALT) を用いる場合を考える。

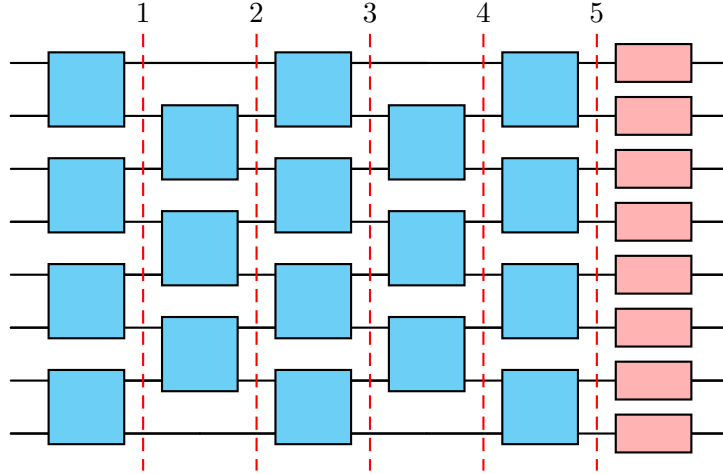


図 5.4.1: 青いゲートブロックは入力回路の Alternating Layered Ansatz (8 量子ビット、5 層の場合) を表す。赤いゲートブロックは学習回路の Tensor Product Ansatz を表す。

5.1 節で見たように、 $\text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}(\theta)]$ の上界は次のように与えられる。

$$\text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}(\theta)] \leq 2 \max_{i, \theta} [(\partial_{\ell_i}(\theta) f)^2] \times r_{n,s} \times \int_{U \in \mathcal{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$$

学習回路とオブザーバブルは、命題 5.1.4 の仮定と同じであるとする。すなわち、学習回路は図 5.4.1 の赤いゲートブロックの構造を持ち、各 $s = 1$ 量子ビット上のゲートブロックはユニタリ 2-デザインを成すとする。また、コスト関数は次で与えられるとする。

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i=1}^N (\ell_i(\theta) - y_i)^2, \quad \ell_i(\theta) = \text{Tr}[\rho_i(\theta) O_L], \quad O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbf{1}_j$$

ただし、 $y_i \in \{0, 1\}$, $\ell_i(\theta) \in [0, 1]$ である。このとき、 $2 \max_{i, \theta} [(\partial_{\ell_i}(\theta) f)^2] = 8$, $r_{n,s} = \frac{s \cdot 2^{3(s-1)}}{n^2(2^{2s}-1)^2}$ となる。残る項 $\int_{U \in \mathcal{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$ はデータセットと入力回路の構造に依存するため、これ評価することで、上界のスケールリングが分かる。

$\int_{U \in \mathcal{U}} dU D_{\text{HS}}(\rho^{(h)}, 1/2^s)$ を入力回路がローカルユニタリ 2-デザインを成すという仮定の下で、各量子ビット数 n 、入力回路の各層数 L に対して、上界を計算しプロットしたものが図 5.4.5 の右である¹。破線は Alternating Layered Ansatz の回路を十分深くした場合の収束先、つまり、5.2 節で得たグローバルユニタリ 2-デザインの場合の結果 $D_{\text{Haar}} = \frac{2^s - 2^{-s}}{2^{n+1}}$ である (図 5.4.5 の左における破線もこれと同じである)。

さらに、同様の設定の下、量子回路シミュレーション (測定回数 ∞) によって、アヤメのデータセットの分類問題²におけるコスト関数の勾配の分散を計算した。その結果が、図 5.4.5 の左である。図 5.4.5 の左右を比較すると、確かに右のプロットは上界として機能していることがわかる。

ただし、Alternating Layered Ansatz のゲートブロックには図 5.4.2 の構造を用いた。また、学習回路は図 5.4.3 の Tensor Product Ansatz を使い、深さを 10 層とした。10 層は各量子ビットがユニタリ

¹ここでは Random Tensor Network Integrator (RTNI) [67] を用いて、シンボリックな計算を行った。具体的な計算方法については、補足の B.1 節に示す。

²用いたアヤメのデータセットは 4 つの特徴量と 2 つのラベルからなり、それぞれのラベルに対応するサンプル数は 50 である。

2-デザインを成すのに十分な層数である。これは、図 5.4.4 において、1 量子ビットで 10 層の Tensor Product Ansatz のフレームポテンシャルが、ユニタリ 2-デザインのフレームポテンシャルに十分に収束していることからわかる（サンプル数は 2×10^4 ）。よって、10 層あれば、各量子ビットがローカルユニタリ 2-デザインを成すとみなせる。

また、図 5.4.5 の左の勾配の分散の計算のために、全てのパラメーターを 10 回ランダムにサンプリングした。そして、プロットしたのは、学習回路の各パラメーターに関する勾配の分散 $\text{Var}[\partial_{\theta_i} \mathcal{L}(\theta)]$ を i について平均したものである。

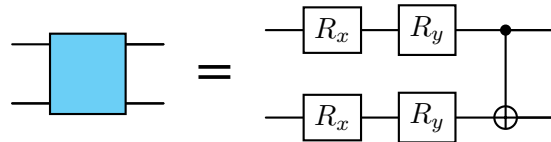


図 5.4.2: ALT のゲートブロックとして用いた構造

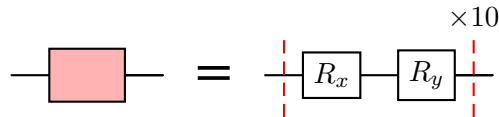


図 5.4.3: TPA のゲートブロックとして用いた構造

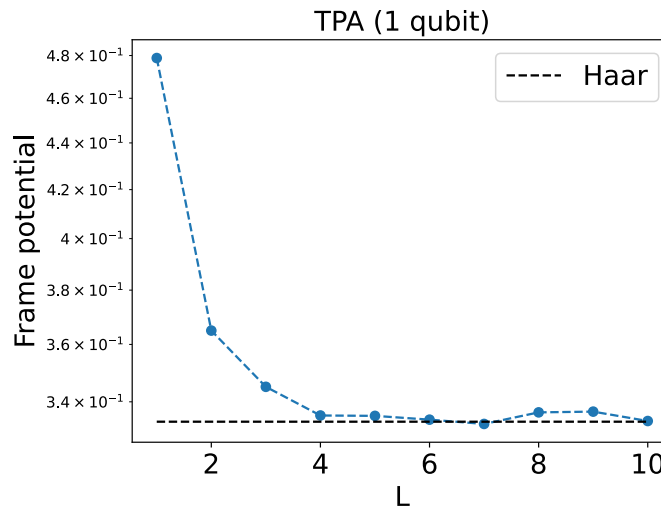


図 5.4.4: 1 量子ビットの Tensor Product Ansatz のフレームポテンシャル。黒の破線は量子回路がユニタリ 2-デザインを成す場合のフレームポテンシャル ($= 1/3$) である。

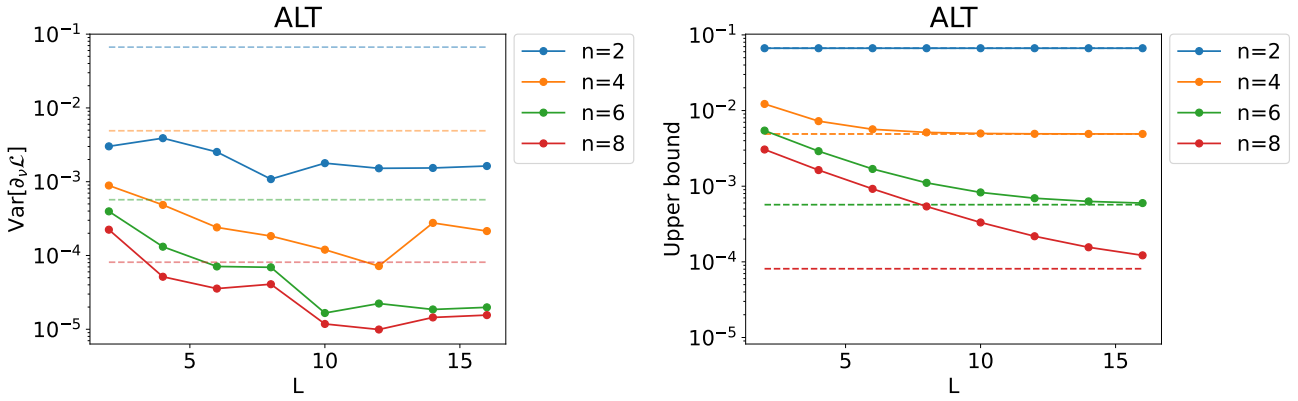


図 5.4.5: 入力回路に ALT を用いた場合のコスト関数の勾配の分散 (左) とその上界 (右)

我々は、特にデータの inputs が勾配の分散に与える影響に興味があった。そこで、上界の表式のうちデータ入力に依存する項 $\int_{U \in \mathcal{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s)$ だけに注目してプロットすると、図 5.4.6 のようになる。

解析的には示せていないが、図から曲線はある直線 β^{-L} ($1 < \beta$) を下回らないことが分かる。よって、上界が量子ビット数 n に対して指数関数的に落ちないためには、ある γ に対して $n^{-\gamma} < \beta^{-L}$ 、すなわち、 $L < \gamma \log n / \log \beta$ である必要がある。これが、入力回路に Alternating Layered Ansatz を用いた場合に、バレンプラトーが起きないための必要条件である。以上から、バレンプラトーが起きないためには学習回路が浅いことに加えて、入力回路も $\mathcal{O}(\log n)$ 程度に浅くしなければならないことが分かる。

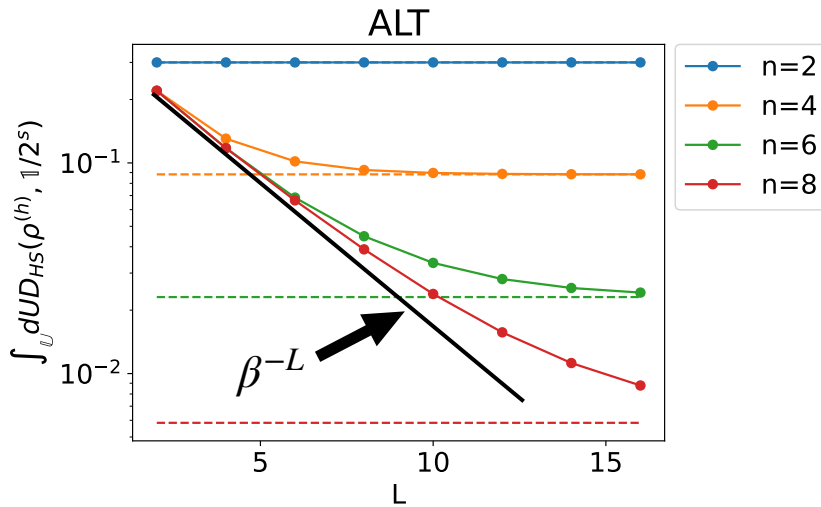


図 5.4.6: 入力回路に ALT を用いた時の $\int_{U \in \mathcal{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbb{1}/2^s)$ の計算結果

5.5 ノイズの影響

ここまでのコスト関数の勾配の分散の計算においては、ノイズは考慮していなかった。しかし、実際の量子コンピューターにはノイズが存在し、ノイズはバレンプラトーの原因の 1 つでもある。そこで、ノイズの影響も考慮した場合に、コスト関数の勾配の分散の上界がどのように変化するかを調べる。具体的には、5.4 節で行った計算に Depolarizing ノイズを仮定して、コスト関数の勾配の分散の上界を計算する。各層のユニタリチャネル $\mathcal{E}_i(\rho) = U_i \rho U_i^\dagger$, $i = 1, 2, \dots, L$ に付随して、次の Depolarizing noise

channel \mathcal{N}_p を作用させることを考える。

$$\mathcal{N}_p(\rho) = (1-p)\rho + \frac{p}{d}\mathbb{1}$$

このとき、全体のチャネル $\mathcal{N} = \bigcirc_{i=1}^L \mathcal{N}_p \circ \mathcal{E}_i$ は次のように表されることを既に確認した (2.9 節)。

$$\mathcal{N}(\rho) = (1-p)^L \mathcal{E}(\rho) + \{1 - (1-p)^L\} \frac{\mathbb{1}}{2^n}$$

ただし、 $\mathcal{E} := \bigcirc_{i=1}^L \mathcal{E}_i$ である。

$D_{HS}(\rho, \mathbb{1}/d) = \text{Tr}[\rho^2] - \frac{1}{d}$ であることを用いると、

$$\begin{aligned} D_{HS}(\mathcal{N}(\rho), \mathbb{1}/d) &= \text{Tr}[\mathcal{N}(\rho)^2] - \frac{1}{d} \\ &= (1-p)^L \left\{ (1-p)^L \text{Tr}[\rho^2] - \frac{1}{d} \right\} \\ &\leq (1-p)^L \left\{ \text{Tr}[\rho^2] - \frac{1}{d} \right\} \\ &= (1-p)^L D_{HS}(\rho, \mathbb{1}/d). \end{aligned} \quad (5.5.1)$$

したがって、ノイズの影響を考慮した場合のコスト関数の勾配の分散の上界は、次のようになる。

$$\begin{aligned} \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{noise}}(\boldsymbol{\theta})] &\leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i}(\boldsymbol{\theta}) f)^2] \times r_{n,s} \times \int_{U \in \mathcal{U}} dU D_{HS}(\mathcal{N}(\rho^{(h)}), \mathbb{1}/2^s) \\ &\leq 2 \max_{i, \boldsymbol{\theta}} [(\partial_{\ell_i}(\boldsymbol{\theta}) f)^2] \times r_{n,s} \times (1-p)^L \int_{U \in \mathcal{U}} dU D_{HS}(\rho^{(h)}, \mathbb{1}/2^s) \end{aligned} \quad (5.5.2)$$

図 5.4.5 の右はノイズがある場合の上界 (5.5.2) をプロットしたものである。また、5.4 節と同じ設定の下、Depolarizing ノイズ ($p = 0.03$) を加えてアヤメの分類問題のコスト関数の勾配の分散を計算すると、図 5.5.1 の左のようになった。勾配の分散の計算のために、全てのパラメーターを 10 回ランダムにサンプリングした。ただし、プロットしたのは、学習回路の各パラメーターに関する勾配の分散 $\text{Var}[\partial_{\theta_i} \mathcal{L}(\boldsymbol{\theta})]$ を i について平均したものである。図 5.4.5 と比較すると、コスト関数の勾配の分散がさらに小さくなっていることが分かる。上界のノイズに依る項 $(1-p)^L$ だけを見ても、バレンプラトーが起きないためには $L \in \mathcal{O}(\log n)$ でなければならないことが分かる。

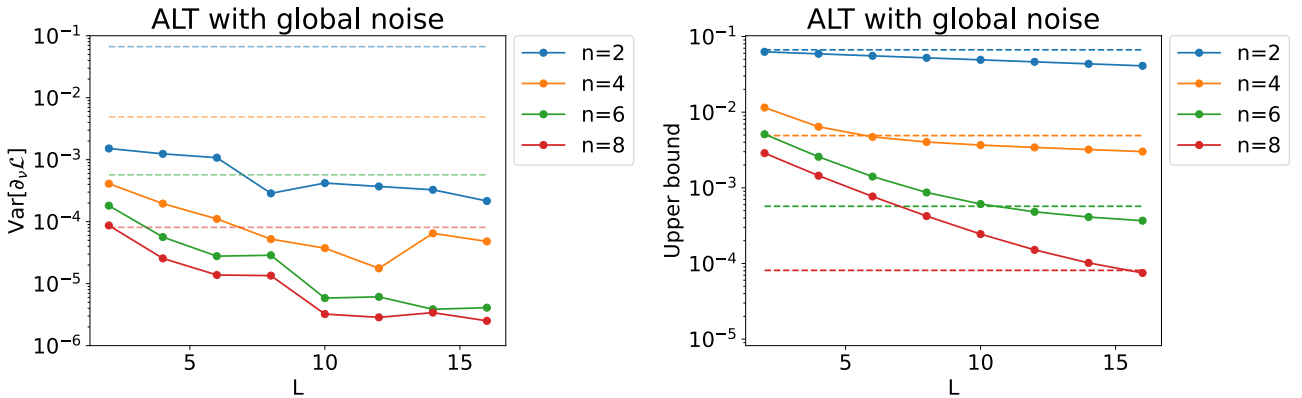


図 5.5.1: ノイズのある入力回路に ALT を用いた場合のコスト関数の勾配の分散 (左) とその上界 (右)

第6章 量子機械学習におけるコスト関数の勾配の分散の下界

この章では、コスト関数の勾配の分散の下界のスケーリングについて調べる。6.1 節では、入力データがガウス分布に従うという仮定の下、絶対誤差のコスト関数 $\mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N |l_i(\boldsymbol{\theta}) - y_i|$ の勾配の分散の下界を計算する。6.2 節では、二乗誤差のコスト関数 $\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N (l_i(\boldsymbol{\theta}) - y_i)^2$ や交差エントロピー誤差のコスト関数 $\mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N [-y_i \log l_i(\boldsymbol{\theta}) - (1 - y_i) \log(1 - l_i(\boldsymbol{\theta}))]$ の勾配の分散を、絶対誤差のコスト関数の勾配の分散から推定する。

6.1 絶対誤差のコスト関数の勾配の分散

この節では、量子回路の構造は図 5.1.1 で与えられるものとする。すなわち、学習回路は ξ 個の s 量子ビットユニタリからなり、全量子ビット数は $n = s \times \xi$ である。そして、2 値分類 $y_i \in \{0, 1\}$ の問題設定において、絶対誤差のコスト関数の勾配の分散の下界を計算する。

コスト関数が

$$\mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N |l_i(\boldsymbol{\theta}) - y_i|, \quad l_i(\boldsymbol{\theta}) = \text{Tr}[\rho_i(\boldsymbol{\theta}) O_L], \quad O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$$

で与えられるとき、コスト関数の勾配は、 $\partial_\nu := \partial/\partial\theta_\nu$, $w_i := \text{sgn}(l_i(\boldsymbol{\theta}) - y_i)$ として、

$$\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N w_i \cdot \partial_\nu l_i(\boldsymbol{\theta}) \quad (6.1.1)$$

と表せる。ただし、 $l_i(\boldsymbol{\theta}) \in [0, 1]$ であることから w_i と y_i は次のように対応する。

$$w_i = \begin{cases} 1 & \text{if } y_i = 0 \\ -1 & \text{if } y_i = 1 \end{cases} \quad (6.1.2)$$

ここで、学習回路の各 s 量子ビットユニタリがユニタリ 2-デザインを成すと仮定する。また、証明 5.1 で示したように、勾配の平均は 0 である。よって、勾配の分散は次のように計算できる。

$$\begin{aligned} \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] &= \text{E}_{V(\boldsymbol{\theta})}[(\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}))^2] \\ &= \frac{1}{N^2} \sum_{i,j=1}^N w_i w_j \text{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta}) \cdot \partial_\nu l_j(\boldsymbol{\theta})] \end{aligned} \quad (6.1.3)$$

$\text{E}_{V(\boldsymbol{\theta})}[\partial_\nu l_i(\boldsymbol{\theta}) \cdot \partial_\nu l_j(\boldsymbol{\theta})]$ は、論文 [27] の結果を拡張することで次のように計算できる。略証は C.1 節を参照。

系 6.1.1. (論文 [27] の Supplementary Note 5, B の拡張された結果) 量子回路の構造は図 5.1.1 で与えられ、学習回路の各 s 量子ビットユニタリがユニタリ 2-デザインを成すとき、次の等式が成り立つ。

$$E_{V(\theta)}[\partial_\nu \ell_i(\theta) \cdot \partial_\nu \ell_j(\theta)] = r_{n,s} \left(\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right] - \frac{1}{2^s} \right), \quad r_{n,s} := \frac{s 2^{3(s-1)}}{n^2 (2^{2s} - 1)^2} \quad (6.1.4)$$

よって、式 (6.1.3) は次のように書き換えられる。

$$\text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)] = \frac{r_{n,s}}{N^2} \sum_{i,j=1}^N w_i w_j \left(\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right] - \frac{1}{2^s} \right) \quad (6.1.5)$$

$$= \frac{r_{n,s}}{N^2} \left[\sum_{w_i=w_j}^N \left(\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right] - \frac{1}{2^s} \right) - \sum_{w_i \neq w_j}^N \left(\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right] - \frac{1}{2^s} \right) \right] \quad (6.1.6)$$

上の表式から、入力状態に依存する $\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right]$ が重要であり、 $\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right]$ が同じラベルの場合と異なるラベルの場合でどのようにスケールするかが、勾配の分散に影響することがわかる¹。

式 (6.1.5) をさらに変形すると、次のようになる。

$$\text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)] = r_{n,s} \left(\text{Tr}[(\bar{\rho})^2] - \frac{w^2}{2^s} \right) \quad (6.1.7)$$

ただし、 $w := \frac{1}{N} \sum_{i=1}^N w_i$, $\bar{\rho} := \frac{1}{N} \sum_{i=1}^N w_i \rho_i^{(h)}$ とした。前者は $w_i = \pm 1$ であることに注意すると、 $|w| \leq 1$ である。特に、 $\#\{y_i | y_i = 0\} = \#\{y_i | y_i = 1\} = N/2$ ならば $w = 0$ であり、 $\#\{y_i | y_i = 0\} = N$ または $\#\{y_i | y_i = 1\} = N$ ならば $|w| = 1$ である。後者は、論文 [69] において述べられているように Kernel Alignment と類似した量である。簡単のため、 $\bar{\rho}$ と表しているが、この量自体は密度演算子とは限らないことに注意する。なぜなら、 $\text{Tr}[\bar{\rho}] = w$ だからである。

さらに、 $\bar{\rho}_\pm := \frac{1}{N} \sum_{i, (w_i = \pm 1)} \rho_i^{(h)}$ として、 $\bar{\rho}$ を変形すると次のようになる。

$$\text{Tr}[(\bar{\rho})^2] = \text{Tr}[(\bar{\rho}_+ - \bar{\rho}_-)^2] = D_{HS}(\bar{\rho}_+, \bar{\rho}_-) \quad (6.1.8)$$

したがって、それぞれのラベルの平均入力状態である $\bar{\rho}_+$ と $\bar{\rho}_-$ の差が大きいほど、下界も大きくなることがわかる。

6.1.1 下界のスケール (1)

ここでは、先行研究 [70] で得られた結果を拡張し、コスト関数の勾配の分散の下界を計算する。入力データの各成分はガウス分布に従い、入力回路は図 6.1.1 のように R_y ゲートのみで構成されているとする。このとき、次の定理が成り立つ。証明は C.3 節を参照。

¹論文 [68] においては、入力状態が十分に分離していることを、同じラベルの入力状態間のカーネルと異なるラベルの入力状態間のカーネルが次のようにスケールことと述べている。

$$\begin{aligned} \text{Tr}[\rho_i \rho_{i'}] &\in \Omega\left(\frac{1}{\text{poly}(n)}\right), & \text{if } y_i = y_{i'} \\ \text{Tr}[\rho_i \rho_{i'}] &\in \mathcal{O}\left(\frac{1}{2^n}\right), & \text{if } y_i \neq y_{i'} \end{aligned}$$

定理 6.1.2. (論文 [70] の Theorem 1 の拡張された結果) $\mathcal{X} = \{\mathbf{x}\}$, $\mathcal{Z} = \{\mathbf{z}\}$ はそれぞれラベル $y_i = 0$ と $y_i = 1$ に属する入力データセットであり、 $|\mathcal{X}| : |\mathcal{Z}| = p : q$ ($p + q = 1$) であるとする。そして、各データサンプルを次のように表記する： $\mathbf{x} = (x_j)_{j=1}^n$, $\mathbf{x}_j = (x_{j,d})_{d=1}^L$, $\mathbf{z} = (z_j)_{j=1}^n$, $\mathbf{z}_j = (z_{j,d})_{d=1}^L$ 。異なるラベルに属する入力データを量子回路に同時には入力しないとする。入力データの各成分の分布は $x_{j,d} \sim \mathcal{N}(\mu_{x|j,d}, \sigma_{x|j,d}^2)$, $z_{j,d} \sim \mathcal{N}(\mu_{z|j,d}, \sigma_{z|j,d}^2)$ であり、分散は $\sigma_{\min} \leq \sigma_{x|j,d}, \sigma_{z|j,d} \leq \sigma_{\max}$ を満たすとする。また、量子回路は図 6.1.1 のように構成されている。ただし、 $1 \leq j \leq n$, $1 \leq d \leq L$ である。このとき、 $\bar{\rho} = p \mathbb{E}_{\mathcal{X}}[\rho^{(h)}(\mathbf{x})] - q \mathbb{E}_{\mathcal{Z}}[\rho^{(h)}(\mathbf{z})]$ について次の不等式が成り立つ。

$$\frac{1}{2^s} \left[w^2 + \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \right] \leq \text{Tr}[(\bar{\rho})^2] \leq \frac{1}{2^s} (1 + e^{-L\sigma_{\min}^2})^s \quad (6.1.9)$$

ただし、 n, s は量子ビット数 ($n = s \times \xi$)、 L は入力回路の層数、 $w = p - q$, $\Sigma_{x|j} := \sum_{d=1}^L \sigma_{x|j,d}^2$, $\Sigma_{z|j} := \sum_{d=1}^L \sigma_{z|j,d}^2$ である。

この定理においては、 θ_ν が含まれる h ($1 \leq h \leq \xi$) 番目の s 量子ビットについてだけ考えればよいことに注意する。

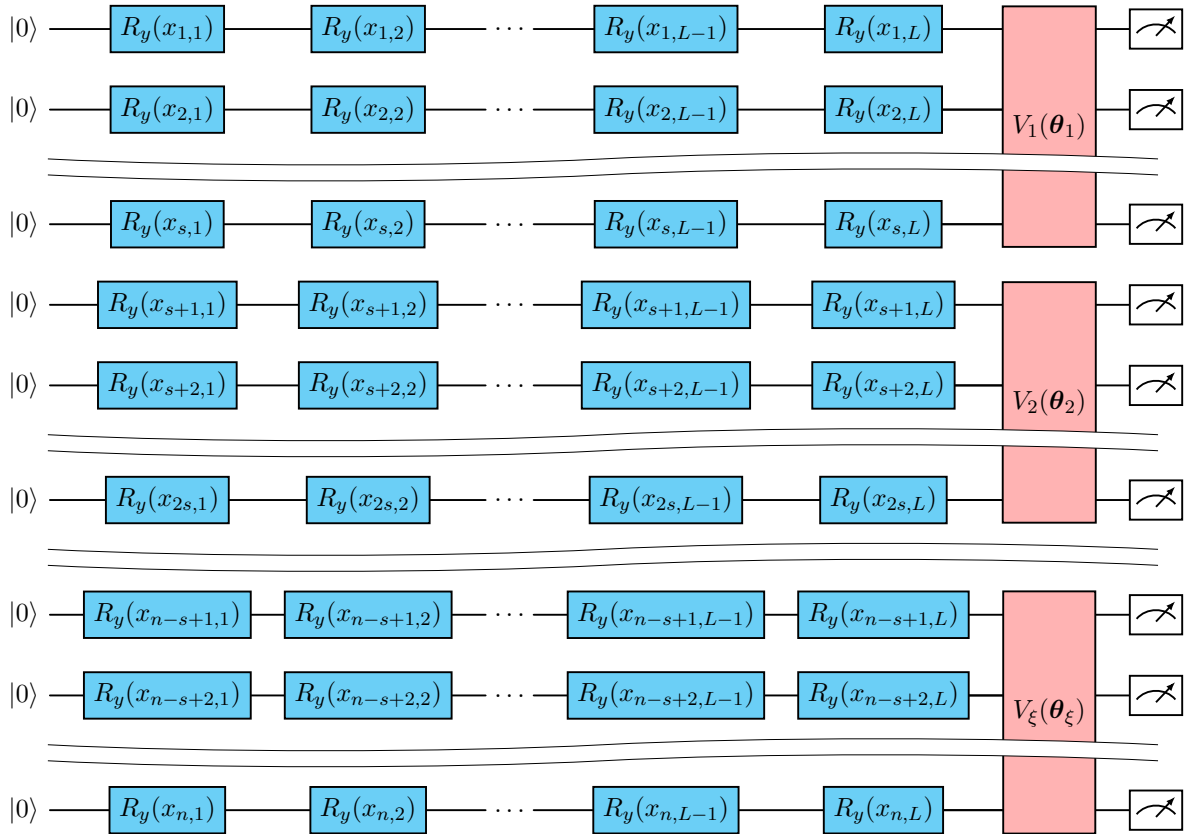


図 6.1.1: $n = s \times \xi$ 量子ビットからなる量子回路。入力回路は L 層の R_y ゲートからなる。学習回路は ξ 個の s 量子ビットユニタリからなる。

定理 6.1.2 と式 (6.1.7) より、直ちに次の定理が導かれる。

定理 6.1.3. 系 6.1.1 と定理 6.1.2 の条件のもとで、絶対誤差のコスト関数の勾配の分散は上下から次のように抑えられる。

$$\frac{r_{n,s}}{2^s} \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \leq \text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)] \leq \frac{r_{n,s}}{2^s} \left[\left(1 + e^{-L\sigma_{\min}^2} \right)^s - w^2 \right] \quad (6.1.10)$$

ただし、 $r_{n,s} := \frac{s 2^{3(s-1)}}{n^2(2^{2s}-1)^2}$, $w = p - q$ である。

$|\mathcal{X}| : |\mathcal{Z}| = 1 : 1$ の場合

$|\mathcal{X}| : |\mathcal{Z}| = 1 : 1 \iff p = q = 1/2, w = 0$ の場合、定理 6.1.3 は次のようになる。

$$\frac{r_{n,s}}{2^{s+2}} \sum_{j=1}^s \left(e^{-\Sigma_{x|j}/2} - e^{-\Sigma_{z|j}/2} \right)^2 \leq \text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)] \leq r_{n,s} \left(\frac{1 + e^{-L\sigma_{\min}^2}}{2^s} \right)^s \leq r_{n,s} \quad (6.1.11)$$

これより $e^{-\Sigma_{x|j}/2}, e^{-\Sigma_{z|j}/2}$ の差が大きいほど、下界が大きくなることがわかる。しかし、 $(e^{-\Sigma_{x|j}/2} - e^{-\Sigma_{z|j}/2})^2$ は層数 L を増やすと指数関数的に小さくなる。また、すべての j について $e^{-\Sigma_{x|j}/2} = e^{-\Sigma_{z|j}/2}$ である場合、下界は 0 となり意味をなさない²。

一方で、 $s \in \mathcal{O}(\text{poly}(n))$ ならば、上界は指数関数的に減少するため、バレンプラトーとなることがわかる。

$|\mathcal{X}| : |\mathcal{Z}| = 1 : 0$ の場合

異常検知などに用いられる One class classification [71] の場合は、 $|\mathcal{X}| : |\mathcal{Z}| = 1 : 0 \iff p = 1, q = 0, w = 1$ となる。このとき、定理 6.1.3 は次のようになる。

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

$\Sigma_{x|j} \leq L\sigma_{\max}^2$ であるから、 $s e^{-L\sigma_{\max}^2} \leq \sum_{j=1}^s e^{-\Sigma_{x|j}}$ が成り立つ。これと、 $[(1 + e^{-L\sigma_{\min}^2})^s - 1]/2^s \leq 1$ であることから、定理 6.1.3 は次のようになる。

$$\frac{r_{n,s}s}{2^s} e^{-L\sigma_{\max}^2} \leq \text{Var}_{V(\theta)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\theta)] \leq r_{n,s} \quad (6.1.12)$$

s と σ_{\max}^2 が固定されているとすると、勾配の分散の下界が量子ビット数 n に関して $\mathcal{O}(1/\text{poly}(n))$ でスケールするためには、 $L \in \mathcal{O}(\log n)$ でなければならない。ところで、3.2.3 節の量子カーネル法の説明で述べた Bandwidth というハイパーパラメーターを導入すれば、入力するデータの分散を調整することができる。よって分散の最大値が $\sigma_{\max}^2 \in \mathcal{O}(\log n)/L$ を満たすように Bandwidth を調整すれば、 $L\sigma_{\max}^2 \in \mathcal{O}(\log n)$ となるようにできる。しかし、この手法によってバレンプラトーが回避できても、汎化性能に悪影響を及ぼす可能性があることに注意する。

一方、上界については $s \in \mathcal{O}(\text{poly}(n))$ のとき、上界は指数関数的に減少するため、バレンプラトーとなることがわかる。

²極端な例として、 $\mathcal{X} = \mathcal{Z}$ の場合は、 $\mathcal{L}_{\text{MAE}}(\theta) = 1/2$ となるため、勾配は 0、したがって勾配の分散も 0 となる。ゆえに、下界が 0 となることもあり得る。

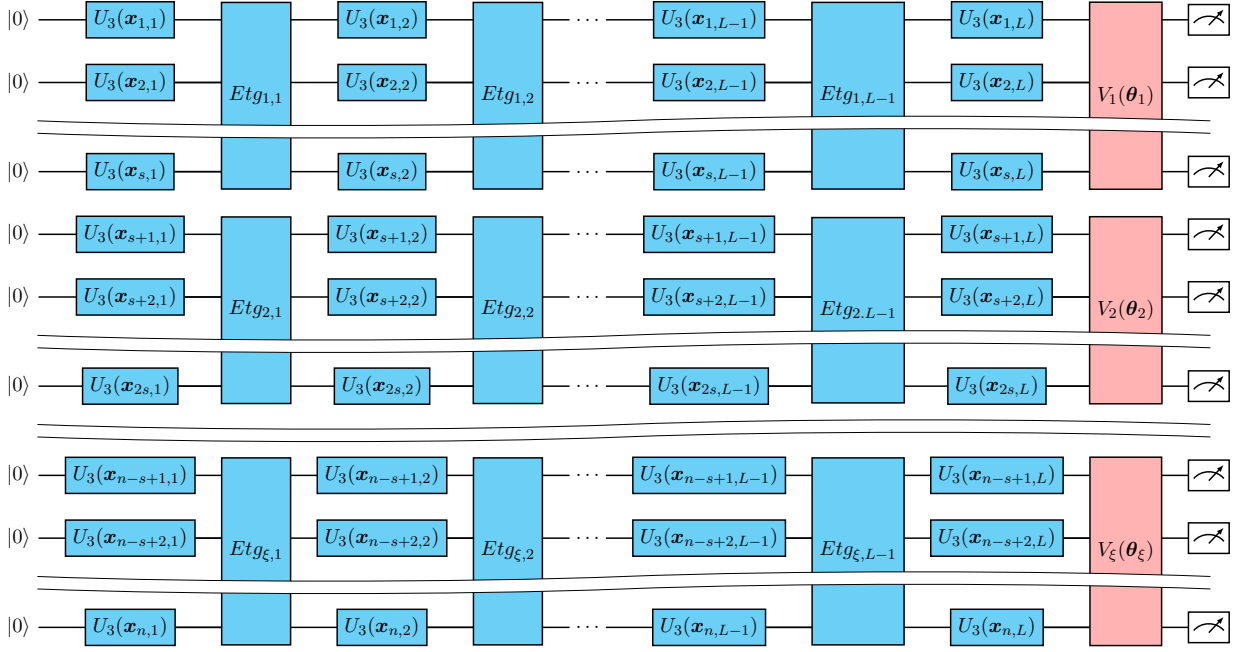


図 6.1.2: $n = s \times \xi$ 量子ビットからなる量子回路。入力回路は、データを入力する $U_3(\mathbf{x}_{j,d})$ ゲートの L 層と、CNOT ゲートや CZ ゲートによってエンタングルメントを生成する層 $Etg_{g,d'}$ の $L-1$ 層からなる。ただし、 $1 \leq j \leq n$, $1 \leq d \leq L$, $1 \leq g \leq \xi$, $1 \leq d' \leq L-1$ である。

6.1.2 下界のスケーリング (2)

ここでも先程と同様に、先行研究 [70] で得られた結果を拡張し、コスト関数の勾配の分散の下界を計算する。ただし、この節では $p = 1$, $q = 0$ である場合のみを考える。入力データの各成分はガウス分布に従い、入力回路は図 6.1.2 のように構成されているとする。このとき、次の定理が成り立つ。略証は C.4 節を参照。

定理 6.1.4. (論文 [70] の Theorem 2 の拡張された結果) $\mathcal{X} = \{\mathbf{x}\}$ はラベル $y_i = 0$ に属する入力データセットであるとする。そして、各データサンプルを次のように表記する： $\mathbf{x} = (\mathbf{x}_j)_{j=1}^n$, $\mathbf{x}_j = (\mathbf{x}_{j,d})_{d=1}^L$, $\mathbf{x}_{j,d} = (x_{j,d,1}, x_{j,d,2}, x_{j,d,3})$ 。入力データの各成分の分布は $x_{j,d,k} \sim \mathcal{N}(\mu_{x_{j,d,k}}, \sigma_{x_{j,d,k}}^2)$ であり、分散は $\sigma_{\min} \leq \sigma_{x_{j,d,k}} \leq \sigma_{\max}$ を満たすとする。また、量子回路は図 6.1.2 のように構成されている。ただし、 $1 \leq j \leq n$, $1 \leq d \leq L$, $1 \leq k \leq 3$ である。このとき、 $\bar{\rho} := \mathbb{E}_{\mathcal{X}}[\rho^{(h)}(\mathbf{x})]$ について次の不等式が成り立つ。

$$\frac{1}{2^s} + e^{-3sL\sigma_{\max}^2} \leq \text{Tr}[(\bar{\rho})^2] \leq \frac{1}{2^s} + e^{-L\sigma_{\min}^2} \quad (6.1.13)$$

ただし、 n, s は量子ビット数 ($n = s \times \xi$)、 L は入力回路の層数である。

この定理においては、 θ_h が含まれる h ($1 \leq h \leq \xi$) 番目の s 量子ビットについてだけ考えればよいことに注意する。

定理 6.1.4 と式 (6.1.7) より、直ちに次の定理が導かれる。

定理 6.1.5. 系 6.1.1 と定理 6.1.4 の条件のもとで、絶対誤差のコスト関数の勾配の分散は上下から次のように抑えられる。

$$r_{n,s} e^{-3sL\sigma_{\max}^2} \leq \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] \leq r_{n,s} e^{-L\sigma_{\min}^2} \quad (6.1.14)$$

ただし、 $r_{n,s} := \frac{s 2^{3(s-1)}}{n^2(2^{2s}-1)^2}$ である。

定理 6.1.5 の不等式の下界は 6.1.1 項の $p = 1$ の場合と同じ形であるため、同様に議論できる。すなわち、 s と σ_{\max}^2 が固定されているとすると、 $L \in \mathcal{O}(\log n)$ ならば、勾配の分散の下界は $\mathcal{O}(1/\text{poly}(n))$ となり、バレンプラトーにならない。一方、上界においては、 $s \in \mathcal{O}(\text{poly}(n))$ のとき、あるいは、 σ_{\min} が一定の下で $L \in \mathcal{O}(\text{poly}(n))$ のとき、バレンプラトーとなることがわかる。

6.2 他のコスト関数の勾配の分散に関する考察

前節では、絶対誤差のコスト関数を用いたときの勾配の分散について考察した。ここでは、それ以外のコスト関数を用いたときの勾配の分散について考察する。まず、絶対誤差のコスト関数 $\mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{i=1}^N |\ell_i(\boldsymbol{\theta}) - y_i|$ について、その勾配は次のように計算されるのであった。

$$\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{i=1}^N \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \quad (6.2.1)$$

一方、二乗誤差のコスト関数 $\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{i=1}^N (\ell_i(\boldsymbol{\theta}) - y_i)^2$ について、その勾配は次のように計算される。

$$\begin{aligned} \partial_\nu \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) &= \frac{1}{N} \sum_{i=1}^N 2(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{i=1}^N 2|\ell_i(\boldsymbol{\theta}) - y_i| \cdot \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \end{aligned} \quad (6.2.2)$$

また、交差エントロピー誤差のコスト関数 $\mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) := \frac{1}{N} \sum_{i=1}^N [-y_i \log \ell_i(\boldsymbol{\theta}) - (1 - y_i) \log(1 - \ell_i(\boldsymbol{\theta}))]$ について、その勾配は次のように計算される。

$$\begin{aligned} \partial_\nu \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta}) &= \frac{1}{N} \sum_{i=1}^N -\frac{y_i}{\ell_i(\boldsymbol{\theta})} \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) + \frac{1 - y_i}{1 - \ell_i(\boldsymbol{\theta})} \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{y_i=0} \frac{1}{1 - \ell_i(\boldsymbol{\theta})} \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) + \frac{1}{N} \sum_{y_i=1} \frac{1}{-\ell_i(\boldsymbol{\theta})} \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{1 - \ell_i(\boldsymbol{\theta}) - y_i} \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{|1 - \ell_i(\boldsymbol{\theta}) - y_i|} \cdot \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{1 - |\ell_i(\boldsymbol{\theta}) - y_i|} \cdot \text{sgn}(\ell_i(\boldsymbol{\theta}) - y_i) \cdot \partial_\nu \ell_i(\boldsymbol{\theta}) \end{aligned} \quad (6.2.3)$$

ただし、 $\ell_i(\boldsymbol{\theta}) \in [0, 1]$, $y_i \in \{0, 1\}$ であることを用いた。

ここで、 $\ell_i(\boldsymbol{\theta})$ の平均と分散に関して次の補題が成り立つ。

補題 6.2.1. 学習回路 V がユニタリ 2-デザインを成す場合、 $l_i(\boldsymbol{\theta}) = \text{Tr}[V\rho_iV^\dagger O_L]$ の平均は $1/2$ 、分散は $1/(4n(2^n + 1))$ となる。ただし、 ρ_i は入力状態、 $O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbf{1}_{\bar{j}}$ は測定に用いるオブザーバブルである。

証明. まず、第 j 番目の量子ビットにおける Pauli Z を Z_j とし、状態 $V\rho_iV$ による Z_j の期待値を $\langle Z_j \rangle := \text{Tr}[V\rho_iVZ_j]$ と表す。このとき、 $\langle Z_j \rangle$ の V に関する平均と分散、そして共分散 $\text{Cov}[\langle Z_j \rangle, \langle Z_{j' \neq j} \rangle]$ は、 V がユニタリ 2-デザインからサンプリングされるとき、公式 (4.1.7), (4.1.5) を用いて次のように計算される。

$$\mathbb{E}_{U(d)}[\langle Z_j \rangle] = 0 \quad (6.2.4)$$

$$\text{Var}_{U(d)}[\langle Z_j \rangle] = \frac{1}{2^n + 1} \quad (6.2.5)$$

$$\text{Cov}_{U(d)}[\langle Z_j \rangle, \langle Z_{j' \neq j} \rangle] = 0 \quad (6.2.6)$$

ここで、 Z_j はオブザーバブル $O_{j|0} := |0\rangle\langle 0|_j$ によって $Z_j := 2O_{j|0} - \mathbf{1}$ と表すことができる。つまり、 $\langle Z_j \rangle = 2\langle O_{j|0} \rangle - 1$ となる。この等式と $\langle Z_j \rangle$ の平均と分散、共分散から、 $\langle O_{j|0} \rangle$ の平均と分散、共分散は次で与えられる。

$$\mathbb{E}_{U(d)}[\langle O_{j|0} \rangle] = \frac{1}{2} \quad (6.2.7)$$

$$\text{Var}_{U(d)}[\langle O_{j|0} \rangle] = \frac{1}{4(2^n + 1)} \quad (6.2.8)$$

$$\text{Cov}_{U(d)}[\langle O_{j|0} \rangle, \langle O_{j'|0} \rangle] = 0 \quad (6.2.9)$$

ところで、 $l_i(\boldsymbol{\theta})$ は $\langle O_{j|0} \rangle$ を用いて次のように表される。

$$l_i(\boldsymbol{\theta}) = \text{Tr} \left[V\rho_iV \left(\frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbf{1}_{\bar{j}} \right) \right] = \frac{1}{n} \sum_{j=1}^n \langle O_{j|0} \rangle \quad (6.2.10)$$

したがって、 $l_i(\boldsymbol{\theta})$ の平均と分散は次で与えられる。

$$\mathbb{E}_{U(d)}[l_i(\boldsymbol{\theta})] = \frac{1}{2} \quad (6.2.11)$$

$$\text{Var}_{U(d)}[l_i(\boldsymbol{\theta})] = \frac{1}{4n(2^n + 1)} \quad (6.2.12)$$

■

補題 6.2.1 とチェビシェフの不等式より、 $\Pr[|l_i(\boldsymbol{\theta}) - 1/2| \geq \epsilon] \leq 1/(\epsilon^2 \cdot 4n(2^n + 1))$ が成り立ち、 $l_i(\boldsymbol{\theta})$ は指数関数的に $1/2$ に集中することがわかる。そこで、すべての i について $l_i(\boldsymbol{\theta}) \sim 1/2$ 、すなわち $|l_i(\boldsymbol{\theta}) - y_i| \sim 1/2$ と近似する。このとき、式 (6.2.2), (6.2.3) から $\partial_\nu \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})$, $\partial_\nu \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta})$ は $\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})$ によって次のように近似される。

$$\partial_\nu \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) \sim \partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta}) \quad (6.2.13)$$

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

ゆえに、勾配の分散の比は次のように近似される³。

$$\begin{aligned} \text{Var}_{\mathcal{U}(d)}[\partial_\nu \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})] / \text{Var}_{\mathcal{U}(d)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] &\sim 1 \\ \text{Var}_{\mathcal{U}(d)}[\partial_\nu \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta})] / \text{Var}_{\mathcal{U}(d)}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})] &\sim 4 \end{aligned} \tag{6.2.14}$$

これにより、二乗誤差と交差エントロピー誤差のコスト関数の勾配の分散は、絶対誤差のコスト関数の勾配の分散と同程度のスケージングであると考えられる。

ここで得られた近似式 (6.2.14) は、学習回路が十分深いときに成り立つ。しかし、学習回路が浅い場合には、この近似式は成り立たない可能性がある。そこで、その適用範囲を調べるために、数値実験を行った。

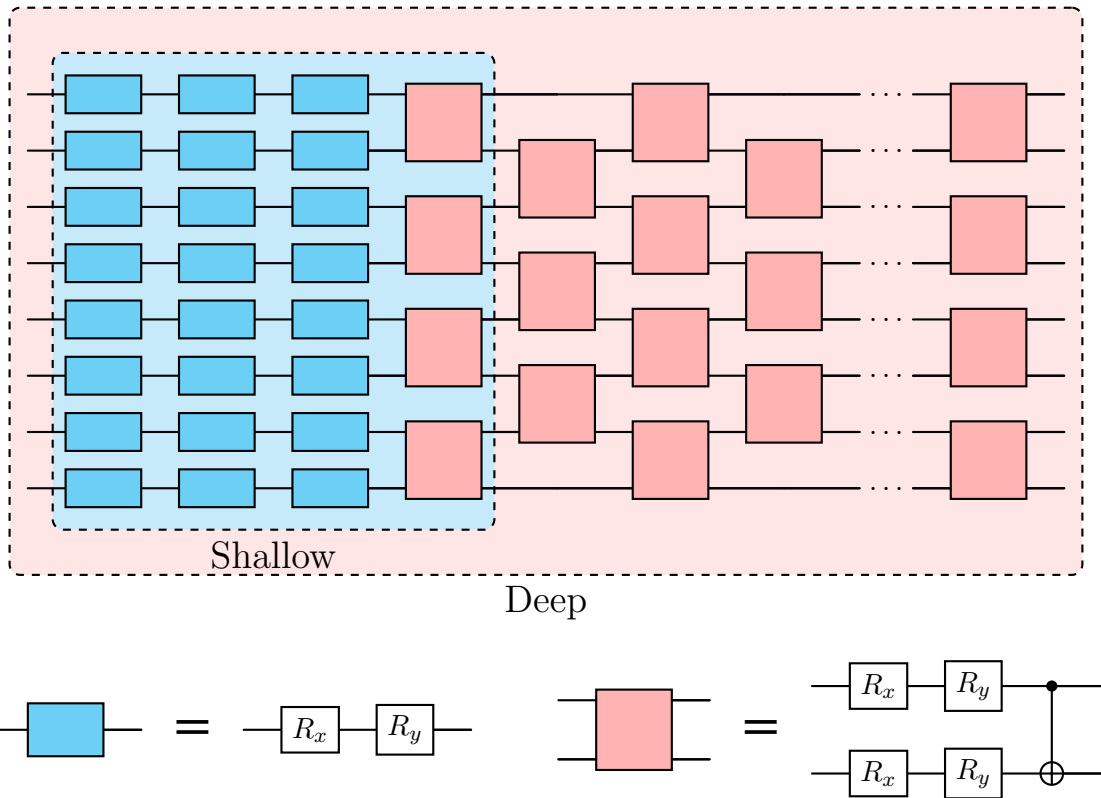


図 6.2.1: コスト関数の勾配の分散の比を計算するために用いた量子回路の構造。青いゲートブロックは Tensor Product Ansatz の入力回路、赤いゲートブロックは Alternating Layered Ansatz の学習回路を表す。

実験の設定は次の通りである。量子回路は、図 6.2.1 にあるように、Tensor Product Ansatz 3 層の入力回路と、Alternating Layered Ansatz の学習回路で与えられる。5.4 節において、アヤメのデータセットを用いてコスト関数の勾配の分散を計算したが、ここでも同じデータセットを用いて、各量子ビット数 n 、学習回路の各層数 L_{red} に対して 3 つのコスト関数の勾配の分散を計算した。これらから得られる勾配の分散の比 $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})] / \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})]$, $\text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{LOG}}(\boldsymbol{\theta})] / \text{Var}_{V(\boldsymbol{\theta})}[\partial_\nu \mathcal{L}_{\text{MAE}}(\boldsymbol{\theta})]$ をプロットしたものを図 6.2.2 に示す。

ただし、勾配の分散の計算のために、全てのパラメーターを 100 回ランダムにサンプリングした。そ

³厳密な議論は今後の課題とする

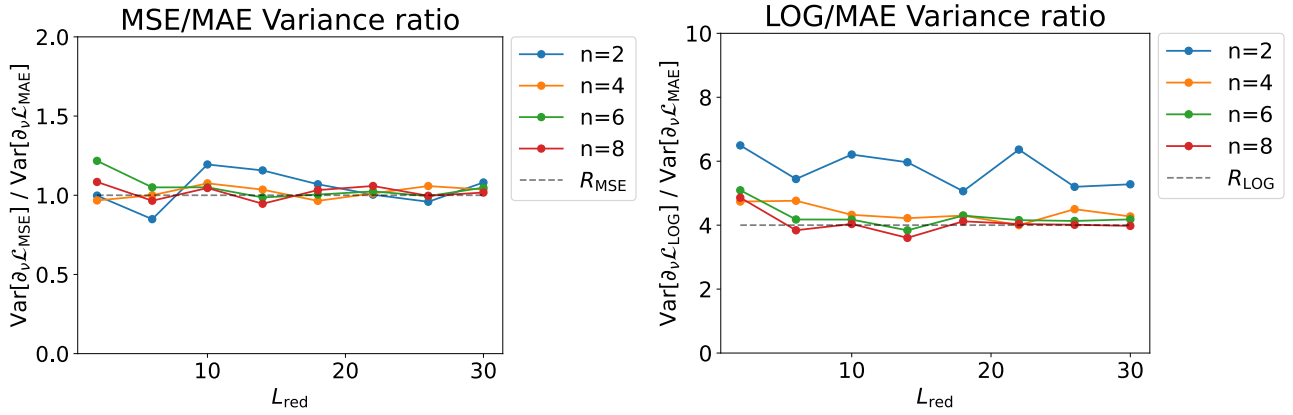


図 6.2.2: 図 6.2.1 の量子回路を用いて計算したそれぞれのコスト関数の勾配の分散から、絶対誤差の勾配の分散との比を計算した。左は、二乗誤差と絶対誤差の勾配の分散の比、右は、交差エントロピー誤差と絶対誤差の勾配の分散の比をプロットしたものである。破線は、式 (6.2.14) で得られた近似比 (左: $R_{\text{MSE}} = 1$ 、右: $R_{\text{LOG}} = 4$) を表している。

して、学習回路の各パラメーターに関する勾配の分散 $\text{Var}[\partial_{\theta_i} \mathcal{L}(\boldsymbol{\theta})]$ を i について平均したものをプロットした。

図 6.2.2 の左は、二乗誤差と絶対誤差の勾配の分散の比、右は、交差エントロピー誤差と絶対誤差の勾配の分散の比である。破線は、式 (6.2.14) の近似比 (1 または 4) を表している。両方の数値計算の結果とも、量子ビット数 $n = 2$ のときはややばらつきがあるが、各量子ビット数 n 、学習回路の各層数 L_{red} において式 (6.2.14) の近似比 (1 または 4) に近い値をとっていることがわかる。特に、学習回路の層数 L_{red} が少なく、ユニタリ 2-デザインを成すには不十分な場合にも、式 (6.2.14) の近似比 (1 または 4) に近い値をとっている。量子ビット数 $n = 2$ のときはばらつきがやや大きいのは、量子ビット数が少ないほど $l_i(\boldsymbol{\theta})$ の分散が大きくなり、 $l_i(\boldsymbol{\theta}) \sim 1/2$ の近似が成り立たないためであると考えられる。

このことから、数値的にも、二乗誤差と交差エントロピー誤差のコスト関数の勾配の分散は、絶対誤差のコスト関数の勾配の分散と同程度のスケールングであると考えられる。したがって、バレンプラトートの解析においては、いずれかひとつについてのみ考えれば、他のコスト関数についても同様の結果が得られると推測される。

以上の議論では $O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_j$ をオブザーバブルに用いたが、一般的に、Pauli 演算子の線形結合で表されるオブザーバブルを用いても、同様に $l_i(\boldsymbol{\theta})$ の値の集中が示される。例えば、測定するオブザーバブルを $O = c_0 \mathbb{1} + \sum_{\{i\} \neq \emptyset} c_i P_i$ と表す。ただし、 $P_i = \bigotimes_{j=1}^n P_{i_j}$, $i_j \in \{0, 1, 2, 3\}$ であり、 P_{i_j} は第 j 番目の量子ビットにおける Pauli 演算子、 c_i は実数である。このとき、 $l_i(\boldsymbol{\theta}) = \text{Tr}[V \rho_i V O]$ の平均と分散は次のように計算される。

$$E_{U(d)}[l_i(\boldsymbol{\theta})] = c_0 \quad (6.2.15)$$

$$\text{Var}_{U(d)}[l_i(\boldsymbol{\theta})] = \frac{1}{2^n + 1} \sum_{\{i\} \neq \emptyset} c_i^2 \quad (6.2.16)$$

よって、 $\sum_{\{i\} \neq \emptyset} c_i^2 \in \mathcal{O}(\text{poly}(n))$ であれば、チェビシエフの不等式から $l_i(\boldsymbol{\theta})$ は指数関数的に c_0 に集中することがわかる。したがって、この節における議論は、 O_L に限らず、一般的なオブザーバブルについても成り立つと考えられる。

第7章 結論

近年、機械学習はさまざまな分野で応用され、著しい発展を遂げている。一方で、計算量が膨大であり、計算資源の不足が懸念されている。そこで、量子コンピューターによって機械学習を効率化するための研究が盛んに行われている。本研究は、NISQ 上での実装が可能な変分量子アルゴリズムの文脈において、教師あり量子機械学習の効率化に関する研究を行った。

量子機械学習アルゴリズムの効率化においては、データ入力が必要であることが指摘されている一方で、バレンプラトーという勾配消失問題との関連も指摘されている。バレンプラトーには、量子回路の深さ、オブザーバブルの局所性、ノイズ、データの入力に関係していることが知られている。どれか一つの原因でも存在すると、計算量は量子ビット数に関して指数関数的に増大してしまい、学習がほとんど進まない。しかしながら、これまでのバレンプラトーについては主に学習回路の構造に着目して研究されてきたため、データ入力が必要であることがバレンプラトーに与える影響は十分には理解されていなかった。そこで、本研究は、量子機械学習に特有のデータ入力が必要であることがバレンプラトーに与える影響を評価した。特に、量子機械学習における識別モデルのバレンプラトーについて、コスト関数の勾配の分散のスケールを評価することを目的とした。

第5章では、量子機械学習における識別モデルのコスト関数のバレンプラトーについて、学習回路やオブザーバブルの局所性によるバレンプラトーが起きない設定の下で、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、数値的にも検証を行った。その結果、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ほど、コスト関数の勾配の分散の上界が小さくなり、バレンプラトーにつながる事が明らかになった。また、量子回路に Depolarizing ノイズを加えた場合に、その上界が更に小さくなることを示した。

第6章では、量子機械学習における識別モデルのコスト関数のバレンプラトーについて、特に、絶対誤差のコスト関数の勾配の分散の下界を、ガウス分布に従う入力データの観点から導出した。これにより、絶対誤差のコスト関数の勾配の分散の下界において、入力データの分散が重要な役割を果たすことを示した。そして、特定の場合において、バレンプラトーを引き起こさないための訓練データの分散に関する十分条件を得た。また、二乗誤差や交差エントロピー誤差のコスト関数の勾配の分散は、絶対誤差のコスト関数の勾配の分散と同じようにスケールすることを数値的に示した。これにより、量子機械学習におけるバレンプラトーの解析では、どれか1つのコスト関数の勾配の分散を評価すれば十分であると推測される。

今後必要となる研究には、次の2点が挙げられる。まず、本研究では学習回路に単純な構造を仮定したが、より複雑な構造を用いた場合のスケールを調べる事が望まれる。また、本研究では量子回路学習モデルのバレンプラトーについてのみ議論したが、より一般的な議論のためには、Data reuploading モデルについても解析が必要である。

しかし、訓練可能性だけが、量子機械学習の性能を決定するわけではない。データ入力のための量子回路の構造には、古典コンピューターによる効率的な計算が難しいこと、汎化性能が高いことなども重要である。将来的に、これらの性能についても評価した上で、量子機械学習のデータ入力を設計することが期待される。

謝辞

本研究では多くの方にサポートして頂きました。皆様に心より感謝申し上げます。指導教官である寺師弘二准教授、永野廉人特任助教には、研究の方向性・問題点・計算に関する議論・資料の校正まで何度も付き合ってもらい、多くの助言を頂きました。また、量子コンピューティンググループのミーティングを通して、田中純一教授、澤田龍准教授、飯山悠太郎助教、Wai Yuen Chan 特任研究員、加地 俊瑛 特任研究員には研究についての質問や指摘、研究発表へのアドバイスを頂きました。張元豪さん、寺尾萌里乃さん、尾嶋陸斗さん、山崎峻さんにはゼミやミーティングでの議論、日頃の相談に付き合ってもらいました。そして、ICEPP 物理事務室のみなさまには修士過程の生活を送る上で各種サポートをして頂きました。最後に、日々の生活を支えてくれた家族に心から感謝します。

参考文献

- [1] P. Benioff, “The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines”, *Journal of Statistical Physics* **22**, 563–591 (1980) (cited on page 5).
- [2] “Quantum theory, the church–turing principle and the universal quantum computer”, *Proc. R. Soc. Lond. A* **400**, 97–117 (1985) (cited on page 5).
- [3] R. P. Feynman et al., “Simulating physics with computers”, *International Journal of Theoretical Physics* **21**, 467–488 (1982) (cited on page 5).
- [4] P. W. Shor, “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer”, *SIAM J. Comput.* **26**, 1484–1509 (1997) (cited on page 5).
- [5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, “Quantum supremacy using a programmable superconducting processor”, *Nature* **574**, 505–510 (2019) (cited on page 5).
- [6] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, “Evidence for the utility of quantum computing before fault tolerance”, *Nature* **618**, 500–505 (2023) (cited on page 5).
- [7] L. K. Grover, “A fast quantum mechanical algorithm for database search”, in *Proceedings of the twenty-eighth annual acm symposium on theory of computing* (1996), pages 212–219 (cited on page 5).
- [8] A. W. Harrow, A. Hassidim, and S. Lloyd, “Quantum algorithm for solving linear systems of equations”, *Physical Review Letters* **103**, 150502 (2009) (cited on pages 5, 24).
- [9] J. Preskill, *Quantum Computing in the NISQ era and beyond*, (July 31, 2018) preprint (cited on pages 5, 20).

- [10] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, “Variational Quantum Algorithms”, *Nature Reviews Physics* **3**, 625–644 (2021) (cited on page 5).
- [11] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, “A variational eigenvalue solver on a photonic quantum processor”, *Nature Communications* **5**, 4213 (2014) (cited on page 6).
- [12] E. Farhi, J. Goldstone, and S. Gutmann, *A Quantum Approximate Optimization Algorithm*, (Nov. 14, 2014) preprint (cited on page 6).
- [13] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, “Variational ansatz-based quantum simulation of imaginary time evolution”, *npj Quantum Information* **5**, 10.1038/s41534-019-0187-2 (cited on page 6).
- [14] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, “Theory of variational quantum simulation”, *Quantum* **3**, 191 (2019) (cited on page 6).
- [15] I. M. Georgescu, S. Ashhab, and F. Nori, “Quantum simulation”, *Reviews of Modern Physics* **86**, 153 (2014) (cited on page 6).
- [16] M. C. Bañuls, R. Blatt, J. Catani, A. Celi, J. I. Cirac, M. Dalmonte, L. Fallani, K. Jansen, M. Lewenstein, S. Montangero, C. A. Muschik, B. Reznik, E. Rico, L. Tagliacozzo, K. Van Acoleyen, F. Verstraete, U.-J. Wiese, M. Wingate, J. Zakrzewski, and P. Zoller, “Simulating lattice gauge theories within quantum technologies”, *The European Physical Journal D* **74**, 10.1140/epjd/e2020-100571-8 (2020) (cited on page 6).
- [17] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, “Quantum-assisted quantum compiling”, *Quantum* **3**, 140 (cited on page 6).
- [18] K. Sharma, S. Khatri, M. Cerezo, and P. J. Coles, “Noise resilience of variational quantum compiling”, *New Journal of Physics* **22**, 043006 (2020) (cited on page 6).
- [19] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning”, *Physical Review A* **98**, 032309 (2018) (cited on pages 6, 22–24).
- [20] E. Farhi and H. Neven, *Classification with Quantum Neural Networks on Near Term Processors*, (Aug. 30, 2018) preprint (cited on pages 6, 24).
- [21] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum enhanced feature spaces”, *Nature* **567**, 209–212 (2019) (cited on pages 6, 24–25).
- [22] M. Schuld and N. Killoran, “Quantum machine learning in feature Hilbert spaces”, *Physical Review Letters* **122**, 040504 (2019) (cited on pages 6, 24–25).
- [23] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes”, *Nature Communications* **9**, 4812 (2018) (cited on pages 6, 26–28, 72).

- [24] A. Arrasmith, M. Cerezo, P. Czarnik, L. Cincio, and P. J. Coles, “Effect of barren plateaus on gradient-free optimization”, *Quantum* **5**, 558 (2021) (cited on pages 6, 28).
- [25] C. O. Marrero, M. Kieferová, and N. Wiebe, *Entanglement Induced Barren Plateaus*, (Mar. 10, 2021) <http://arxiv.org/abs/2010.15968>, preprint (cited on pages 6, 28).
- [26] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, “Connecting ansatz expressibility to gradient magnitudes and barren plateaus”, *PRX Quantum* **3**, 010313 (2022) (cited on pages 6, 28, 30, 37–38, 48).
- [27] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, “Cost function dependent barren plateaus in shallow parametrized quantum circuits”, *Nature Communications* **12**, 1791 (2021) (cited on pages 6, 28, 30, 32, 43–44, 53–54, 78).
- [28] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, “Noise-Induced Barren Plateaus in Variational Quantum Algorithms”, *Nature Communications* **12**, 6961 (2021) (cited on pages 6, 27–28).
- [29] S. Thanasilp, S. Wang, N. A. Nghiem, P. J. Coles, and M. Cerezo, “Subtleties in the trainability of quantum machine learning models”, [10.48550/arXiv.2110.14753](https://arxiv.org/abs/10.48550/arXiv.2110.14753) (2021) (cited on pages 6, 28, 40–41).
- [30] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, *Exponential concentration and untrainability in quantum kernel methods*, (Aug. 23, 2022) <http://arxiv.org/abs/2208.11060>, preprint (cited on pages 6, 26, 28, 40, 42).
- [31] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning”, *Nature* **549**, 195–202 (2017) (cited on page 6).
- [32] I. Cong, S. Choi, and M. D. Lukin, “Quantum Convolutional Neural Networks”, *Nature Physics* **15**, 1273–1278 (2019) (cited on pages 6, 30).
- [33] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, “Power of data in quantum machine learning”, *Nature Communications* **12**, 2631 (2021) (cited on pages 6, 26).
- [34] M. C. Caro, E. Gil-Fuster, J. J. Meyer, J. Eisert, and R. Sweke, “Encoding-dependent generalization bounds for parametrized quantum circuits”, *Quantum* **5**, 582 (2021) (cited on page 6).
- [35] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*, 10th anniversary ed (Cambridge University Press, Cambridge ; New York, 2010), 676 pages (cited on page 18).
- [36] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, “Quantum Natural Gradient”, *Quantum* **4**, 269 (2020) (cited on page 22).
- [37] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Faehrmann, B. Meynard-Piganeau, and J. Eisert, “Stochastic gradient descent for hybrid quantum-classical optimization”, *Quantum* **4**, 314 (2020) (cited on page 22).

- [38] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization”, (2017) (cited on page 22).
- [39] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, “Evaluating analytic gradients on quantum hardware”, *Physical Review A* **99**, 032331 (2019) (cited on pages 22–23).
- [40] J. A. Nelder and R. Mead, “A simplex method for function minimization”, *The computer journal* **7**, 308–313 (1965) (cited on page 23).
- [41] M. J. Powell, “An efficient method for finding the minimum of a function of several variables without calculating derivatives”, *The computer journal* **7**, 155–162 (1964) (cited on page 23).
- [42] M. J. Powell, “A direct search optimization method that models the objective and constraint functions by linear interpolation”, (1994) (cited on page 23).
- [43] J. Spall, “Multivariate stochastic approximation using a simultaneous perturbation gradient approximation”, *IEEE Transactions on Automatic Control* **37**, 332–341 (1992) (cited on page 23).
- [44] J. Platt, “Sequential minimal optimization: a fast algorithm for training support vector machines”, (1998) (cited on page 23).
- [45] K. M. Nakanishi, K. Fujii, and S. Todo, “Sequential minimal optimization for quantum-classical hybrid algorithms”, *Physical Review Research* **2**, 043158 (2020) (cited on page 23).
- [46] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, “A jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization”, (2019) (cited on page 23).
- [47] M. Ostaszewski, E. Grant, and M. Benedetti, “Structure optimization for parameterized quantum circuits”, *Quantum* **5**, 391 (2021) (cited on page 23).
- [48] W. Guan, G. Perdue, A. Pesah, M. Schuld, K. Terashi, S. Vallecorsa, and J.-R. Vlimant, “Quantum Machine Learning in High Energy Physics”, *Machine Learning: Science and Technology* **2**, 011003 (2021) (cited on page 24).
- [49] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, “Data re-uploading for a universal quantum classifier”, *Quantum* **4**, 226 (2020) (cited on pages 24–25).
- [50] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, “Quantum embeddings for machine learning”, Feb. 10, 2020 (cited on pages 24–25).
- [51] M. Schuld, R. Sweke, and J. J. Meyer, “The effect of data encoding on the expressive power of variational quantum machine learning models”, *Physical Review A* **103**, 032430 (2021) (cited on pages 24–25).
- [52] M. Schuld, *Supervised quantum machine learning models are kernel methods*, (Apr. 17, 2021) <http://arxiv.org/abs/2101.11020>, preprint (cited on pages 24–25).
- [53] J. M. Kübler, S. Buchholz, and B. Schölkopf, *The Inductive Bias of Quantum Kernels*, (Nov. 9, 2021) <http://arxiv.org/abs/2106.03747>, preprint (cited on page 26).

- [54] R. Shaydulin and S. M. Wild, “Importance of kernel bandwidth in quantum machine learning”, *Physical Review A* **106**, 10.1103/physreva.106.042407 (2022) (cited on page 26).
- [55] A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, and R. Shaydulin, *Bandwidth Enables Generalization in Quantum Kernel Models*, (June 18, 2023) <http://arxiv.org/abs/2206.06686>, preprint (cited on page 26).
- [56] A. Arrasmith, Z. Holmes, M. Cerezo, and P. J. Coles, “Equivalence of quantum barren plateaus to cost concentration and narrow gorges”, *Quantum Science and Technology* **7**, 045015 (2022) (cited on page 27).
- [57] M. Cerezo and P. J. Coles, “Higher Order Derivatives of Quantum Neural Networks with Barren Plateaus”, *Quantum Science and Technology* **6**, 035006 (2021) (cited on page 28).
- [58] L. Leone, S. F. E. Oliviero, L. Cincio, and M. Cerezo, *On the practical usefulness of the Hardware Efficient Ansatz*, (Nov. 2, 2022) <http://arxiv.org/abs/2211.01477>, preprint (cited on page 28).
- [59] A. Uvarov and J. Biamonte, “On barren plateaus and cost function locality in variational quantum algorithms”, *Journal of Physics A: Mathematical and Theoretical* **54**, 245301 (2021) (cited on page 28).
- [60] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, “Layerwise learning for quantum neural networks”, *Quantum Machine Intelligence* **3**, 5 (2021) (cited on page 30).
- [61] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, “An initialization strategy for addressing barren plateaus in parametrized quantum circuits”, *Quantum* **3**, 214 (2019) (cited on page 30).
- [62] K. Zhang, L. Liu, M.-H. Hsieh, and D. Tao, *Escaping from the Barren Plateau via Gaussian Initializations in Deep Variational Quantum Circuits*, (Dec. 1, 2022) <http://arxiv.org/abs/2203.09376>, preprint (cited on page 30).
- [63] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, “Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices”, *Physical Review X* **10**, 021067 (2020) (cited on page 30).
- [64] E. C. Martín, K. Plekhanov, and M. Lubasch, *Barren plateaus in quantum tensor network optimization*, (Sept. 20, 2022) <http://arxiv.org/abs/2209.00292>, preprint (cited on page 30).
- [65] D. Gross, K. Audenaert, and J. Eisert, “Evenly distributed unitaries: on the structure of unitary designs”, *Journal of Mathematical Physics* **48**, 052104 (2007) (cited on pages 34–35).
- [66] L. Zhang, *Matrix integrals over unitary groups: An application of Schur-Weyl duality*, (Sept. 1, 2015) <http://arxiv.org/abs/1408.3782>, preprint (cited on page 36).

- [67] M. Fukuda, R. Koenig, and I. Nechita, “RTNI - A symbolic integrator for Haar-random tensor networks”, *Journal of Physics A: Mathematical and Theoretical* **52**, 425303 (2019) (cited on page 49).
- [68] D. García-Martín, M. Larocca, and M. Cerezo, *Deep quantum neural networks form Gaussian processes*, (May 17, 2023) preprint (cited on page 54).
- [69] T. Hubregtsen, D. Wierichs, E. Gil-Fuster, P.-J. H. S. Derks, P. K. Faehrmann, and J. J. Meyer, “Training quantum embedding kernels on near-term quantum computers”, *Physical Review A* **106**, 042431 (2022) (cited on page 54).
- [70] G. Li, R. Ye, X. Zhao, and X. Wang, *Concentration of Data Encoding in Parameterized Quantum Circuits*, (June 16, 2022) <http://arxiv.org/abs/2206.08273>, preprint (cited on pages 54–55, 57, 79, 82).
- [71] S. S. Khan and M. G. Madden, “One-class classification: taxonomy of study and review of techniques”, *The Knowledge Engineering Review* **29**, 345–374 (2014) (cited on page 56).

付録A 第3章の補足

A.1 機械学習の概要

機械学習とは、コンピューターに問題を解くためのルールを明示することなく、学習を通して最適なモデルを作る方法である。機械学習の分野は、教師あり学習、教師なし学習、強化学習の3つの大きなカテゴリに分けられる。

- **教師あり学習**とは、与えられた入力値と出力値のペアの集合（訓練データセット）に基づいて、入力値を出力値にマッピングする関数を学習するタスクである。これにより、関数が未知のデータサンプルの出力値を正しく決定できるようにする。
- **教師なし学習**とは、出力値を持たない入力値のみのデータセットからの推論を行うために使用される機械学習のタスクである。最も一般的な教師なし学習の方法はクラスター分析であり、これはデータ内の隠れたパターンやグループを見つけるための探索的データ分析に使用される。
- **強化学習**とは、ソフトウェア上のエージェントが環境との相互作用を通じて学習する機械学習のタスクである。強化学習は、教師あり学習とは異なり、正しい出力が何であるかを明示的に教えられない。代わりに、エージェントは、試行錯誤のプロセスを通じて報酬または罰則を受け取り、その結果、最適な行動を決定するための行動戦略を学習する。

A.1.1 教師あり学習のモデル

教師あり学習の問題を記述するために、入力の集合 X と出力の集合 Y を定義する。また、仮説の集合 H を定義する。 H は X から Y へのすべての関数の集合の部分集合である。 H を仮説空間と呼ぶ。また、 $S = \{(x, y) | x \in X, y \in Y\}$ とし、訓練サンプルの集合 D は S の部分集合であるとする。同様に、テストサンプルの集合 T も S の部分集合であるとする。ただし、 D と T は共通の要素を持たないとする。教師あり学習の目的は、 D に基づいて H の中の最適な仮説 h を見つけることである。最適な仮説 h は、 T に基づいてテストされる。 h が T でうまく機能する場合、 h は未知の入力 x に対してもうまく機能すると予想される。

損失関数 L は、仮説 h と訓練サンプル (x, y) を入力として受け取り、 h が x に対して y を正しく予測しなかった場合のペナルティを計算する関数である。損失関数の単純な例は、次のように定義される二乗誤差である。

$$L(h, (x, y)) = (h(x) - y)^2. \tag{A.1.1}$$

教師あり学習は、訓練データセット D と損失関数 L によって定義される経験損失 $E[L(h, (x, y))]$ を最小化する仮説 h を見つける。

$$E[L(h, (x, y))] = \sum_{(x, y) \in D} L(h, (x, y)). \quad (\text{A.1.2})$$

A.1.2 教師なし学習のモデル

教師なし学習では、空間 X のベクトルとして表されるデータサンプルの集合 D が与えられる。教師なし学習の目的は、データの構造を発見することである。構造は通常、クラスターの集合によって記述される。各クラスターは D の部分集合であり、すべてのクラスターの和集合は D である。教師なし学習の目的は、 D の良い分割であるクラスターの集合 C を見つけることである。良い分割とは、同じクラスター内の任意の 2 点間の距離が小さいことと、異なるクラスター内の任意の 2 点間の距離が大きいことを意味する。コスト関数 J は、クラスターの集合 C を入力として受け取り、クラスターの集合 C が D の良い分割である場合に小さい値を返す関数である。

$$J(C) = \sum_{C_i \in C} \sum_{x, y \in C_i} \|x - y\|^2. \quad (\text{A.1.3})$$

A.1.3 強化学習のモデル

強化学習では、アクションの集合 A 、状態の集合 S 、および報酬関数 R が与えられる。報酬関数は、エージェントが特定の状態 $s \in S$ にいて特定のアクション $a \in A$ を取った場合の報酬を計算する。強化学習の目的は、状態を入力として受け取り、アクションを返す関数である方策 π のなかで、期待される総報酬を最大化する方策を見つけていることである。具体的には、状態価値関数や行動価値関数を用いて、方策を最適化する。

状態価値関数 $V^\pi(s)$ は、最初に状態 s にあり、その後方策 π に従って行動を選択することを繰り返した場合の期待される総報酬である。

$$V^\pi(s) = E_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \mid s_0 = s \right]. \quad (\text{A.1.4})$$

ここで、 $\gamma \in [0, 1]$ は割引率である。

行動価値関数 $Q^\pi(s, a)$ は、最初に状態 s にあり、行動 a を選択し、その後方策 π に従って行動を選択することを繰り返した場合の期待される総報酬である。

$$Q^\pi(s, a) = E_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \mid s_0 = s, a_0 = a \right]. \quad (\text{A.1.5})$$

これらの価値関数に対して、価値反復法や Q 学習などのアルゴリズムを用いて、最適な方策を見つけている。

A.2 ユニタリ 2-デザインにおけるバレンプラト-の定理と証明

先行研究 [23] はバレンプラト-という現象を最初に示した。ここでは、その論文の結果と証明を紹介する。ユニタリ 2-デザインの定義については、4.2 節を参照されたい。

定理 A.2.1. 量子回路 $V(\boldsymbol{\theta})$ は n 個の量子ビットからなり、2つの部分 $V(\boldsymbol{\theta}) = V_+V_-$, $V_+ := \prod_{l=k+1}^L U_l(\theta_l)W_l$, $V_- := \prod_{l=1}^k U_l(\theta_l)W_l$ に分解できるとする。 V_+ , V_- の少なくとも一方がユニタリ 2-デザインであるとき、コスト関数 $C(\boldsymbol{\theta}) = \text{Tr}[V(\boldsymbol{\theta})\rho V^\dagger(\boldsymbol{\theta})O]$ はバレンプラト-となる [23]。

$$\text{Var}_{V(\boldsymbol{\theta})} \left[\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} \right] = \begin{cases} -\frac{1}{4} \frac{\langle \text{Tr}[P_-^2] \rangle_{V_-}}{d^2-1} \left(\text{Tr}[O^2] - \frac{\text{Tr}[O]^2}{d} \right) & \text{if } V_+ \text{ is 2-design} \\ -\frac{1}{4} \frac{\langle \text{Tr}[P_+^2] \rangle_{V_+}}{d^2-1} \left(\text{Tr}[\rho^2] - \frac{1}{d} \right) & \text{if } V_- \text{ is 2-design} \\ \frac{d^2}{2(d^2-1)^2} \left(\text{Tr}[O^2] - \frac{\text{Tr}[O]^2}{d} \right) \left(\text{Tr}[\rho^2] - \frac{1}{d} \right) & \text{if } V_+, V_- \text{ are 2-design} \end{cases} \quad (\text{A.2.1})$$

ここで、 $U_k(\theta_k) = \exp(-\frac{i}{2}\theta_k P_k)$, $(P_k^2 = I, \text{Tr}[P_k] = 0)$, $P_+ := [P_k, V_+^\dagger O V_+]$, $P_- := [P_k, V_- \rho V_-^\dagger]$ とした。

量子回路がユニタリ t -デザインならば、ユニタリ $(t-1)$ -デザインでもあるので、ユニタリ $t \geq 2$ -デザインならば、バレンプラト-が起きる。

証明. $U_k(\theta_k) = \exp(-\frac{i}{2}\theta_k P_k)$ と表され、 $P_k \in \{X, Y, Z\}$ とする。このとき、 $\partial_k V(\boldsymbol{\theta}) = -\frac{i}{2} V_+ P_k V_-$ であるから、

$$\partial_k C(\boldsymbol{\theta}) := \frac{\partial C(\boldsymbol{\theta})}{\partial \theta_k} = \frac{i}{2} \left(\text{Tr} \left[V_+ V_- \rho V_-^\dagger P_k V_+^\dagger O \right] - \text{Tr} \left[V_+ P_k V_- \rho V_-^\dagger V_+^\dagger O \right] \right) \quad (\text{A.2.2})$$

$$= -\frac{i}{2} \text{Tr} \left[V_+^\dagger O V_+ [P_k, V_- \rho V_-^\dagger] \right] \quad (\text{A.2.3})$$

$$= \frac{i}{2} \text{Tr} \left[V_- \rho V_-^\dagger [P_k, V_+^\dagger O V_+] \right] \quad (\text{A.2.4})$$

最後の等式では $\text{Tr}[A[B, C]] = -\text{Tr}[C[B, A]]$ であることを用いた。

まず、勾配の平均が 0 になることを示す。 V_+ がユニタリ 2-デザインであるとする、 V_+ はユニタリ 1-デザインでもあるから、

$$\mathbb{E}_{V_+} [\partial_k C(\boldsymbol{\theta})] = -\frac{i}{2} \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \text{Tr} \left[V_+^\dagger O V_+ [P_k, V_- \rho V_-^\dagger] \right] \quad (\text{A.2.5})$$

$$= -\frac{i}{2} \text{Tr} \left[\left(\int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) V_+^\dagger O V_+ \right) [P_k, V_- \rho V_-^\dagger] \right] \quad (\text{A.2.6})$$

$$= -\frac{i}{2} \text{Tr} \left[\frac{\text{Tr}[\rho] \mathbb{1}}{d} \times [P_k, V_- \rho V_-^\dagger] \right] \quad \because (4.1.7) \quad (\text{A.2.7})$$

$$= 0 \quad (\text{A.2.8})$$

同様に、 V_- がユニタリ 2-デザインであるとする、 V_- はユニタリ 1-デザインでもあるから、

$$\mathbb{E}_{V_-} [\partial_k C(\boldsymbol{\theta})] = \frac{i}{2} \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_-) \text{Tr} \left[V_- \rho V_-^\dagger [P_k, V_+^\dagger O V_+] \right] \quad (\text{A.2.9})$$

$$= \frac{i}{2} \text{Tr} \left[\left(\int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_-) V_- \rho V_-^\dagger \right) [P_k, V_+^\dagger O V_+] \right] \quad (\text{A.2.10})$$

$$= \frac{i}{2} \operatorname{Tr} \left[\frac{\operatorname{Tr}[\rho] \mathbb{1}}{d} \times [P_k, V_+^\dagger O V_+] \right] \quad \because (4.1.7) \quad (\text{A.2.11})$$

$$= 0 \quad (\text{A.2.12})$$

故に、どちらの場合も $E_{V(\boldsymbol{\theta})}[\partial_k C(\boldsymbol{\theta})] = 0$ である。

次に、勾配の分散が量子ビット数に関して指数関数的に小さくなることを示す。 V_- がユニタリ 2-デザインであるとすると、 $P_+ := [P_k, V_+^\dagger O V_+]$ として、

$$\operatorname{Var}_{V_+} [\partial_k C(\boldsymbol{\theta})] = -\frac{1}{4} \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_-) \operatorname{Tr} [V_- \rho V_-^\dagger P_+] \operatorname{Tr} [V_- \rho V_-^\dagger P_+] \quad (\text{A.2.13})$$

公式 (4.1.5) において、 $A = C = \rho$, $B = D = P_+$ とすると、

$$\begin{aligned} \operatorname{Var}_{V_-} [\partial_k C(\boldsymbol{\theta})] &= -\frac{1}{4} \left(\frac{1}{d^2 - 1} (\operatorname{Tr}[\rho] \operatorname{Tr}[P_+] \operatorname{Tr}[\rho] \operatorname{Tr}[P_+] + \operatorname{Tr}[\rho^2] \operatorname{Tr}[P_+^2]) \right. \\ &\quad \left. - \frac{1}{d(d^2 - 1)} (\operatorname{Tr}[\rho^2] \operatorname{Tr}[P_+] \operatorname{Tr}[P_+] + \operatorname{Tr}[\rho] \operatorname{Tr}[\rho] \operatorname{Tr}[P_+^2]) \right) \end{aligned} \quad (\text{A.2.14})$$

$$= -\frac{1}{4} \frac{\operatorname{Tr}[P_+^2]}{d^2 - 1} \left(\operatorname{Tr}[\rho^2] - \frac{1}{d} \right) \quad (\because \operatorname{Tr}[P_+] = 0, \operatorname{Tr}[\rho] = 1) \quad (\text{A.2.15})$$

V_+ がユニタリ 2-デザインとすると、 $P_- := [P_k, V_- \rho V_-^\dagger]$ として、

$$\operatorname{Var}_{V_+} [\partial_k C(\boldsymbol{\theta})] = -\frac{1}{4} \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \operatorname{Tr} [V_+^\dagger O V_+ P_-] \operatorname{Tr} [V_+^\dagger O V_+ P_-] \quad (\text{A.2.16})$$

公式 (4.1.5) を用いて同様に計算すると、

$$\operatorname{Var}_{V_+} [\partial_k C(\boldsymbol{\theta})] = -\frac{1}{4} \frac{\operatorname{Tr}[P_-^2]}{d^2 - 1} \left(\operatorname{Tr}[O^2] - \frac{\operatorname{Tr}[O]^2}{d} \right) \quad (\because \operatorname{Tr}[P_-] = 0) \quad (\text{A.2.17})$$

最後に V_+, V_- の両方がユニタリ 2-デザインであるとする。 $\operatorname{Var}_{V_-} [\partial_k C(\boldsymbol{\theta})]$ を計算したので、あとは、 $\operatorname{Tr}[P_+^2]$ を V_+ で積分すればよい。

$$\int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \operatorname{Tr}[P_+^2] = \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \operatorname{Tr} [[P_k, V_+^\dagger O V_+]^2] \quad (\text{A.2.18})$$

$$\begin{aligned} &= \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \operatorname{Tr} [P_k V_+^\dagger O V_+ P_k V_+^\dagger O V_+ + V_+^\dagger O V_+ P_k V_+^\dagger O V_+ P_k \\ &\quad - P_k V_+^\dagger O V_+ V_+^\dagger O V_+ P_k - V_+^\dagger O V_+ P_k P_k V_+^\dagger O V_+] \end{aligned} \quad (\text{A.2.19})$$

$$= \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) 2 \times \left(\operatorname{Tr} [P_k V_+^\dagger O V_+ P_k V_+^\dagger O V_+] - \operatorname{Tr}[O^2] \right) \quad (\text{A.2.20})$$

$$= \int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) 2 \times \left(\operatorname{Tr} [O V_+ P_k V_+^\dagger O V_+ P_k V_+^\dagger] - \operatorname{Tr}[O^2] \right) \quad (\text{A.2.21})$$

最後から 2 番目の等式では、 $P_k^2 = \mathbb{1}$, $V_+ V_+^\dagger = \mathbb{1}$ を用いた。公式 (4.1.8) において、 $A = C = O$, $B = D = P_k$, $E = \mathbb{1}$ とし、全体のトレースを取ると、

$$\int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \operatorname{Tr} [P_k V_+^\dagger O V_+ P_k V_+^\dagger O V_+] = \frac{1}{d^2 - 1} (\operatorname{Tr}[P_k] \operatorname{Tr}[P_k] \operatorname{Tr}[O^2] + \operatorname{Tr}[P_k^2] \operatorname{Tr}[O] \operatorname{Tr}[O])$$

$$-\frac{1}{d(d^2-1)}(\text{Tr}[P_k^2] \text{Tr}[O^2] + \text{Tr}[P_k] \text{Tr}[O] \text{Tr}[P_k] \text{Tr}[O]) \quad (\text{A.2.22})$$

$$= \frac{1}{d^2-1}(d \text{Tr}[O]^2 - \text{Tr}[O^2]) \quad (\text{A.2.23})$$

最後の等式では、 $\text{Tr}[P_k] = 0$, $\text{Tr}[P_k^2] = \text{Tr}[\mathbb{1}] = d$ を用いた。よって、

$$\int_{\mathcal{U}(d)} d\mu_{\text{Haar}}(V_+) \text{Tr}[P_+^2] = 2 \times \left(\frac{1}{d^2-1}(d \text{Tr}[O]^2 - \text{Tr}[O^2]) - \text{Tr}[O^2] \right) \quad (\text{A.2.24})$$

$$= -\frac{2d^2}{d^2-1} \left(\text{Tr}[O^2] - \frac{\text{Tr}[O]^2}{d} \right) \quad (\text{A.2.25})$$

したがって、

$$\text{Var}_{V(\theta)} [\partial_k C(\theta)] = \frac{d^2}{2(d^2-1)^2} \left(\text{Tr}[O^2] - \frac{\text{Tr}[O]^2}{d} \right) \left(\text{Tr}[\rho^2] - \frac{1}{d} \right) \quad (\text{A.2.26})$$

■

付録B 第5章の補足

B.1 ローカルユニタリ 2-デザイン: ALT の計算例

ここでは、ローカルユニタリ 2-デザインの仮定の下での $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s)$ の計算について、量子ビット数が 6、層数が 3、 $s = 1$ の場合の例を用いて説明する。 $D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^s) = \text{Tr}[(\rho^{(h)})^2] - 1/2^s$ であるから、 $\int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2]$ を計算すれば良い。 $\text{Tr}[(\rho^{(h)})^2]$ をテンソルネットワークで表現すると、図 B.1.1 の一番上のネットワークのようになる。これを、各ゲートブロックがユニタリ 2-デザインを成すとして積分する。以降の図では、 $UU^\dagger = \mathbf{1}$ となって消える部分を灰色で表現している。図の最初の等号では、灰色の部分を消した。2つ目の等号は、 U_1 のみを、図 4.1.6 を用いて積分した。

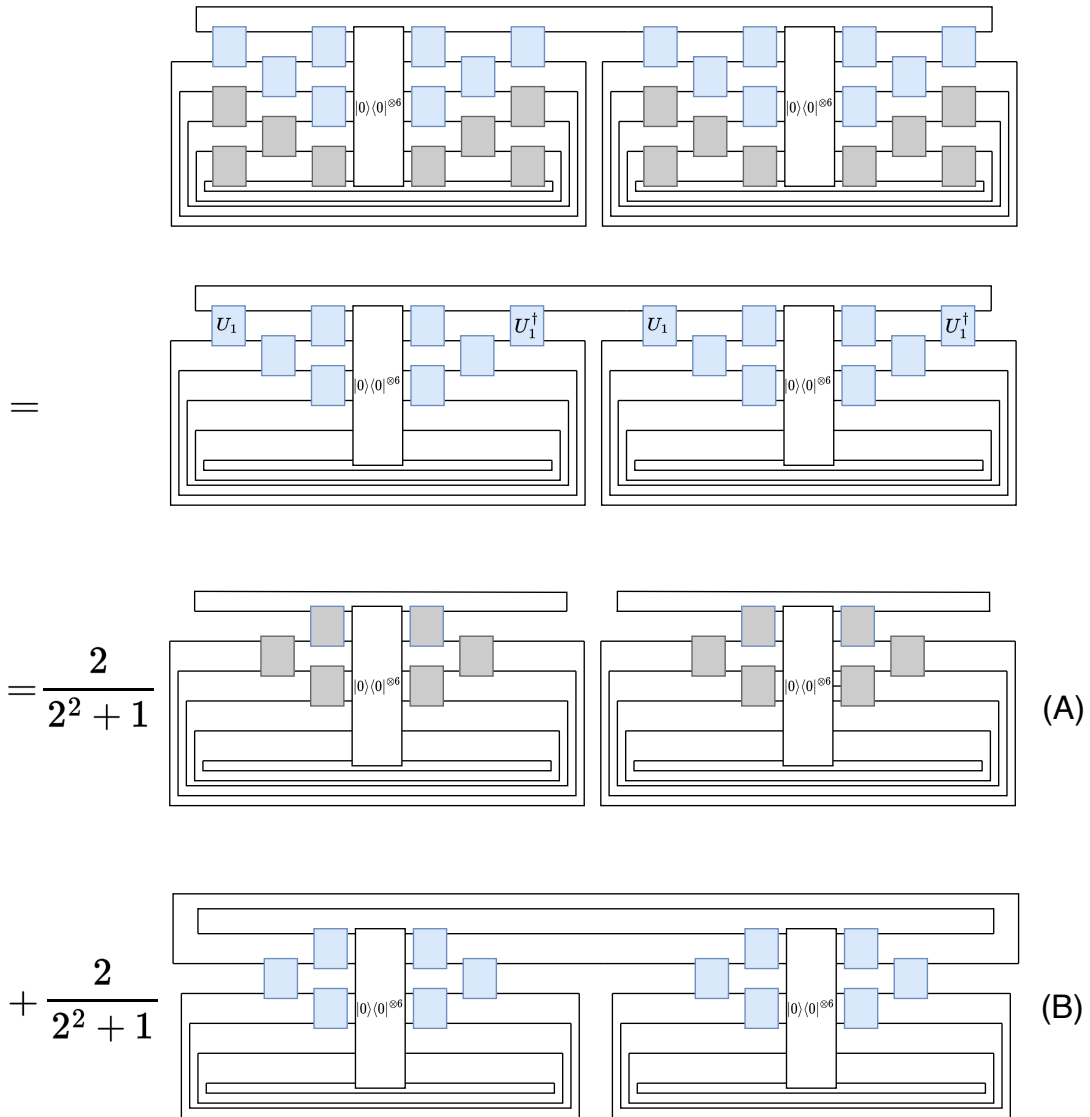


図 B.1.1: 量子ビット数が 6、層数が 3、 $s = 1$ の ALT において、 U_1 を図 4.1.6 を用いて積分

これにより出てきた 2 つの項 (A) と (B) について：まず、項 (A) は $(\text{Tr}[U |0\rangle\langle 0|^{\otimes 6} U^\dagger])^2 = 1$ である。次に項 (B) については、図 B.1.2 の一番上のネットワークのゲートブロック U_2 を積分すると項 (C), (D)

に分解される。

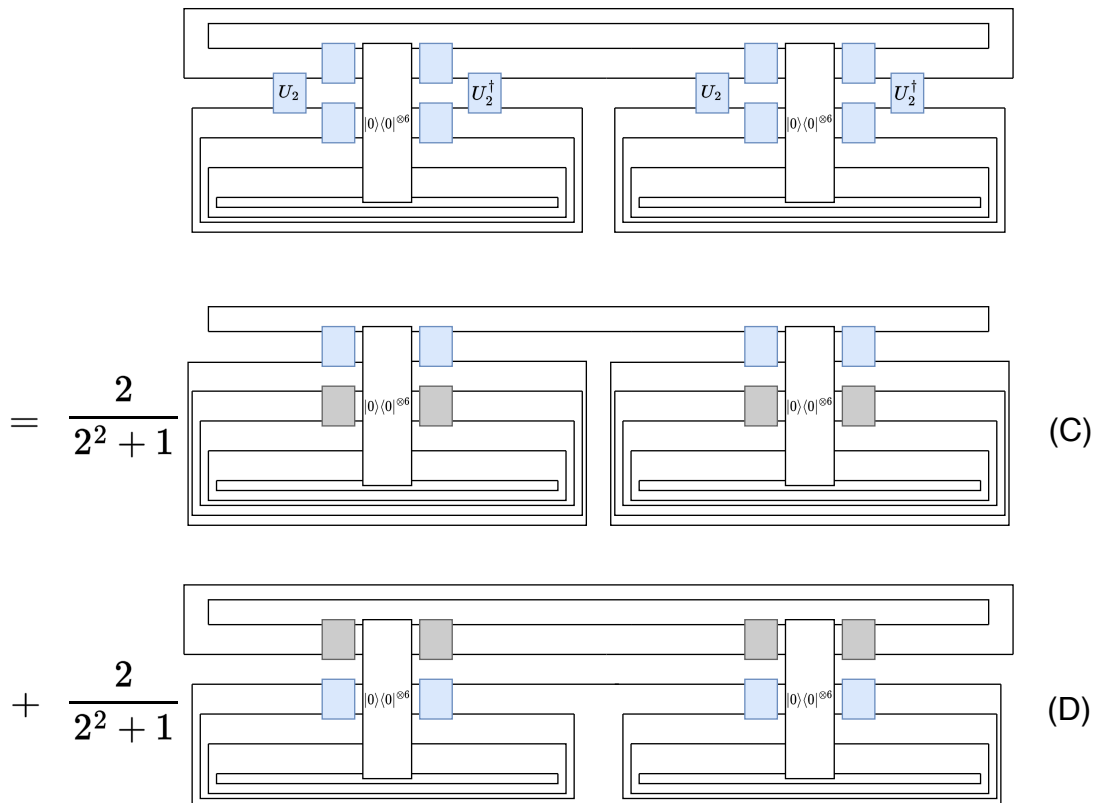
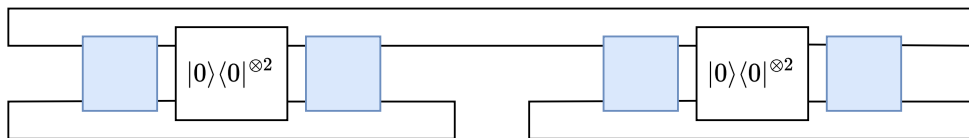


図 B.1.2: 項 (B) について、 U_2 を図 4.1.6 を用いて積分した計算の例。

以上の計算で出てきた、項 (C), (D) は、 $UU^\dagger = 1$, $\text{Tr}[|0\rangle\langle 0|] = 1$ であることを用いると、次の図のようになる。



この計算は、定理 5.2.1 において、 $n = 2, s = 1$ とした場合に相当するので、 $\frac{2^{2-1}+2^1}{2^2+1} = \frac{4}{5}$ となる。したがって、 $\int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2]$ は、

$$\begin{aligned}
 \int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2] &= \frac{2}{2^2+1} \times (\text{A}) + \frac{2}{2^2+1} \times (\text{B}) \\
 &= \frac{2}{2^2+1} \times (\text{A}) + \frac{2}{2^2+1} \left[\frac{2}{2^2+1} \times (\text{C}) + \frac{2}{2^2+1} \times (\text{D}) \right] \\
 &= \frac{2}{2^2+1} \times 1 + \frac{2}{2^2+1} \left[\frac{2}{2^2+1} \times \frac{4}{5} + \frac{2}{2^2+1} \times \frac{4}{5} \right] \\
 &= \frac{82}{125}
 \end{aligned}$$

$\int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2]$ を各量子ビット数 n 、各層数 L について、同様の計算を行った結果が、表 B.1.1 である。これから $\int_{U \in \mathbb{U}} dU D_{\text{HS}}(\rho^{(h)}, \mathbf{1}/2^{s=1}) = \int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2] - 1/2^{s=1}$ を計算してプロットしたものが図 5.4.6 である。

表 B.1.1: 各量子ビット数 n 、各層数 L 、 $s = 1$ の ALT における $\int_{U \in \mathbb{U}} dU \text{Tr}[(\rho^{(h)})^2]$ の計算結果。

$n \setminus L$	2	4	6	8	10	12	14	16
2	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$	$\frac{4}{5}$
4	$\frac{18}{5^2}$	$\frac{394}{5^4}$	$\frac{9402}{5^6}$	$\frac{231466}{5^8}$	$\frac{5757978}{5^{10}}$	$\frac{143720074}{5^{12}}$	$\frac{3591166842}{5^{14}}$	$\frac{89764490986}{5^{16}}$
6	$\frac{18}{5^2}$	$\frac{386}{5^4}$	$\frac{8882}{5^6}$	$\frac{212834}{5^8}$	$\frac{5210258}{5^{10}}$	$\frac{128929346}{5^{12}}$	$\frac{3207308402}{5^{14}}$	$\frac{79991607074}{5^{16}}$
8	$\frac{18}{5^2}$	$\frac{386}{5^4}$	$\frac{8850}{5^6}$	$\frac{210498}{5^8}$	$\frac{5116018}{5^{10}}$	$\frac{125901602}{5^{12}}$	$\frac{3120244306}{5^{14}}$	$\frac{77633338882}{5^{16}}$

付録C 第6章の補足

C.1 定理6.1.1の略証

系 C.1.1. (論文 [27] の Supplementary Note 5, B の拡張された結果) 図 5.1.1 の回路に対して、次の式が成り立つ。

$$\mathbb{E}_{V(\theta)}[\partial_{\nu} \ell_i \cdot \partial_{\nu} \ell_j] = \frac{2^{2s-1} \Delta \Omega}{(2^{2s} - 1)^2} \left(\text{Tr} \left[\rho_i^{(h)} \rho_j^{(h)} \right] - \frac{1}{2^s} \right) \quad (\text{C.1.1})$$

ただし、 $\Delta \Omega = \text{Tr}[(\hat{O}_L^{(h)})^2] - \frac{1}{2^s} (\text{Tr}[\hat{O}_L^{(h)}])^2$

証明. まず、論文 [27] の「Supplementary Note 4: Variance of the cost function partial derivative」の結果を拡張する。具体的には、 $\rho \rightarrow \rho_i$ というように入力状態にインデックスを付け、勾配の共分散を計算する。

定理 C.1.2. (論文 [27] の「Supplementary Note 4: Variance of the cost function partial derivative」の拡張された結果)

$$\mathbb{E}_{V(\theta)}[\partial_{\nu} \ell_i \cdot \partial_{\nu} \ell_j] = \frac{2^{m-1} \text{Tr}[\sigma_{\nu}^2]}{(2^{2m} - 1)^2} \sum_{\substack{pq \\ p'q'}} \left\langle \Delta \Omega_{i;pq}^{j;p'q'} \right\rangle_{V_R} \left\langle \Delta \Psi_{i;pq}^{j;p'q'} \right\rangle_{V_L}, \quad (\text{C.1.2})$$

$$\Delta \Omega_{pq}^{p'q'} := \text{Tr}[\Omega_{qp} \Omega_{p'q'}] - \frac{\text{Tr}[\Omega_{qp}] \text{Tr}[\Omega_{p'q'}]}{2^m},$$

$$\Delta \Psi_{i;pq}^{j;p'q'} := \text{Tr}[\Psi_{i;pq} \Psi_{j;p'q'}] - \frac{\text{Tr}[\Psi_{i;pq}] \text{Tr}[\Psi_{j;p'q'}]}{2^m},$$

$$\Omega_{qp} := \text{Tr}_{\bar{w}} \left[(|p\rangle\langle q| \otimes \mathbb{1}_w) V_R^{\dagger} O V_R \right],$$

$$\Psi_{i;pq} := \text{Tr}_{\bar{w}} \left[(|q\rangle\langle p| \otimes \mathbb{1}_w) V_L \rho_i V_L^{\dagger} \right].$$

この結果を用いて、「Supplementary Note 5: Variance of the cost function partial derivative for a single layer of the Alternating Layered Ansatz (B. Variance of the local cost function partial derivative)」において、 $\Delta \Psi_{i;pq}^{j;p'q'} = \text{Tr}_{\bar{h}}[\rho_i] =: \rho_i^{(h)}$ となることを示せるので、拡張された定理を得る。 ■

C.2 定理6.1.2の証明のための補題

補題 C.2.1. $X \sim \mathcal{N}(\mu, \sigma^2)$ とするとき、

$$\mathbb{E}[\cos(X)] = e^{-\sigma^2/2} \cos(\mu), \quad \mathbb{E}[\sin(X)] = e^{-\sigma^2/2} \sin(\mu) \quad (\text{C.2.1})$$

証明. $X \sim \mathcal{N}(\mu, \sigma^2)$ とするとき、 $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$ であるから、 $X = \mu + \sigma Z$ と表せる。よって、

$$\mathbb{E}[e^{aX}] = \mathbb{E}\left[e^{a(\mu+\sigma Z)}\right] = \mathbb{E}[e^{a\mu} e^{a\sigma Z}] = e^{a\mu} \mathbb{E}[e^{a\sigma Z}] \quad (\text{C.2.2})$$

となる。 $f(a) = \mathbb{E}[e^{aZ}]$ とおくと、 $e^{a\mu} \mathbb{E}[e^{a\sigma Z}] = e^{a\mu} f(a\sigma)$ である。ゆえに、 $f(a) = \mathbb{E}[e^{aZ}]$ を求めれば、 $\mathbb{E}[e^{aX}]$ が分かる。 Z の確率密度関数は $\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ であるから、

$$f(a) = \mathbb{E}[e^{aZ}] = \int_{-\infty}^{\infty} e^{az} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \quad (\text{C.2.3})$$

$$= \exp(a^2/2) \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-(z^2 + 2az + a^2)/2) dz \quad (\text{C.2.4})$$

$$= \exp(a^2/2) \cdot 1 \quad (\text{C.2.5})$$

よって、 $f(a) = \mathbb{E}[e^{aZ}] = \exp(a^2/2)$ となる。ゆえに、

$$\mathbb{E}[e^{aX}] = e^{a\mu} f(a\sigma) = e^{a\mu} \exp(a^2\sigma^2/2) \quad (\text{C.2.6})$$

である。 $a = i$ とおくと、

$$\mathbb{E}[e^{iX}] = e^{i\mu} \mathbb{E}[e^{i\sigma Z}] = e^{-\sigma^2/2} e^{i\mu} \iff \begin{cases} \mathbb{E}[\cos(X)] = e^{-\sigma^2/2} \cos(\mu) \\ \mathbb{E}[\sin(X)] = e^{-\sigma^2/2} \sin(\mu) \end{cases} \quad (\text{C.2.7})$$

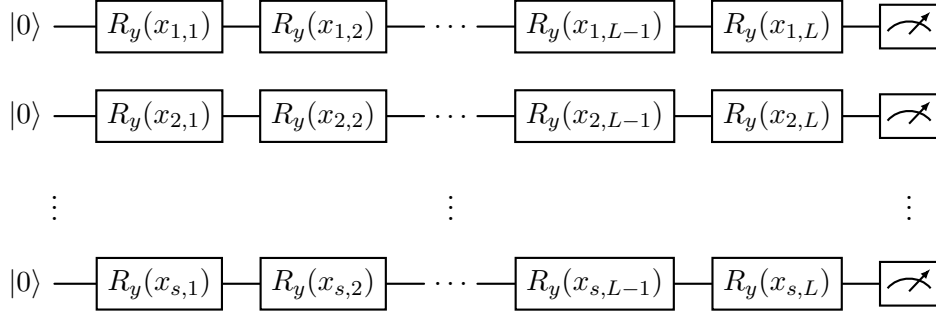
■

C.3 定理 6.1.2 の証明

定理 C.3.1. (論文 [70] の Theorem 1 の拡張された結果) $\mathcal{X} = \{\mathbf{x}\}$, $\mathcal{Z} = \{\mathbf{z}\}$ はそれぞれラベル $y_i = 0$ と $y_i = 1$ に属する入力データセットであり、 $|\mathcal{X}| : |\mathcal{Z}| = p : q$ ($p + q = 1$) であるとする。そして、各データサンプルを次のように表記する： $\mathbf{x} = (\mathbf{x}_j)_{j=1}^s$, $\mathbf{x}_j = (x_{j,d})_{d=1}^L$, $\mathbf{z} = (\mathbf{z}_j)_{j=1}^s$, $\mathbf{z}_j = (z_{j,d})_{d=1}^L$ 。異なるラベルに属する入力データを量子回路に同時には入力しないとする。入力データの各成分の分布は $x_{j,d} \sim \mathcal{N}(\mu_{x|j,d}, \sigma_{x|j,d}^2)$, $z_{j,d} \sim \mathcal{N}(\mu_{z|j,d}, \sigma_{z|j,d}^2)$ であり、分散は $\sigma_{\min} \leq \sigma_{x|j,d}, \sigma_{z|j,d}$ を満たすとする。また、入力回路は次の図 C.3.1 のように構成されているとする。ただし、 $1 \leq j \leq s$, $1 \leq d \leq L$ である。このとき、 $\bar{\rho} = p \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x})] - q \mathbb{E}_{\mathcal{Z}}[\rho(\mathbf{z})]$ について次の不等式が成り立つ。

$$\frac{1}{2^s} \left[w^2 + \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \right] \leq \text{Tr}[(\bar{\rho})^2] \leq \frac{1}{2^s} (1 + e^{-L\sigma_{\min}^2})^s \quad (\text{C.3.1})$$

ただし、 s は量子ビット数、 L は入力回路の層数、 $w = p - q$, $\Sigma_{x|j} := \sum_{d=1}^L \sigma_{x|j,d}^2$, $\Sigma_{z|j} := \sum_{d=1}^L \sigma_{z|j,d}^2$ である。論文内では、 $s = n$, $L = D$ である。


 図 C.3.1: s 量子ビットからなる量子回路。 L 層の R_y ゲートからなる。

証明. 回路の構造上、データ入力後の状態はエンタングルメントを持たないので、 $\rho(\mathbf{x}) = \rho(\mathbf{x}_1) \otimes \rho(\mathbf{x}_2) \otimes \cdots \otimes \rho(\mathbf{x}_s)$ と表せる。また、 R_y のみを用いているから、 $\rho(\mathbf{x}_j) = R_y(x_{j,1} + x_{j,2} + \cdots + x_{j,L}) |0\rangle\langle 0| R_y^\dagger(x_{j,1} + x_{j,2} + \cdots + x_{j,L})$ である。入力データの成分は独立であるから、

$$\mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x})] = \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x}_1)] \otimes \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x}_2)] \otimes \cdots \otimes \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x}_s)] \quad (\text{C.3.2})$$

が成り立つ。 \mathcal{Z} についても同様。 \mathcal{X} と \mathcal{Z} のデータセットのサイズの比率が $|\mathcal{X}| : |\mathcal{Z}| = p : q$ なので、 $\bar{\rho} = \bar{\rho}_+ - \bar{\rho}_-$ は次で与えられる。

$$\bar{\rho} = p \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x})] - q \mathbb{E}_{\mathcal{Z}}[\rho(\mathbf{z})] \quad (\text{C.3.3})$$

$$= p \bigotimes_{j=1}^s \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x}_j)] - q \bigotimes_{j=1}^s \mathbb{E}_{\mathcal{Z}}[\rho(\mathbf{z}_j)] \quad (\text{C.3.4})$$

$$=: p \bigotimes_{j=1}^s \bar{\rho}_{x|j} - q \bigotimes_{j=1}^s \bar{\rho}_{z|j} \quad (\text{C.3.5})$$

よって、 $\text{Tr}[(\bar{\rho})^2]$ は次のように書き換えられる。

$$\text{Tr}[(\bar{\rho})^2] = \text{Tr} \left[\left(p \bigotimes_{j=1}^s \bar{\rho}_{x|j} - q \bigotimes_{j=1}^s \bar{\rho}_{z|j} \right)^2 \right] \quad (\text{C.3.6})$$

$$= p^2 \text{Tr} \left[\left(\bigotimes_{j=1}^s \bar{\rho}_{x|j} \right)^2 \right] + q^2 \text{Tr} \left[\left(\bigotimes_{j=1}^s \bar{\rho}_{z|j} \right)^2 \right] - 2pq \text{Tr} \left[\left(\bigotimes_{j=1}^s \bar{\rho}_{x|j} \bar{\rho}_{z|j} \right) \right] \quad (\text{C.3.7})$$

$$= p^2 \prod_{j=1}^s \text{Tr}[(\bar{\rho}_{x|j})^2] + q^2 \prod_{j=1}^s \text{Tr}[(\bar{\rho}_{z|j})^2] - 2pq \prod_{j=1}^s \text{Tr}[\bar{\rho}_{x|j} \bar{\rho}_{z|j}] \quad (\text{C.3.8})$$

$X_j = \sum_{d=1}^L x_{j,d}$ として $\bar{\rho}_{x|j}$ の行列は次で与えられる。

$$\bar{\rho}_{x|j} = \frac{1}{2} \begin{bmatrix} 1 + \mathbb{E}[\cos(X_j)] & \mathbb{E}[\sin(X_j)] \\ \mathbb{E}[\sin(X_j)] & 1 - \mathbb{E}[\cos(X_j)] \end{bmatrix} \quad (\text{C.3.9})$$

これより、

$$\text{Tr}[(\bar{\rho}_{x|j})^2] = \frac{1}{2} [1 + \mathbb{E}^2[\cos(X_j)] + \mathbb{E}^2[\sin(X_j)]] \quad (\text{C.3.10})$$

ここで、補題 C.2.1 と $X_j \sim \mathcal{N}(\sum_{d=1}^L \mu_{x|j,d}, \sum_{d=1}^L \sigma_{x|j,d}^2)$ であることから、 $\Sigma_{x|j} := \sum_{d=1}^L \sigma_{x|j,d}^2$ とおくと、

$$\begin{aligned} \text{Tr}[(\bar{\rho}_{x|j})^2] &= \frac{1}{2} \left[1 + e^{-\Sigma_{x|j}} \cos^2 \left(\sum_{d=1}^L \mu_{x|j,d} \right) + e^{-\Sigma_{x|j}} \sin^2 \left(\sum_{d=1}^L \mu_{x|j,d} \right) \right] \\ &= \frac{1}{2} (1 + e^{-\Sigma_{x|j}}) \end{aligned} \quad (\text{C.3.11})$$

同様に $Z_j = \sum_{d=1}^L z_{j,d}$, $\Sigma_{z|j} := \sum_{d=1}^L \sigma_{z|j,d}^2$ とすると、 $\text{Tr}[(\bar{\rho}_{z|j})^2]$ は次のようになる。

$$\text{Tr}[(\bar{\rho}_{z|j})^2] = \frac{1}{2} (1 + e^{-\Sigma_{z|j}}) \quad (\text{C.3.12})$$

また、 $\text{Tr}[\bar{\rho}_{x|j} \bar{\rho}_{z|j}]$ は次のようになる。

$$\begin{aligned} \text{Tr}[\bar{\rho}_{x|j} \bar{\rho}_{z|j}] &= \frac{1}{2} [1 + \text{E}[\cos(X_j)] \text{E}[\cos(Z_j)] + \text{E}[\sin(X_j)] \text{E}[\sin(Z_j)]] \\ &= \frac{1}{2} \left[1 + e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2} \cos \left(\sum_j \mu_{x|j,d} - \sum_j \mu_{z|j,d} \right) \right] \end{aligned} \quad (\text{C.3.13})$$

これより、 $\text{Tr}[(\bar{\rho}_{x|j} \bar{\rho}_{z|j})]$ は次の不等式を満たす。

$$\frac{1}{2} \left(1 - e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2} \right) \leq \text{Tr}[(\bar{\rho}_{x|j} \bar{\rho}_{z|j})] \leq \frac{1}{2} \left(1 + e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2} \right) \quad (\text{C.3.14})$$

したがって、 $\text{Tr}[(\bar{\rho})^2]$ は、次の不等式を満たす。

$$\begin{aligned} &\frac{1}{2^s} \left[p^2 \prod_{j=1}^s (1 + e^{-\Sigma_{x|j}}) + q^2 \prod_{j=1}^s (1 + e^{-\Sigma_{z|j}}) - 2pq \prod_{j=1}^s (1 + e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2}) \right] \leq \text{Tr}[(\bar{\rho})^2] \\ &\leq \frac{1}{2^s} \left[p^2 \prod_{j=1}^s (1 + e^{-\Sigma_{x|j}}) + q^2 \prod_{j=1}^s (1 + e^{-\Sigma_{z|j}}) - 2pq \prod_{j=1}^s \left(1 - e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2} \right) \right] \end{aligned} \quad (\text{C.3.15})$$

不等式 (C.3.15) の左辺の展開した各項について次が成り立つ。 ($1 \leq j_p \leq s$)

$$p^2 \cdot 1^{s-t} \prod_{p=1}^t e^{-\Sigma_{x|j_p}} + q^2 \cdot 1^{s-t} \prod_{p=1}^t e^{-\Sigma_{z|j_p}} - 2pq \cdot 1^{s-t} \prod_{p=1}^t e^{-(\Sigma_{x|j_p} + \Sigma_{z|j_p})/2} \quad (\text{C.3.16})$$

$$= \left(p \prod_{p=1}^t e^{-\Sigma_{x|j_p}/2} - q \prod_{p=1}^t e^{-\Sigma_{z|j_p}/2} \right)^2 \geq 0 \quad (\text{C.3.17})$$

この不等式より、不等式 (C.3.15) の左辺は次の式で下から抑えられる。

$$\frac{1}{2^s} \left[p^2 \prod_{j=1}^s (1 + e^{-\Sigma_{x|j}}) + q^2 \prod_{j=1}^s (1 + e^{-\Sigma_{z|j}}) - 2pq \prod_{j=1}^s (1 + e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2}) \right] \quad (\text{C.3.18})$$

$$\begin{aligned} &\geq \frac{1}{2^s} \left[p^2 (1 + e^{-\Sigma_{x|1}} + e^{-\Sigma_{x|2}} + \dots + e^{-\Sigma_{x|s}}) + q^2 (1 + e^{-\Sigma_{z|1}} + e^{-\Sigma_{z|2}} + \dots + e^{-\Sigma_{z|s}}) \right. \\ &\quad \left. - 2pq (1 + e^{-(\Sigma_{x|1} + \Sigma_{z|1})/2} + e^{-(\Sigma_{x|2} + \Sigma_{z|2})/2} + \dots + e^{-(\Sigma_{x|s} + \Sigma_{z|s})/2}) \right] \end{aligned} \quad (\text{C.3.19})$$

$$= \frac{1}{2^s} \left[(p - q)^2 + \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \right] \quad (\text{C.3.20})$$

また、 $L\sigma_{\min} \leq \Sigma_{x|j}, \Sigma_{z|j}$ により、不等式 (C.3.15) の右辺は次の式で上から抑えられる。

$$\frac{1}{2^s} \left[p^2 \prod_{j=1}^s (1 + e^{-\Sigma_{x|j}}) + q^2 \prod_{j=1}^s (1 + e^{-\Sigma_{z|j}}) - 2pq \prod_{j=1}^s (1 - e^{-(\Sigma_{x|j} + \Sigma_{z|j})/2}) \right] \quad (\text{C.3.21})$$

$$\leq \frac{1}{2^s} (1 + e^{-L\sigma_{\min}^2})^s \quad (\text{C.3.22})$$

以上より、 $w = p - q$ として $\text{Tr}[(\bar{\rho})^2]$ は上下から次の式で抑えられる。

$$\frac{1}{2^s} \left[w^2 + \sum_{j=1}^s \left(p e^{-\Sigma_{x|j}/2} - q e^{-\Sigma_{z|j}/2} \right)^2 \right] \leq \text{Tr}[(\bar{\rho})^2] \leq \frac{1}{2^s} (1 + e^{-L\sigma_{\min}^2})^s \quad (\text{C.3.23})$$

■

C.4 定理 6.1.4 の略証

定理 C.4.1. (論文 [70] の Theorem 2 の拡張された結果) $\mathcal{X} = \{\mathbf{x}\}$ はラベル $y_i = 0$ に属する入力データセットであるとする。そして、各データサンプルを次のように表記する： $\mathbf{x} = (\mathbf{x}_j)_{j=1}^s$, $\mathbf{x}_j = (\mathbf{x}_{j,d})_{d=1}^L$, $\mathbf{x}_{j,d} = (x_{j,d,1}, x_{j,d,2}, x_{j,d,3})$ 。入力データの各成分の分布は $x_{j,d,k} \sim \mathcal{N}(\mu_{x|j,d,k}, \sigma_{x|j,d,k}^2)$ であり、分散は $\sigma_{\min} \leq \sigma_{x|j,d,k} \leq \sigma_{\max}$ を満たすとする。また、入力回路は次の図 C.4.1 のように構成されているとする。ただし、 $1 \leq j \leq s, 1 \leq d \leq L, 1 \leq k \leq 3$ である。このとき、 $\bar{\rho} := \mathbb{E}_{\mathcal{X}}[\rho(\mathbf{x})]$ について次の不等式が成り立つ。

$$\frac{1}{2^s} + e^{-3sL\sigma_{\max}^2} \leq \text{Tr}[(\bar{\rho})^2] \leq \frac{1}{2^s} + e^{-L\sigma_{\min}^2} \quad (\text{C.4.1})$$

ただし、 s は量子ビット数、 L は入力回路の層数である。論文内では、 $s = n, L = D$ である。

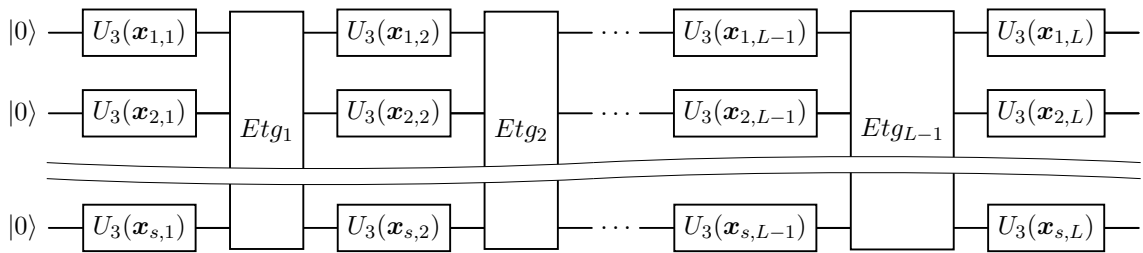


図 C.4.1: s 量子ビットからなる量子回路。入力回路は、データを入力する $U_3(\mathbf{x}_{j,d})$ ゲートの L 層と、CNOT ゲートや CZ ゲートによってエンタングルメントを生成する層 $\text{Etg}_{d'}$ の $L-1$ 層からなる。ただし、 $1 \leq j \leq s, 1 \leq d \leq L, 1 \leq d' \leq L-1$ である。

証明. 論文 [70] 内の式 (S53) における s_m^2 の固有値が $e^{-\sigma_{\min}^2}$ を超えないことと同様に、 $e^{-2\sigma_{\max}^2} \leq A_2^2 A_3^2$ かつ $e^{-\sigma_{\max}^2/2} \leq A_1$ であることから $e^{-3\sigma_{\max}^2}$ を下回らないことがわかる。よって、 $\bigotimes_{j=1}^s T_{j,d}$ の固有値は $e^{-3s\sigma_{\max}^2}$ を下回らない。

$$\text{Tr}[(\bar{\rho})^2] = \frac{1}{2^s} + 2^s \overset{\circ}{\pi}_0 \mathcal{T}_1 \mathcal{E}_1 \cdots \mathcal{T}_{L-1} \mathcal{E}_{L-1} \mathcal{T}_L \mathcal{E}_L \cdots \mathcal{E}_{L-1} \mathcal{T}_{L-1} \cdots \mathcal{E}_1 \mathcal{T}_1 \overset{\circ}{\pi}_0^\top \quad (\text{C.4.2})$$

であり、第二項は次のように上から抑えられる。

$$\overset{\circ}{\pi}_0 \mathcal{T}_1 \mathcal{E}_1 \cdots \mathcal{T}_{L-1} \mathcal{E}_{L-1} \mathcal{T}_L \mathcal{T}_L^\top \mathcal{E}_{L-1}^\top \mathcal{T}_{L-1}^\top \cdots \mathcal{E}_1^\top \mathcal{T}_1^\top \overset{\circ}{\pi}_0^\top \leq \left| \overset{\circ}{\pi}_0 \right|_2^2 e^{-L\sigma_{\min}^2} = \frac{2^s - 1}{2^{2s}} e^{-L\sigma_{\min}^2} \leq \frac{e^{-L\sigma_{\min}^2}}{2^s} \quad (\text{C.4.3})$$

同様に、第二項は次のように下から抑えられる。

$$\frac{e^{-3sL\sigma_{\max}^2}}{2^s} \leq \frac{2^s - 1}{2^{2s}} e^{-3sL\sigma_{\max}^2} \leq \left| \overset{\circ}{\pi}_0 \right|_2^2 e^{-3sL\sigma_{\max}^2} \leq \overset{\circ}{\pi}_0 \mathcal{T}_1 \mathcal{E}_1 \cdots \mathcal{T}_{L-1} \mathcal{E}_{L-1} \mathcal{T}_L \mathcal{T}_L^\top \mathcal{E}_{L-1}^\top \mathcal{T}_{L-1}^\top \cdots \mathcal{E}_1^\top \mathcal{T}_1^\top \overset{\circ}{\pi}_0^\top \quad (\text{C.4.4})$$

ゆえに、拡張された定理を得る。 ■