

機械学習の適用による
量子アルゴリズムの実装性向上に関する研究

2023 年度

手塚宙之

学位論文 博士（工学）

機械学習の適用による
量子アルゴリズムの実装性向上に関する研究

2023 年度

慶應義塾大学大学院理工学研究科

手塚 宙之

要旨

本論文は、全 7 章により構成される。

第 1 章では、研究背景と問題提起、構成について記す。本論文では、量子計算を実行する場合に一般的に直面する課題に着目し、それらの解法の一つを記す。具体的には、データ入力、データ処理アルゴリズムの実装、データ読出し、量子機械学習における学習コスト、ハードウェアヘルスチェックの観点について議論する。

第 2 章では、量子計算の基礎事項について説明する。

第 3 章では、効率的なデータロード手法である「Approximate amplitude encoding (AAE)」について説明する。量子計算における最も重要な問題の一つが、いかに効率的にデータを量子状態に埋め込むかである。既存手法では量子回路サイズのスケールアップに対し、必要な計算ステップが指数関数的に増加してしまう問題を抱えていた。本アルゴリズムでは、生成される状態に有限の誤差を許容できれば、機械学習的な手法を用いることで計算ステップを既存手法に比べて指数関数的に抑えることが可能になる。

第 4 章では、「ブラックボックスの無い Grover's algorithm (GA)」について説明する。GA は量子振幅増幅アルゴリズムとも呼ばれ、量子計算において不可欠なアルゴリズムである。重要な応用として検索タスクが挙げられる。重ね合わせ状態の中からクエリに合致する量子状態のみを抽出し、振幅の増幅を行う。しかし、既存手法ではアルゴリズムの中核にブラックボックス回路を含んでおり、汎用かつ実用的な実装ができなかった。本アルゴリズムでは、第 3 章で説明した AAE を応用することで、このブラックボックスを不要とした。具体的な実装例として、量子画像検索を行い性能評価を行った。

第 5 章では、「NTK を利用した量子-古典ハイブリッドニューラルネットワーク (qcNN)」について説明する。NTK (neural tangent kernel) は、元々は古典機械学習 (ML) の領域で発見され、ノード数が無限大の極限で時不変となる性質をもつ。すなわち、学習過程の解析的評価が可能になる。本章では、これが成立するような qcNN を提案し、理論的な解析により量子優位性を示しうる条件を議論する。さらに数値実験により優位性を示す具体的な実装例を示す。本モデルは、学習対象を意図的に古典 NN 部に限定しており、洗練された古典 ML の手法が適用可能、かつ量子 ML で深刻な問題である勾配消失の心配が無い。一方で、出力関数が projected quantum kernel (PQK) の非線形変換に対応しており、量子 ML の利点と実装性を両立する。

第 6 章では、「最適輸送を用いた量子生成モデル」について説明する。生成モデルの特徴は、所望のクラスに属するが学習データセットに含まれない状態を生成できる点である。古典 ML では画像生成や異常検知で高い性能を発揮している。本章では量子状態を扱う生成モデルを提案する。しかし、既存手法の適用範囲は混合状態の学習に限られ、量子アンサンブルの学習が不可能だった。そこで、個々の純粋状態を保ったまま扱うための損失関数として、量子アンサンブルに対する最適輸送距離を新たに導入し、適切に学習可能なモデルを提案する。本章では、異常検知タスクをモチーフとして、数値実験により提案アルゴリズムの有効性を確認した。

第 7 章では、本論文のまとめと展望を記す。

目次

第 1 章	序章	1
1.1	背景	1
1.2	問題提起	2
1.2.1	データロード・初期状態準備の課題	2
1.2.2	Oracle の実装性の課題	3
1.2.3	効率的なデータ読み出しの課題	3
1.2.4	学習過程の効率化の課題（学習コストの観点）	3
1.2.5	ハードウェアのヘルスチェックの課題（量子異常検知）	4
1.3	課題解決へのアプローチと本論文の構成	4
第 2 章	基本的な事項の説明	6
2.1	量子ビット	6
2.2	量子ビットに対する操作	7
2.3	量子回路	9
2.4	エンコーディング・符号化	9
2.4.1	Basis encoding	10
2.4.2	Amplitude encoding	10
2.4.3	Angle encoding	11
第 3 章	効率的なデータロード手法：Approximate amplitude encoding (AAE)	12
3.1	先行技術の課題：指数関数的なリソース	12
3.2	提案アルゴリズム	13
3.2.1	コンセプト	13
3.2.2	問題設定	13
3.2.3	既存手法の問題点：符号が指定できない	14
3.2.4	提案アルゴリズム	14
3.2.5	エンコーダー $U(\theta)$ の学習方法	15
3.3	結論	17
第 4 章	明示的な Oracle の実装：Grover's algorithm	18
4.1	従来の Grover's algorithm と問題点	18
4.2	提案手法	20
4.2.1	問題設定：量子パターンマッチング	20
4.2.2	提案手法：Black box less Grover's algorithm	20
4.3	実装例：量子画像検索	25

4.3.1	量子状態を用いた画像の表現：Quantum image representation	26
4.3.2	データベースとクエリの量子画像表現（一般系）	27
4.3.3	問題設定	28
4.3.4	実験の詳細	29
4.3.5	結果と考察	30
4.3.6	回路深さと multi qubit gate の数の比較	33
4.3.7	エンコーディング誤差の影響確認	33
4.4	結論	35
第 5 章	学習の効率化：NTK を利用した量子-古典ハイブリッドニューラルネットワーク	37
5.1	先行研究とそれらの問題点	37
5.2	アプローチ：NTK 理論の導入と提案モデルのアウトライン	38
5.3	関連研究	38
5.4	NTK 理論	39
5.4.1	問題設定	39
5.4.2	NTK の定義	40
5.4.3	定理	40
5.4.4	cNN が適切に学習できないケース（本研究の動機）	43
5.5	提案手法	44
5.5.1	提案手法：量子-古典ハイブリッドニューラルネットワーク	45
5.5.2	Quantum neural tangent kernel	47
5.5.3	QNTK に関する定理	47
5.5.4	提案モデルの優位性	49
5.6	数値実験：QNN, CNN との比較	51
5.6.1	幅有限 qcNN と幅無限大 qcNN の比較	51
5.6.2	Locality の機械学習性能への影響	55
5.6.3	純古典モデルと純量子モデルに対する qcNN の優位性	57
5.7	結論	61
第 6 章	効率的な量子異常検知の実装：Quantum generative model with optimal transport	63
6.1	導入	63
6.1.1	Implicit Generative Model	63
6.1.2	最適輸送距離（Optimal Transport Loss; OTL）	64
6.2	提案手法	66
6.2.1	Local ground cost を用いた Optimal transport loss	66
6.2.2	学習アルゴリズム	68
6.3	数値実験：損失関数とその勾配	69
6.3.1	データ数による近似誤差	70
6.3.2	Shot 数による近似誤差	73
6.3.3	勾配消失の回避	76
6.4	デモンストレーション：量子異常検知	77
6.4.1	量子異常検知タスク	77
6.4.2	分散したデータの場合	78

6.4.3	局在したデータの場合	79
6.5	結論	82
第7章	結論と展望	83
7.1	結論	83
7.2	展望	84
謝辞		88
付録		89
A	5.6.3 項の数値実験の詳細	89
B	式 (6.14) のシミュレーション結果	91
C	図 6.4 の直感的な理解	91
D	6.3.3 項のシミュレーション結果	92
参考文献		95

目次

2.1	Bloch 球	7
2.2	量子回路の一例	9
3.1	パラメタ付き量子回路の一例	13
3.2	アルゴリズムの適用イメージ	15
4.1	Grover’s algorithm のイメージ	19
4.2	パターンマッチを行う量子回路の全体像. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	22
4.3	パターンマッチを行う量子回路の全体像の詳細 (n_D, n_I) = (3, 3). Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	26
4.4	Database state (左) と Query state (右) . Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	28
4.5	4 pixel-binary data (“h”は 16 進数の意) . Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	29
4.6	振幅増幅前のパターンマッチの結果. 縦軸は, データベースの index (index 0,1,2,3,4,5,6,7)=(data 0h,2h,4h,6h,8h,Ah,Ch,Eh). 棒グラフは, 観測確率 $P(\text{index} = k) = {}_I\langle k 0^{n_D}\Psi\rangle ^2$ を表す. 色は, 青 (上) がケース (i), 灰色 (中) がケース (ii), 橙 (下) がケース (iii) を表す. “Others”は, 失敗確率, すなわち $ 0\rangle^{\otimes n_D}$ 以外の状態に $ \Psi\rangle$ が射影された確率の合計を表す. “match”は, データベースの中で Query と一致しているものの index を示し, “closest”は, 一致はしていないが Query に最も近いものに対応する index を表す. 図中のインセットは Query データを表す. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	31
4.7	振幅増幅後のパターンマッチの結果 (iteration 数 $t = 5$) . 色は, 青 (上) がケース (i), 橙 (下) がケース (iii) を表す. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	31
4.8	振幅増幅後の観測確率と iteration 数の関係. プロットは数値実験の結果を表し, 破線は式 (4.14) による理論曲線を表す. データベースは図 4.6, Query データは “0h”である. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	32
4.9	振幅増幅後の観測各地ると iteration の関係. Query データが “1h”であり, それ以外の設定は Fig. 4.8 と同様. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	32

4.10	エンコード時にノイズを与えてパターンマッチを行った結果 (FRQI) . 縦軸はクエリ, 横軸は類似度, インセットは正解の数字を表し, 色違いはノイズの大きさの違いに対応している. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	36
4.11	エンコード時にノイズを与えてパターンマッチを行った結果 (NEQR) . 縦軸はクエリ, 横軸は類似度, インセットは正解の数字を表し, 色違いはノイズの大きさの違いに対応している. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.	36
5.1	提案 qcNN モデルの概要. 前半の量子パートはデータ \mathbf{x}^a をエンコードするためのユニタリ演算子 $U_{\text{enc}}(\mathbf{x}^a)$ と, その後に続くランダムユニタリ演算子 U_i , そして Observable O による測定で構成され, 量子状態 $f^Q(\mathbf{x}^a)_i$ の特徴量を抽出する. n_0 個の異なる量子回路により特徴量ベクトル $\mathbf{f}^Q(\mathbf{x}^a) = (f^Q(\mathbf{x}^a)_1, f^Q(\mathbf{x}^a)_2, \dots, f^Q(\mathbf{x}^a)_{n_0})$ を構成し, 後半の古典パートに入力する. 古典パートは n_0 ノードの multi-layered NN で構成される. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	45
5.2	$U_{\text{enc}}(\mathbf{x})$ の構成. 各 qubit に Hadamard gate を作用させた後, RZ-gates の回転角として正規化されたデータ値 x_i ($i = 1, \dots, n$) をエンコードする. さらに, (a) と (c) には entangler として CNOT-gates を作用させる. また, (a) と (b) には RZ-gates の回転角として2つのデータ値の積をエンコードする (表 5.1 中の “Cross-term”) . 尚, RZ(x) の回転角に入力する際の係数が (a) と (b) では $2\pi x$ である点, Ansatz-A4 と Ansatz-A4ne では破線で囲った回路 (shown as “Depth=1”) を4回繰り返す点にも注意されたい. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	53
5.3	回帰タスクにおける学習曲線. コスト関数の時間発展は, 数値シミュレーションは $n_0 = 10^3$, 理論曲線は $n_0 \rightarrow \infty$ で得られたもの. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	54
5.4	分類タスクの学習曲線. それぞれの ansatz に対して同じデータセットを用いた結果. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	55
5.5	テストデータに対する誤答率 ($L = 1, n_0 = 10^3$). 棒グラフは100試行の median を表し, エラーバーの上端 (下端) は最高 (最低) 値を表す. 破線は2クラス分類における random guess の誤答率 50% を示す. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	56
5.6	Heart Disease Data Set の分類タスクの学習曲線 (qcNN の理論曲線) . 量子回路は12-qubit, qcNN の古典パートの層数は $L = 1$. 他の設定は, コスト関数を含め 5.6.1 項と同様. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	57
5.7	5.6.1 項で用いたモデル. (a) 量子データを生成する量子回路. (b) 量子-古典ハイブリッドニューラルネットワーク (qcNN), (c) 量子ニューラルネットワーク (qNN), (d) 古典ニューラルネットワーク (cNN). (c) は $L_q = 10$ とした (10 層の qNN) . 図中の M という箱は測定を表し, $U_3(\alpha, \beta, \gamma)$ は3つの Euler 角に関する single-qubit 回転ゲートである. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	58

5.8	回帰タスクの結果. (a) 学習データ, (b) テストデータ. 縦軸は root mean squared errors (RMSE), 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	60
5.9	分類タスクの結果. (a) 学習データ, (b) テストデータ. 縦軸は正解率, 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	61
5.10	(a) 回帰タスクにおける shot noise の影響 (学習データ). 縦軸は root mean squared errors (RMSE), 横軸は qubit 数. (b) 分類タスクにおける shot noise の影響 (学習データ). 縦軸は正解率, 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	61
6.1	6.3 節と 6.4 節で用いるパラメタ付き量子回路 (ansatz). この ansatz は同じ構造をもつ層を繰り返して構成される. 第 ℓ 層目には, 回転角 $\{\theta_{\ell,j} \cdot z_{\eta_{\ell,j}}\}_{j=1}^n$ で回転方向 $\{\xi_{\ell,j}\}_{j=1}^n$ の single-qubit 回転演算子が各 qubit に作用された後, ladder 構造の controlled-Z gate が作用される. 回転方向 $\xi = \{\xi_{\ell,j}\}_{\ell,j=1}^{N_L \cdot n}$ と潜在変数の index パラメタ $\eta = \{\eta_{\ell,j}\}_{\ell,j=1}^{N_L \cdot n}$ は学習過程の前にランダムに選択され, 学習過程では固定される. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	70
6.2	学習データ数と経験的 OTL の関係 (数値シミュレーション結果). 破線は M^{-1/N_s} に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	72
6.3	フィッティングパラメタ b のシミュレーション結果. (a) qubit 数依存性, (b) 潜在次元依存性. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	73
6.4	測定回数と経験的 OTL の関係 (数値シミュレーション結果). 左列図中の点線は $M^{-1/2}$ に従う関係を表し, 破線は $N_s = 128$ の結果を $\sqrt{c_1 \ln(M) + c_2}$ でフィッティングした結果を表す. 右列図中の点線は $N_s^{-1/2}$ に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	74
6.5	損失関数の勾配と qubit 数の関係 (数値シミュレーション). 図中の点線は M^{-1} に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	76
6.6	分散データ, $n = 2$ における異常検知の結果 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.	79
6.7	分散データ, $n = 10$ における異常検知の結果 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.	80
6.8	学習曲線 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.	81
6.9	局在データ, $n = 6, 10$ における異常検知の結果 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.	81
A	ランダムユニタリ回路. $U_3(\alpha, \beta, \gamma)$ は 3 つのオイラー角による single-qubit rotation gate であり, 角度 α, β, γ は一様分布 $U(0, 2\pi)$ からランダムに選択される. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.	90
B	提案 OTL の近似誤差 (6.14) と学習データ数 M の関係 (数値シミュレーション). プロットは経験的 OTL の数値実験結果, 点線はフィッティング結果. フィッティングによって得られた式 (6.15) のパラメタ b は点線の凡例に記した. Reprinted figure from [46]. Reproduced with permission from Springer Nature.	93

D	qubit 数と local cost に基づく提案 OTL の分散の関係 (数値シミュレーション) . Reprinted figure from [46]. Reproduced with permission from Springer Nature.	94
---	---	----

表目次

2.1	代表的な量子ゲート	8
5.1	$U_{\text{enc}}(\mathbf{x})$ の比較	52
6.1	6.3.1 項の数値シミュレーションにおけるパラメタ設定	71
6.2	6.3.2 項の数値シミュレーションにおけるパラメタ設定	73
6.3	6.3.3 項の数値シミュレーションにおけるパラメタ設定	75
6.4	6.4.2 項の数値シミュレーションにおけるパラメタ設定	79
6.5	6.4.3 項の数値シミュレーションにおけるパラメタ設定	82
A.1	第 5 章 5.6.3 項で用いたモデルパラメタ数	89
A.2	第 5 章 5.6.3 項の数値実験で用いた iteration 数と学習に必要な回路数	89

第 1 章

序章

1.1 背景

近年、従来型のコンピュータ（古典コンピュータ）の計算性能を凌駕しうる次世代の計算機として量子コンピュータが注目されている。量子コンピュータは、量子力学の法則を利用し、古典コンピュータでは実現できない状態表現（重ね合わせや量子もつれ）を活用することで、ある種の問題を効率的に解けることが知られている。歴史的には 1982 年に Richard P. Feynman が、古典コンピュータでは指数関数的な時間がかかる量子系のシミュレーションを量子力学に基づいて動作するコンピュータで効率的に行える可能性を指摘し [1], 1985 年に David Deutsch が量子計算の定式化を行った [2]。それ以降、1994 年の Peter W. Shor による素因数分解の解法（Shor's algorithm [3]; 古典比で指数加速）、1996 年の Lov K. Grover によるデータベース検索（Grover's algorithm [4]; 二乗加速）、2009 年の Harrow らによる線形方程式系の解法（HHL algorithm [5]; 指数加速）など、古典コンピュータに対する優位性が理論的に示されるアルゴリズムが提案された。初期のアルゴリズムは数学的な問題設定の色が濃く抽象度が高かったが、次第に具体的な産業応用へ向けたアルゴリズムが多数提案されるようになり、量子化学 [6, 7], 金融 [8, 9, 10], 機械学習 [11, 12, 13, 14] などへの応用例が示された。

しかし、2023 年現在で前段落で挙げたようなアルゴリズムを実用的な規模でハードウェア実装することはできない。これは、デバイスの有する様々なノイズの影響や扱える量子ビットの数が限られていることが原因である。このようなデバイスを、2018 年に John Preskill は NISQ (Noisy intermediate-scale quantum) [15] と名付けた。NISQ は、エラー訂正を行わずに使用する数百 qubit 程のスケールの量子計算機である。一般的な NISQ デバイス（例えば超電導型）では、single-qubit gate のエラー率は 10^{-4} 程度、double-qubit gate のエラー率は 10^{-3} 程度、読出しエラー率が 10^{-3} 程度である。この値から逆算すると、使用する qubit 数にも依るが、error mitigation や error correction を行わない場合、せいぜい数 100 層（double-qubit gate の層数）程度の量子回路を実行すると、ノイズに埋もれて意味のある計算結果を読み出すことができなくなる。すなわち、エラー訂正が行えないという制約から理論的に優れたアルゴリズムであっても、その実装には大きな制限が課せられる。ハードウェアを用いて量子計算を行うにあたり、特に影響が大きいのが演算処理を行う際に生じるノイズである。この影響を最小限に抑えつつ、量子計算機の利点を活用するために様々な試みがなされている。その中で最も代表的なものが変分量子アルゴリズム（variational quantum algorithm; VQA [16]）である。これは量子計算機と古典計算機のハイブリッドアルゴリズムであり、量子状態がノイズで毀損されるまでの時間で行った処理の結果を古典計算機へ渡し、それに基づいて量子計算機のパラメタ更新を行うということを何度も繰り返すことで最終的な解状態を求めるというフレームワークであり、量子化学計算への応用（variational quantum eigensolver; VQE [17]）をはじめ、多くの応用が研究されている。他にも、ノイズありの測定結果から元の情報を推定するエラー緩和

(error mitigation) の手法 [18], 回路を分解して得られた測定結果を統合することで元回路の情報を復元する手法 (circuit cutting) [19] など, NISQ を用いた実利的な量子計算へ向けて様々な手法が開発されている。また, ハードウェアノイズ以外の問題点への対応や, NISQ のフレームワークにおける産業応用の可能性の探究も広く進められている [20]。

1.2 問題提起

本節では, 量子計算を実際に行うことを考えた際に直面する課題について考える。量子計算を情報処理のツールとしてシンプルに捉えると, ワークフローの全体は「入力」, 「処理」, 「出力」の3要素に分解できる。産業応用において実利が得られるような状況を考えると, これらが一連の流れとして効率的に行える必要がある。前述の通り「処理」の部分に関しては, 古くから学術的な関心も高く, 産学の両者により数多くの研究がなされており, 量子優位性を発揮しうる条件や問題設定, データ構造, そして量子計算が好適な産業領域の候補などが深く検討されている。一方, 「入力」や「出力」に関しては, 「処理」の領域に比べ, その重要度を鑑みても未だに研究が不十分であろう。もちろん, 後述の通り, 様々なアプローチの提案がある。しかし, 未だ決定的なものはない。また, 「処理」の部分に関しても, 理論上は優位性が示されていながら実装性・実現性に乏しいものも存在している。

また, 社会実装までの距離が比較的近い領域として量子機械学習がある。これは量子計算機を特徴量マッピングやサンプラーとして用いることで, 量子特有の性質を活用するものである。しかし, 一般に量子計算では高次元系を扱うことから, 学習の効率性, 特に機械学習では繰り返しパラメータ更新を行う必要があるために, 読出しを含めた効率性が重要になる。逆説的には, そのような高次元系を適切に扱うことのできるモデル・アルゴリズムの開発ができれば, それは優位性に繋がるともいえる。量子計算技術の社会実装に向けた取り組みとして, このような観点も意識する必要がある。

本研究は以上の状況を鑑みて行われた。本研究では, 実装性や効率性を考慮したアルゴリズムを, 主に NISQ に適用可能なものを中心に検討したが, コンセプトや考え方をを用いることで誤り訂正を前提とした量子コンピュータ (Fault Tolerant Quantum Computer; FTQC) における計算効率性の向上にも応用できると考える。実際に量子計算を行う際に問題となる要素は様々あるが, ここでは以下の5点に着目する。

1.2.1 データロード・初期状態準備の課題

量子計算を行うにあたり, 初期化された量子デバイスに対して最初に行うのは, 処理対象のデータの入力・データロードである。入力データが量子状態であれば, 初期状態準備の場合もある。第2章2.4節に記すように, 様々なエンコーディング・符号化手法が開発されている。これらから目的のアルゴリズムに応じて適切なものを選択する必要がある。データエンコードに共通することは, 量子計算を実行するには当然ながら“量子状態”として意図した状態を生成する必要があるという点である。古典計算ではなく量子計算を用いる一つの利点は, その高次元空間を活用することである。高次元空間における状態を適切に制御するためには, それ相応の処理が必要なことは想像に難くない。実際, 汎用的なエンコーディング手法である amplitude encoding を行う場合, qubit 数 n に対して $O(2^n)$ の処理が必要なが分かっている [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]。この問題を回避するために quantum random access memory (QRAM) [33, 34] が仮定されることがある。これは任意の量子状態が $O(n)$ のステップでロードできる装置であるが, 実際に使用できる形での効率的な実装方法は知られていない。仮に実装ができたとしても, NISQ では実用的なスケールでの実装は到底不可能であろう。以上を鑑みるに, NISQ 向けに適用可能で効率的なデータエンコードの手法の必要性は高い。

1.2.2 Oracle の実装性の課題

次に, Oracle と呼ばれるブラックボックス回路について考える. 量子アルゴリズムでは, 入力に対して所望の出力が得られるようなブラックボックス回路を仮定することがある. それらの多くは, 古典計算の要領で四則演算・論理演算を駆使することで実現可能なものであるが, 量子計算機を用いてそれを実装するには非常に多くの計算ステップを要する. すなわち, NISQ ではその実装は事実上不可能であったり, FTQC を想定しても計算コストが重い. 従って, そのようなブラックボックス回路を正攻法とは別のアプローチからの実装方法を検討したり, 何らかの制約の緩和によって実装性を高めることは重要である. そのような方向性の一つとして, 本研究では, 数ある量子アルゴリズムの中でも特に有名かつ重要なアルゴリズムである Grover’s algorithm (GA) に含まれるブラックボックス回路 “Oracle” に着目した. GA は, 重ね合わせ状態にある量子状態の中から特定の量子状態を抽出するためのサブルーチンとして幅広く用いられており, それぞれのタスクに特化した形で Oracle が実装されることはある. しかし, 本研究では検索タスクという汎用的な問題設定に対応する形で明示的に実装可能なアルゴリズムの提案を行う.

1.2.3 効率的なデータ読み出しの課題

データの入力・エンコードと同様に, 計算結果の読み出しも重要である. 量子計算における読み出し手法として最も重要なものの一つが, 量子位相推定 (quantum phase estimation; QPE [35, 36]) である. これは Shor の素因数分解アルゴリズム [3] や線形方程式を効率的に解く HHL アルゴリズム [5] の中核となるサブルーチンである. さらに本質的には, ユニタリ演算における固有値を読み出すアルゴリズムであるため, 量子化学計算における解 (エネルギー準位) の読み出し [6] や量子主成分分析 [11] などにも用いられる. しかし, QPE の実装には Controlled-unitary 演算子や量子フーリエ変換 (quantum Fourier transform; QFT [36]) といった操作が必要になる. これらの実装には多くの gate 操作を必要とするため, 実用的な規模では NISQ での実現は非常に困難である. このような状況下で, NISQ 向けのデータ読み出しのアプローチとして注目されているのが, 乱択測定 (randomized measurement [37]) である. 特に重要なものが “Classical shadow tomography” [38, 39] と呼ばれるアルゴリズムである. このアルゴリズムでは, 同定したい量子状態 ρ に対して, ランダムに選択したユニタリ演算子 U_{rand} を作用させて測定を行い射影 b を得る. 射影 b に対して U_{rand} の逆変換に対応する操作を古典コンピュータ上で行うことで, スナップショット $\hat{\rho}$ と呼ばれる元の量子状態に含まれる情報を部分的に有するものを算出する. このスナップショットを数多く集め, それらを平均化することで元の量子状態 $\rho \simeq \text{mean}[\hat{\rho}_1, \hat{\rho}_2, \dots]$ を推定する. より実用的には, 元の状態 ρ に対する何らかのオブザバブルを効率的に推定することができる. 具体的には, M 個のオブザバブルの推定に必要な測定回数はわずか $O(\log M)$ とされ, NISQ を使用するにあたり非常に強力なツールである.

このように, NISQ で効率的に量子状態を読み出すための研究が広く行われているが, 本研究ではそれをさらに発展させる. 特に, 乱択測定を量子機械学習へ適用することで, その性能や学習効率が向上するようなフレームワークを提案する.

1.2.4 学習過程の効率化の課題 (学習コストの観点)

古典コンピューティングの領域で機械学習 (classical machine learning; CML) が成功を収めていると同様に, 量子コンピューティングの領域でも量子機械学習 (quantum machine learning; QML) が近年注目されている. 尚, QML は勾配法を用いた学習などの CML のテクニックとも親和性が高く, 多くの技術が流用または応用されている. 機械学習では, 一般にモデル構築に大きなコス

トがかかり、このコストを抑えることはアルゴリズムの実装上重要である。特に、量子計算により高次元かつ非線形性の高いデータを扱う場合、モデルの学習に多くのパラメータ更新回数が必要となり、この効率がボトルネックになり得る。まず、量子回路からの出力を用いてパラメータ更新の勾配を求めるとすると、観測値から期待値を推定する方法では、十分な精度を得るには多数の測定回数が必要となる。さらに、QML 特有の問題として知られているものに指数関数的な勾配消失 (barren plateau) の現象もある。加えて懸念されるのは、量子ハードウェアを用いた場合の量子計算機の動作速度である。比較的動作が高速とされる超電導系であっても、現状で 10^3 CLOPS 程度 (CLOPS は 1 秒あたりに実行可能な層数) である。これに加え量子ビットの初期化、読出し、制御用の古典プロセッサとの通信に必要な時間が上乗せされることを考えると、大規模なモデルを学習するには十分に高速とは言えない [40]。以上を踏まえると、実時間で考えた際に優位性が得られるような新たなフレームワークが必要であろう。

1.2.5 ハードウェアのヘルスチェックの課題 (量子異常検知)

最後に、ハードウェアを使用する場合に重要な点に触れておく。シミュレータではなく実デバイスを使用する際には、使用するデバイスが想定した通りに動作することが前提となる。これは実行結果にノイズが含まれることを前提とする NISQ であっても同様である。例えば、我々が日々使用しているパーソナルコンピュータ (PC) でも、RAM (メモリ) やハードディスク/SSD 等のストレージのテストを行うことで、適切な動作の担保や障害の早期検出を行っている。同様に量子計算機が正しく動作していることを担保するためには、簡便なタスクによりデバイスの診断を行う必要がある。すなわち、何か異常が起こった際にはそれを検知する“異常検知”が効率的に行える必要がある。既に量子状態を対象とした異常検知アルゴリズムが、いくつか提案されているが [41, 42]、量子トモグラフィーをベースにした手法であり、異常を診断するために数多くの測定回数が必要である。また、量子計算で想定される高次元系を扱うことを考えると、そのような場合には量子トモグラフィーを用いたアプローチは現実的ではなく、別の方向性からのアプローチが必要である。

1.3 課題解決へのアプローチと本論文の構成

本論文は、全 7 章により構成される。

第 1 章では、本論文の研究背景と問題提起について記した。本論文では、前述の通り、量子計算を実行する場合に直面する課題に着目し、それらの解法を提案する。以降では 1.2 節に記した課題点に順にアプローチしていく。

第 2 章では、量子計算の基礎事項について説明する。

第 3 章では、1.2.1 節を受けて、データロード・初期状態準備の課題について検討する。量子計算における最も重要な問題の一つが、いかに効率的にデータを量子状態に埋め込むかである。既存手法では使用する量子回路のサイズのスケールリングに対し、必要な計算ステップが指数関数的に増加してしまう問題を抱えていた。この問題を解決するための一つのアプローチとして、著者らによって提案された「Approximate amplitude encoding (AAE)」[43] について説明する。既存のデータロードアルゴリズムでは、入力データの厳密な値を量子状態で表現することに注力するために、大きな計算コストが生じていた。そこで本アルゴリズムでは、生成される状態に有限の誤差を許容すること考える。すなわち、機械学習の手法を用いることで、入力データに対応する状態を近似的に生成するユニタリ演算子を構成する。これにより計算ステップを既存手法に比べて指数関数的に抑えることが可能になる。

第 4 章では、1.2.2 節を受けて、Oracle の実装の課題について検討する。特にモチーフとして

Grover's algorithm (GA) に着目する。GA は量子振幅増幅アルゴリズムとも呼ばれ、量子計算において不可欠なアルゴリズムである。重要な応用として探索タスクが挙げられる。このタスクでは、重ね合わせ状態で表現されるデータベース状態の中から、クエリに合致する量子状態のみを抽出し、振幅の増幅を行う。しかし、既存手法ではアルゴリズムの中核にブラックボックス回路を含んでおり、汎用かつ実用的な実装ができなかった。著者らによって提案された本アルゴリズム [44] では、第3章で説明した AAE を応用し、このブラックボックスを不要とした。さらに、データベース状態とクエリ状態の重なりを求め、データ空間の部分的な射影を取ることで、適切に量子データ探索が行えることを示す。具体的な実装例として、量子画像探索による性能評価の結果を示す。

第5章では、1.2.3 節と 1.2.4 節を受けて、効率的なデータ読み出しの課題、学習過程の効率化の課題について検討する。これらの課題を解決しうるフレームワークとして、著者らが提案した「NTK を利用した量子-古典ハイブリッドニューラルネットワーク (qcNN)」 [45] について説明する。NTK (neural tangent kernel) は、元々は古典 ML の領域で発見されたもので、ノード数が無限大の極限で時不変となる性質をもつ。この性質により、学習過程のダイナミクスを記述する方程式が陽に解けるようになり、解析的な評価が可能になる。本章では、これが成立するような qcNN を提案し、理論的な解析により量子優位性を示しうる条件を議論する。さらに数値実験により優位性を示す具体的な実装例を記す。モデルの構成としては、前段の量子回路でデータの特徴量を抽出し、この出力を乱択測定により効率的に古典データとして読み出す。それを後段の古典 NN へ入力することで、非線形変換を行い、機械学習タスクを学習する。また、本モデルは、学習によるパラメタ調整の対象を意図的に古典 NN 部に限定しており、量子 ML で深刻な問題である勾配消失の心配が無い。さらには洗練された古典 ML の手法が適用可能であり、実用上の学習コストは非常に小さくて済む。一方で、モデルからの出力関数が、projected quantum kernel (PQK) の非線形変換となることを示すことができる。PQK は汎化性能がある量子カーネルとして知られており、この結果は量子 ML の利点と実装性の両立を実現しうることを意味する。

第6章では、1.2.5 節を受けて、量子異常検知について検討する。本研究では、古典機械学習の領域で異常検知タスクにおいて高い性能を発揮することが知られている生成モデル (generative model) の考え方を応用することで、効率的な量子異常検知の実現を試みる。戦略として、正常状態に対応する量子状態のアンサンブルをモデルで学習しておき、異常状態に対応する量子状態が入力する際に、何らかのフラグを立てることを考える。しかし、これまでに単一の量子状態を学習する生成モデルの提案は数多くなされているが、量子状態の集合を個別の状態の集合として扱えるものはない。本章では、目的を達するために量子状態に関する新たな最適輸送距離を提案し、これを学習に用いることで効率的にモデルの学習が行なえることを示す ([46] の内容)。まず、理論的考察から、最適輸送距離を用いて学習を行う場合に必要なサンプル数は、システムサイズではなく系の intrinsic dimension にのみ依存することを示す。従って、intrinsic dimension が比較的低いという条件下では、最適輸送距離が高次元系において優れた損失関数になり得ることを示し、これを用いた量子生成モデルを提案する。合わせて、損失関数の local cost 化の手法を導入することで、勾配消失の問題を回避する。続いて、数値実験を用いて行った提案アルゴリズムの性能評価の結果を示す。最後に、実際に量子状態を対象とした異常検知タスクを実装し、数値実験の結果から提案アルゴリズムの有効性を示す。

第7章では、本論文のまとめと展望を記す。

第 2 章

基本的な事項の説明

本章では、量子計算に関する基本的な事項について説明する。より詳細な内容は、量子情報・量子計算の代表的な教科書である [36]などを参照されたい。

2.1 量子ビット

本節では、量子ビットの概念を導入する。古典コンピュータで情報を扱う場合、情報の基本単位はビット (bit) と呼ばれ “0”, “1” の 2 状態のどちらかを取る。一方、量子コンピュータで情報を扱う場合、情報の基本単位は量子ビット (qubit) と呼ばれ、ある量子状態は $|\psi\rangle$ のように表記される。より具体的には、最もシンプルな形として、

$$\begin{aligned} |\psi\rangle &= a_0 |0\rangle + a_1 |1\rangle, & (2.1) \\ a_0, a_1 &\in \mathbb{C}, \quad |a_0|^2 + |a_1|^2 = 1 & (2.2) \end{aligned}$$

と表現される。但し、 $|0\rangle, |1\rangle$ は 2 次元 Hilbert 空間における直交基底であり “計算基底” と呼ばれる。すなわち、式 (2.1) は $|0\rangle, |1\rangle$ という 2 状態が、それぞれ $|a_0|^2, |a_1|^2$ という重みで重ね合わされた状態と見ることができる。さらに、式 (2.2) から実数 θ, ϕ_0, ϕ_1 を用いて、 $a_0 = e^{i\phi_0} \cos \frac{\theta}{2}, a_1 = e^{i\phi_1} \sin \frac{\theta}{2}$ とおけ、これを式 (2.1) に代入することで、

$$\begin{aligned} |\psi\rangle &= a_0 |0\rangle + a_1 |1\rangle \\ &= e^{i\phi_0} \cos \frac{\theta}{2} |0\rangle + e^{i\phi_1} \sin \frac{\theta}{2} |1\rangle \\ &= e^{i\phi_0} \left(\cos \frac{\theta}{2} |0\rangle + e^{i(\phi_1 - \phi_0)} \sin \frac{\theta}{2} |1\rangle \right) & (2.3) \\ &= e^{i\phi_0} \left(\cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \right) \end{aligned}$$

が得られる。但し、 $\phi = \phi_1 - \phi_0$ であり、 $0 \leq \theta \leq \pi, 0 \leq \phi, \phi_0 \leq 2\pi$ である。尚、式 (2.3) の最終行の係数 $e^{i\phi_0}$ は観測に影響しない (ϕ_0 は “global phase” と呼ばれる) ことを示すことができ、多くの場合は省略して考えて問題ない。従って、一般的に 1-qubit の状態は、

$$|\psi\rangle = \cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle$$

と表せる。尚、第二項目の係数 $e^{i\phi}$ に含まれる ϕ は “相対位相 (relative phase)” と呼ばれる。この表記を元に、1-qubit の量子状態を 3 次元空間において図示したものが図 2.1 であり “ブロッホ球 (Bloch sphere)” と呼ばれる。また、ブロッホ球の座標に対応するベクトルを “ブロッホベクトル (Bloch vector)” と呼ぶ。

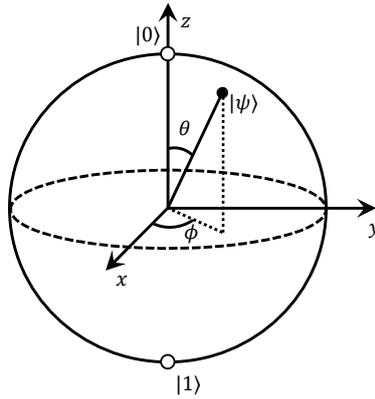


図 2.1: Bloch 球

一般の量子計算では複数の qubit を用いて行うが、各量子ビットが独立に扱える場合は、各状態のテンソル積 (tensor product) として記述される。例えば 2-qubit 系を考えると、

$$\begin{aligned}
 |\varphi\rangle &= |\psi_1\rangle \otimes |\psi_2\rangle = (a_{00}|0\rangle + a_{01}|1\rangle) \otimes (a_{10}|0\rangle + a_{11}|1\rangle) \\
 &= a_{00}a_{10}|00\rangle + a_{00}a_{01}|01\rangle + a_{01}a_{10}|10\rangle + a_{01}a_{11}|11\rangle
 \end{aligned}
 \tag{2.4}$$

という具合である。同様に n -qubit 系を考えると、対応する Hilbert 空間の次元は 2^n となり、qubit 数に対して指数関数的に巨大な状態を扱えることになる。これは n -bit のメモリ空間で n 個のバイナリデータしか扱えない古典コンピュータとは対照的であり、量子計算の優位性の一つの源泉であると言える (量子並列性)。多 qubit 系では、その様子をブロッホ球のように図示することは困難だが、量子状態を興味のある 2 基底に射影することで概要を捉えることができる。

2.2 量子ビットに対する操作

本節では、2.1 節で導入した量子ビットを用いて量子計算を行うにあたり必要となる基本的な操作について説明する。まず、扱いを簡単にするために量子状態をベクトル形式で扱うことを考える。例えば、式 (2.1) を $|\psi\rangle = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$ と記述する。すると、この状態に作用する演算子は 2×2 の複素ユニタリ行列で記述できる。例えば、古典計算で NOT 演算に対応するような操作は、

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

で与えられる。すなわち、

$$X|\psi\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}$$

のように入力状態に作用して $|0\rangle$ と $|1\rangle$ が反転する。尚、量子ビットに作用する演算子を“ゲート (gate)”と呼ぶことがある。例えば、上記の例では、 X ゲート (X gate) という具合である。また、1-qubit だけでなく複数の qubit に作用する操作も同様に定義することができる。特に重要なものとして CNOT (Controlled-not) gate がある。これは 2-qubit gate であり、

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

で与えられる. 式 (2.4) に作用させてみると,

$$\text{CNOT}|\varphi\rangle = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a_{00} \\ a_{01} \\ a_{10} \\ a_{11} \end{pmatrix} = \begin{pmatrix} a_{00} \\ a_{01} \\ a_{11} \\ a_{10} \end{pmatrix}$$

が得られる. すなわち, 第 1qubit が “1” の場合のみ, 第 2qubit を反転するという作用が得られた. このような作用のため, このゲートは CX ゲートとも呼ばれる. 制御ゲートは 2.3 節で示すように量子もつれ (エンタングルメント; entanglement) を生成するために必要なゲートであり, 量子計算特有の情報処理を行ううえで不可欠な操作である. 表 2.1 に代表的な量子ゲートを記した.

表 2.1: 代表的な量子ゲート

ゲート	回路表現	行列表現	意味	ゲート	回路表現	行列表現	意味
X		$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	x -軸周りの反転	$R_x(\theta)$		$\begin{pmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}$	x -軸周りの θ 回転
Y		$\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$	y -軸周りの反転	$R_y(\theta)$		$\begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}$	y -軸周りの θ 回転
Z		$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	z -軸周りの反転	$R_z(\theta)$		$\begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix}$	z -軸周りの θ 回転
H		$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	xz -軸周りの変換	CX		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	制御 X
T		$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\pi/4} \end{pmatrix}$	z -軸周りの $\pi/4$ 回転	CZ		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$	制御 Z

本節の最後に, 測定について説明する. 測定は何らかの基底への射影として扱われる. ここでは例えば, 計算基底への射影として $M_0 = |0\rangle\langle 0|$ を考える. 但し, $\langle \cdot |$ は $\langle \cdot | = (|\cdot\rangle)^\dagger$ で定義される. 尚, \dagger はエルミート共役を表し, 一般に $|\cdot\rangle$ は “ケット”, $\langle \cdot |$ は “ブラ” と呼ばれる. これを用いて式 (2.1) の状態を射影した際に “0” が測定される確率 P_0 は,

$$P_0 = \langle \psi | M_0 | \psi \rangle = |\langle 0 | \psi \rangle|^2 = |a_0|^2$$

と計算される. また, 観測量は量子状態を密度演算子 $\rho = |\psi\rangle\langle \psi|$ で表し, “オブザバブル (observable)” と呼ばれるエルミート演算子を用いて表現されることもある. 例えばオブザバブル O として Z 演算子をとると, 測定結果の期待値 $\langle O \rangle$ は,

$$\langle O \rangle = \text{Tr}\{\rho Z\} = \text{Tr} \begin{pmatrix} a_0^2 & -a_0 a_1 \\ a_0 a_1 & -a_1^2 \end{pmatrix} = a_0^2 - a_1^2$$

と計算される.

2.3 量子回路

本節では、量子操作の表現としての“量子回路 (quantum circuit)”について簡単に記す。量子計算や量子アルゴリズムを視覚的に理解するための方法として、量子回路と呼ばれている表現を用いる。量子回路では、以下のように qubit に施す量子ゲートを順に配置することで量子演算を表現する。利点としては、直感的にアルゴリズムの全体像が把握しやすい点、実装可能性が確認しやすい点などが挙げられるが、欠点として、qubit 数が多くなると記述が煩雑・困難になる点がある。例として、図 2.2 に 3qubit 系の量子回路 (GHZ-state 生成回路) を示した。この図の意味するところは、以下の通りである。

- (基本規則) 横線の数 qubit 数に対応している。
 - (基本規則) 左から右に順に操作を施していく。
1. まず、第 1qubit に Hadamard gate を作用。
 2. 次に、第 1, 第 2qubit 間に CX gate を作用。但し、control bit は第 1, target bit は第 2qubit。
 3. 次に、第 2, 第 3qubit 間に CX gate を作用。但し、control bit は第 2, target bit は第 3qubit。
 4. 最後に 3 つ全ての qubit を計算基底で測定。

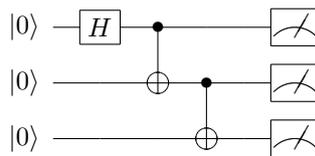


図 2.2: 量子回路の一例

また、状態遷移を量子状態で表現すると以下の通りである。

$$|000\rangle \xrightarrow{1} \frac{|0\rangle + |1\rangle}{\sqrt{2}} |00\rangle \xrightarrow{2} \frac{|00\rangle + |11\rangle}{\sqrt{2}} |0\rangle \xrightarrow{3} \frac{|000\rangle + |111\rangle}{\sqrt{2}}$$

ここで、最右辺の状態がいかなるテンソル積にも分解できないことを指摘しておく。これが量子もつれ (entanglement) なる状態に対応している。図 2.2 を参照すると、この起源が CX gate (制御ゲート) にあることが分かるだろう。

2.4 エンコーディング・符号化

本節では、量子計算で用いられる代表的なエンコーディング・符号化法について記す。所望の量子計算を行うには、初期化された量子ビットに対し、扱いたいデータを入力する必要がある。データの入力にあたっては、量子状態のどのような自由度を用いて情報を表現するかという観点で、様々なデータの埋め込み方が可能である。量子計算では、量子ビットの有する自由度、具体的には基底、振幅、位相といったものに対し、何らかの規則に基づいて符号化することで情報を表現する。問題設定によっては、量子状態をそのまま入力として扱うことあるが、ここでは古典情報を量子状態に埋め込むことを考える。具体的には、 M 個の N 次元ベクトル \mathbf{x} で表現される入力状態 $\mathcal{D} = \{\mathbf{x}^1, \dots, \mathbf{x}^M\}$ を n -qubit に埋め込むことを考える。

2.4.1 Basis encoding

古典バイナリデータの符号化を考える際に、最も直感的で直接的な符号化手法が Basis encoding (BE) である。Basis encoding では、バイナリ列を計算基底に対応させて表現する。例えば、 n -bit の古典ビット列 (0011) は n -qubit の量子ビット列 $|3\rangle = |0011\rangle$ と表現するという具合である。一般化すると、

$$|x\rangle = |b_n \cdots b_1\rangle \leftrightarrow x = (b_n \cdots b_1)$$

と記せる。尚、実数を表現する場合にも、古典コンピュータで用いられているのと同様にバイナリ分数表現を用いれば表現可能である。これを踏まえて先に挙げた入力状態 \mathcal{D} (簡単のため、 $N = 1$ とする) を重ね合わせ状態として表現すると、

$$|\mathcal{D}\rangle = \frac{1}{\sqrt{M}} \sum_{m=1}^M |x^m\rangle$$

となる。但し、 M 個の要素は全て non-zero とした。Basis encoding の利点は、ヒルベルト空間の次元 2^n に比べて埋め込まれているデータの数 M が小さい場合には、その振幅が大きくノイズ耐性が高いという点が挙げられる。一方で、同じ数のデータを表現するのに必要な qubit の数が、他の符号化手法に比べて多いという欠点もある。

2.4.2 Amplitude encoding

Amplitude encoding は、実数データを量子状態の基底の振幅に対応させて表現する手法である。ボルン解釈に基づくと、量子状態の振幅の 2 乗が対応する基底の観測確率に対応するため、積分して 1 になる必要がある。すなわち、手元にある実数データをそのまま埋め込めることは稀であり、規格化の手間が必要となる。しかしながら、HHL algorithm (線形方程式を指数加速するアルゴリズム) [5] や Grover's algorithm [4] のような有望なアルゴリズムを始めとする多くのアルゴリズムで用いられている重要な符号化手法である。規格化後の古典ベクトル $\mathbf{x} \in \mathbb{C}^{2^n}$, $\sum_k |x_k|^2 = 1$ は、量子状態 $|\psi\rangle \in \mathcal{H}$ を用いて、

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{2^n} \end{pmatrix} \leftrightarrow |\psi_x\rangle = \sum_{j=1}^{2^n} x_j |j\rangle$$

と表せる。同様の考え方で先に挙げた入力状態 \mathcal{D} のようなベクトル列、もしくは行列 $A \in \mathbb{C}^{2^n \times 2^m}$ は、

$$\begin{aligned} |\psi_{\mathcal{D}}\rangle &= \sum_{j=1}^M \sum_{i=1}^N a_{ij} |i\rangle |j\rangle \\ &= \sum_{j=1}^M |\psi_j\rangle |j\rangle \end{aligned}$$

と表せる。但し、 $N = 2^n, M = 2^m, \sum_{ij} |a_{ij}|^2 = 1$ とした。Amplitude encoding の利点は、BE とは異なり表現できる精度が使用する qubit 数に依存しない点であり、BE と比較して同じ qubit 数で多くの情報を埋め込むことができる。一方で、振幅というアナログ量を用いて情報を表現するため比較

的ノイズ耐性が低く、状態測定によって精度良く振幅を同定するには、多くの測定回数が必要である。量子状態の振幅を効率的に推定するための研究も盛んになされており、[47]をはじめとする様々な量子振幅推定アルゴリズムが提案されている。

2.4.3 Angle encoding

ここまでの2つの符号化手法は、計算基底や確率振幅という量子状態のもつ自由度に対して、陽に古典情報を埋め込むものだった。これに対し Angle encoding は、ある量子状態に対する回転操作の角度として情報を埋め込む手法である。例えば、1-qubit の初期状態 $|\psi_{\text{init}}\rangle$ に対して、回転操作として $R_z(\theta_2)R_y(\theta_1)$ を選択し、古典ベクトル $\mathbf{x} \in \mathbb{C}^2$ を埋め込む場合、符号化後の状態 $|\psi_{\mathbf{x}}\rangle$ は、

$$\begin{aligned} |\psi_{\mathbf{x}}\rangle &= R_z(x_2)R_y(x_1) |\psi_{\text{init}}\rangle \\ &= U(\mathbf{x}) |\psi_{\text{init}}\rangle \end{aligned}$$

という状態で表しうる。但し、 $U(\mathbf{x})$ はエンコーダー（符号化器）に対応するユニタリー演算子である。このように Angle encoding では、入力データに対応する符号化後の量子状態を規定するのではなく、エンコーダーにより符号化後の量子状態が規定される。これは見方を変えると、エンコーダーを設計することで古典データから量子状態への写像を設計していることになる。また、上の例では、例えば $R_z(\theta_2)R_y(\theta_1)$ へ \mathbf{x} を入力する際に、添字が対応するよう $(\theta_1, \theta_2) = (x_1, x_2)$ と代入を行ったが、代わりに何らかの関数を介し $(\theta_1, \theta_2) = (f(x_1, x_2), g(x_1, x_2))$ とすることで、また別の写像に対応させることもできる。このようなエンコーダーの設計自由度の高さもあり、ヒルベルト空間におけるデータ分布の幾何構造が、アルゴリズムの性能に直結する量子機械学習の領域では、Angle encoding が広く使われている。

他にも目的に応じた符号化方法が各種提案されているが、本論文では説明を割愛する。尚、符号化に関して俯瞰的に整理した教科書として [48] がある。

第3章

効率的なデータロード手法：Approximate amplitude encoding (AAE)

第3章では、NISQを前提とした効率的な振幅エンコーディングの手法について記す。行いたいアルゴリズムに応じて適した符号化手法を考えうが、最も汎用的に用いられるものの一つが、振幅エンコーディング (amplitude encoding; AE) である。筆者らは、このAEに着目し、従来のデータ埋め込み手法に比べて、必要な量子回路の深さを指数的に削減した手法を [49] で提案した。本章では、振幅エンコーディングに関する先行技術とその問題点について記した後、提案アルゴリズムの詳細を説明する。

3.1 先行技術の課題：指数関数的なリソース

これまで、振幅エンコーディングを実現するための数多くのアルゴリズムが提案されてきた。最もナイーブなアプローチとして、multi-controlled rotation gate を使うものがある。これは対応する基底を一つずつ選択しながら、予め計算しておいた所望の振幅に対応する回転軸と回転角を入力することで、目的の量子状態を得る手法である。しかし、この手法を用いて n -qubit に対してデータを埋め込むことを考えると、明らかに $O(2^n)$ の深さの演算が必要なことが分かる。従って、仮にデータを埋め込んだ後のアルゴリズム部分で、古典手法と比べて指数倍の量子加速が得られたとしても、その恩恵がエンコード部分で毀損してしまう。そこでこの問題を回避すべく、別アプローチ (ancillary qubit を導入するなど) の手法も開発されたが、最悪ケースで今度は ancillary qubit が指数個必要になることが知られている [50, 51, 52, 53]。更に別のアプローチとしては、Black-box oracle を用いるものも提案されている [54, 55, 56, 57, 58]。しかしながら、これらを用いても高い精度保証を備えた厳密なデータ埋め込みを行おうとすると、非常に大きな量子回路が必要となるため、実装性に大きな困難が伴うことが知られている。

ここで少し考え方を考えてみる。以上のアルゴリズムでは、厳密な値の埋め込みを目指していた。しかし、世の中の様々な問題に目を向けると、必ずしも非常に高い有効数字が必要な計算ばかりではない。(例えば、多くの数値計算では有効数字が3桁もあれば十分であったり、経済動向を表す金融指標は大まかな変動の傾向で目的が達せられるなど。) そこで我々は、埋め込まれるデータに有限の誤差を許容することで、エンコーダの深さを大幅に抑える手法を開発した。

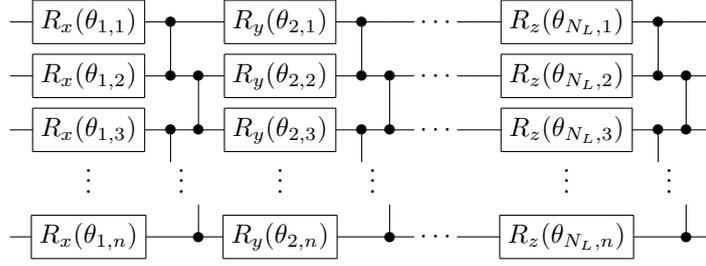


図 3.1: パラメタ付き量子回路の一例

3.2 提案アルゴリズム

3.2.1 コンセプト

提案アルゴリズムでは、例えば、図 3.1 に表されるようなパラメタ付き量子回路 (Parameterized quantum circuit; PQC) を用いて、所望の値を厳密に量子データとして入力することを考えると深くなりがちなエンコーダを浅い量子回路で近似的に表現することを考える。特に、エンコーダの近似を行うために量子機械学習の手法を用いて、PQC のパラメタを求めるという戦略をとる。これはユーザの制約 (許される回路深さ, 求められる精度など) に応じて PQC を構成しつつ、変分法を用いてエンコーダの構築が可能であり、NISQ に適した手法であると言える。

3.2.2 問題設定

何らかの量子アルゴリズムを実行することを想定し、 N 次元の実ベクトル \mathbf{d} が手元にあるとする。これを n -qubit に振幅エンコーディングし、状態 $|\text{Data}\rangle$ を生成する。具体的にはこの状態は、

$$|\text{Data}\rangle = \sum_{j=0}^{N-1} \mathbf{d}_j |j\rangle \quad (3.1)$$

と表される。但し、 j は n -qubit における計算基底に対応、 \mathbf{d}_j は \mathbf{d} の j 番目の成分を表し、 $N = 2^n$ 、 $\sum_j \mathbf{d}_j^2 = 1$ を満たすとする。ここで我々の目標は、 $U(\boldsymbol{\theta}) |0\rangle^{\otimes n} = \sum_{j=0}^{N-1} \mathbf{a}_j |j\rangle$ なるユニタリ回路 $U(\boldsymbol{\theta})$ の下で

$$U(\boldsymbol{\theta}) |0\rangle^{\otimes n} = |\text{Data}\rangle \quad (3.2)$$

すなわち

$$\sum_{j=0}^{N-1} \mathbf{a}_j |j\rangle = \sum_{j=0}^{N-1} \mathbf{d}_j |j\rangle \quad (3.3)$$

の条件を満たすパラメタベクトル $\boldsymbol{\theta}$ を求めることである。(但し、後述する通り、一般の状態を考える際には式 (3.2) は多少拡張して考える必要がある。) この問題のポイントは、式 (3.3) の各要素は絶対値だけでなく符号まで一致させる必要がある点である。実際、[59] では Quantum generative adversarial network (QGAN) を用いることで、機械学習的に古典データの係数の絶対値をエンコードする手法を提案しているが、それでは本条件を満たすことはできない。

3.2.3 既存手法の問題点：符号が指定できない

[59] では, PQC からの出力 $U(\boldsymbol{\theta})|0\rangle$ を計算基底で観測した際の確率分布が所望の分布になるように $U(\boldsymbol{\theta})$ を学習させている. すなわち, 学習結果は以下の関係式を満たすことになる.

$$\begin{aligned} |\mathbf{a}_j|^2 &= |\langle j|U(\boldsymbol{\theta})|0\rangle|^2 = |\mathbf{d}_j|^2, \forall j \in [0, 1, \dots, N-1] \\ \therefore \mathbf{a}_j &= \pm \mathbf{d}_j \end{aligned}$$

つまり, 単純に観測結果だけを用いて学習を行うだけでは, 振幅の符号を指定できないことを表している. [59] では, 式 (3.1) の各測定基底が観測される確率 $|\mathbf{d}_j|^2$ にのみ興味があったため問題がなかったが, 汎用的な符号化手法を構築するには符号の一致まで保証する必要がある.

3.2.4 提案アルゴリズム

そこで我々は「計算基底だけでなく, それに直交する基底への射影も合わせて拘束条件に考慮する」ことで, 上述の問題を回避する手法を提案した. 一般的な状況を想定すると, 以下の2つのケースに分けて考える必要がある.

Case 1) \mathbf{d} の全要素が全て non-positive もしくは non-negative の場合

PQC のパラメタを学習する際の拘束条件として, 以下の2条件を用いる.

$$\begin{aligned} |\langle j|U(\boldsymbol{\theta})|0\rangle^{\otimes n}|^2 &= |\mathbf{d}_j|^2 \quad (\forall j) \\ |\langle j|H^{\otimes n}U(\boldsymbol{\theta})|0\rangle^{\otimes n}|^2 &= \left(\sum_{k=0}^{N-1} \mathbf{d}_j \langle k|H^{\otimes n}|k\rangle \right)^2 \\ &\equiv |\mathbf{d}_j^H|^2 \quad (\forall j) \end{aligned}$$

ここで \mathbf{d}_j^H は Walsh-Hadamard 変換 [60] により $O(N \log N)$ の計算量で効率的に古典計算できることを指摘しておく. 特に \mathbf{d} がスパースな場合, この計算量はさらに圧縮でき, $K = N^\alpha$ ($0 < \alpha < 1$) の下で $O(K \log K \log(N/K))$ となる [61]. 尚, この値は N が増加するに伴い漸近的に 1 へ近づく. このような拘束条件の下で学習された PQC を用いてデータを量子状態へ埋め込み, その後所望の量子アルゴリズムを実行する. (図 3.2 (a))

Case 2) Case 1 以外の場合 (正負が混じっている場合)

まず, 各振幅の符号に着目することで $|\text{Data}\rangle$ を以下のように分解する.

$$|\text{Data}\rangle = |\text{Data}^+\rangle + |\text{Data}^-\rangle \quad (3.4)$$

尚, $|\text{Data}^+\rangle$ は正のもの, $|\text{Data}^-\rangle$ は正以外のものを表す. ここで ancillary qubit を導入し,

$$|\bar{\psi}\rangle \equiv |\text{Data}^+\rangle|0\rangle - |\text{Data}^-\rangle|1\rangle$$

の状態を考えることで全ての振幅を non-negative に直すことができる. ここで計算基底を用いて $|\bar{\psi}\rangle = \sigma_{j=0}^{2N-1} \bar{\mathbf{d}}_j |j\rangle$ と表すことにすると, Case 1 の拘束条件は以下のように書き直せる.

$$\begin{aligned} |\langle j|U(\boldsymbol{\theta})|0\rangle^{\otimes n+1}|^2 &= |\bar{\mathbf{d}}_j|^2 \quad (\forall j) \\ |\langle j|H^{\otimes n+1}U(\boldsymbol{\theta})|0\rangle^{\otimes n+1}|^2 &= \left(\sum_{k=0}^{N-1} \bar{\mathbf{d}}_j \langle k|H^{\otimes n+1}|k\rangle \right)^2 \\ &\equiv |\bar{\mathbf{d}}_j^H|^2 \quad (\forall j) \end{aligned}$$

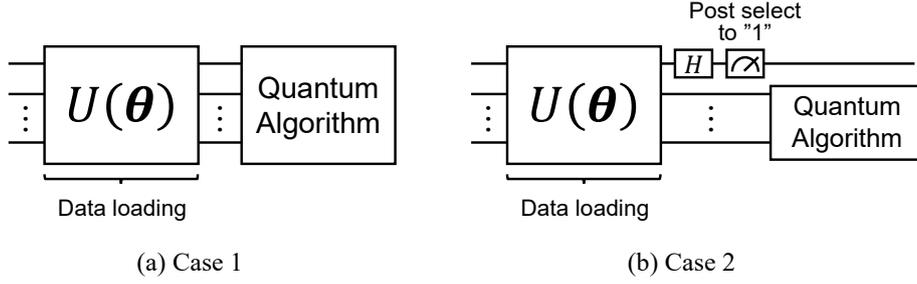


図 3.2: アルゴリズムの適用イメージ

しかし、この状態で単純に学習を行うと $U(\theta) |0\rangle^{\otimes n+1} = \pm |\bar{\psi}\rangle$ となってしまう。そこで最後に以下の処理を行う。まず、学習後の状態において、ancillary qubit に Hadamard 変換を行うと

$$I^{\otimes n} \otimes H |\bar{\psi}\rangle = \frac{|\text{Data}^+\rangle - |\text{Data}^-\rangle}{\sqrt{2}} |0\rangle + \frac{|\text{Data}^+\rangle + |\text{Data}^-\rangle}{\sqrt{2}} |1\rangle$$

となる。従って、ここで ancillary qubit が $|1\rangle$ となる状態を post-select することによって、所望の状態である式 (3.4) が得られる。このような拘束条件の下で学習された PQC を用いてデータを量子状態へ埋め込み、その後 ancillary qubit を除く qubit に対して所望の量子アルゴリズムを実行する。(図 3.2 (b))

3.2.5 エンコーダー $U(\theta)$ の学習方法

本項では、エンコーダー $U(\theta)$ の学習方法の一例として、最急降下法を用いる例を示す。尚、第 4 章でも本手法を用いて学習したエンコーダーを使用する。最急降下法を行うにあたり、最も重要なのがコスト関数の定義である。本問題設定では、量子回路の出力の測定結果に基づいて、量子状態の振幅を所望の値に調整する必要がある。すなわち、確率分布間の距離を適切に評価する必要がある。確率分布間の距離の評価には様々な指標を用いることができるが、本研究では Maximum Mean Discrepancy (MMD) を用いた [62, 63]。これは、1) サンプルングにより効率的に評価ができる点、2) parameter shift rule により効率的に学習が可能のためである。他にも類似の指標として、Stein discrepancy (SD) や Shinkhorn divergence (SHD) [64, 64] を用いることもできる。

今、学習対象のモデルから出力される確率分布を $q_\theta(j)$ 、ターゲットの確率分布を $p(j)$ とすると、コスト関数 $\mathcal{L}_{MMD}(q_\theta(j), p(j))$ は以下のように定義できる。

$$\begin{aligned} \mathcal{L}_{MMD} &\equiv \gamma_{MMD}(q_\theta, p)^2, \\ \gamma_{MMD} &= \left| \sum_{j=0}^{N-1} q_\theta(j) \Phi(j) - \sum_{j=0}^{N-1} p(j) \Phi(j) \right| \end{aligned} \quad (3.5)$$

ここで $\Phi(j)$ は j を特徴量空間へ写像する関数であり、これによって定義されるカーネル関数 $\kappa(j, k) \equiv \Phi(j)^T \Phi(k)$ を Gaussian カーネルに選ぶと、 $\mathcal{L}_{MMD}(q_\theta(j), p(j)) = 0$ は $q_\theta(j) = p(j)$ と等価であることが示されている [62, 63]。 \mathcal{L}_{MMD} をサンプルングから推定することを念頭におくと、式 (3.5) は $\kappa(j, k)$ を用いて

$$\mathcal{L}_{MMD} = \mathbf{E}_{\substack{j \sim q_\theta \\ k \sim q_\theta}} [\kappa(j, k)] - 2 \mathbf{E}_{\substack{j \sim q_\theta \\ k \sim p}} [\kappa(j, k)] + \mathbf{E}_{\substack{j \sim p \\ k \sim p}} [\kappa(j, k)] \quad (3.6)$$

と表現できる。

次に具体的な確率分布を代入する. 例として 3.2.4 項で記した Case 1 について考えると, 学習対象のモデルから出力される確率分布は

$$\begin{aligned} q_\theta(j) &= |\langle j|U(\boldsymbol{\theta})|0\rangle^{\otimes n}|^2, \\ q_\theta^H(j) &= |\langle j|H^{\otimes n}U(\boldsymbol{\theta})|0\rangle^{\otimes n}|^2 \end{aligned}$$

となり, これらがそれぞれ

$$p(j) = \mathbf{d}_j^2, \quad p^H(j) = (\mathbf{d}_j^H)^2$$

となるように学習を進める (Case 2 も同様). 実際の学習では,

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{\mathcal{L}_{MMD}(q_\theta, p) + \mathcal{L}_{MMD}(q_\theta^H, p^H)}{2} \quad (3.7)$$

をコスト関数として, これを最小化するようにパラメタ更新を行うことで目的の状態を求めることができる.

コスト関数 (3.7) を最急降下法を用いて最小化するにあたり, パラメタ θ が single rotation gate の回転角 $\exp(-i\theta_r\sigma_y/2)$ に埋め込まれている状況を仮定する. この場合, q_θ と q_θ^H の θ_r に関する勾配は, parameter shift rule [65] によって

$$\begin{aligned} \frac{\partial q_\theta(j)}{\partial \theta_r} &= q_\theta^+(j) - q_\theta^-(j), \\ \frac{\partial q_\theta^H(j)}{\partial \theta_r} &= q_\theta^{H+}(j) - q_\theta^{H-}(j) \end{aligned}$$

と表せる. 但し, $q_\theta^\pm(j) = |\langle j|U_{r\pm}(\boldsymbol{\theta})|0\rangle|^2$, $q_\theta^{H\pm}(j) = |\langle j|HU_{r\pm}(\boldsymbol{\theta})|0\rangle|^2$ であり, parameter shift されたユニタリ演算子を

$$\begin{aligned} U_{r\pm} &= U_{r\pm}(\theta_1, \dots, \theta_{r-1}, \theta_r, \theta_{r+1}, \dots, \theta_R) \\ &= U(\theta_1, \dots, \theta_{r-1}, \theta_r \pm \pi/2, \theta_{r+1}, \dots, \theta_R) \end{aligned}$$

とおいた. 尚, R はパラメタ総数であり, l を PQC の深さとして $R = ln$ である. 以上を踏まえて式 (3.6) を微分し, \mathcal{L} の勾配を計算すると

$$\begin{aligned} 2 \frac{\partial \mathcal{L}}{\partial \theta_r} &= \mathbf{E}_{\substack{j \sim q_{\theta_r^+} \\ k \sim q_\theta}} [\kappa(j, k)] - \mathbf{E}_{\substack{j \sim q_{\theta_r^-} \\ k \sim q_\theta}} [\kappa(j, k)] \\ &\quad - \mathbf{E}_{\substack{j \sim q_{\theta_r^+} \\ k \sim p}} [\kappa(j, k)] + \mathbf{E}_{\substack{j \sim q_{\theta_r^-} \\ k \sim p}} [\kappa(j, k)] \\ &\quad + \mathbf{E}_{\substack{j \sim q_{\theta_r^{H+}} \\ k \sim q_\theta^H}} [\kappa(j, k)] - \mathbf{E}_{\substack{j \sim q_{\theta_r^{H-}} \\ k \sim q_\theta^H}} [\kappa(j, k)] \\ &\quad - \mathbf{E}_{\substack{j \sim q_{\theta_r^{H+}} \\ k \sim p^H}} [\kappa(j, k)] + \mathbf{E}_{\substack{j \sim q_{\theta_r^{H-}} \\ k \sim p^H}} [\kappa(j, k)] \end{aligned}$$

と表すことができる. これは j, k を確率分布 $q_\theta, q_{\theta_r^+}, q_{\theta_r^-}, q_{\theta_r^{H+}}, q_{\theta_r^{H-}}, p, p^H$ からサンプリングすることで効率的に近似できる. これにより $\mathcal{L}(\boldsymbol{\theta})$ を最小化するパラメタベクトル $\boldsymbol{\theta} = (\theta_1, \dots, \theta_R)$ の更新を最急降下法で行うことができる.

3.3 結論

第3章では, NISQ を前提とした効率的な振幅エンコーディングの手法について記した. 本手法により, qubit 数 n に対して指数の深さが必要だった振幅埋め込みのための量子回路の深さが, 機械学習の手法を導入することで (振幅に有限の誤差を許せば) $O(\text{poly } n)$ ですむことになる. これは多くのデータをエンコードする際には非常に恩恵が大きい, 一方で欠点として, エンコーダの生成に学習過程が必要になるため, データのリアルタイム処理には適していない. しかし, 逆説的にはデータベース検索のような, 一旦巨大なデータベースを構築しておき, それを何度も参照するようなタスクには有用である. また, 今後の発展の方向性としては, そのようなデータベースのオンライン学習や転移学習があり得る. もちろん, PQC の効率的な学習方法や確率分布の距離を効率的に推定する手法の開発などのサブルーチンの開発も期待される. 第4章では, 本手法を応用することで, 歴史的に大前提とされてきた Grover's algorithm における Oracle(ブラックボックス回路) が不要なアルゴリズムの提案について記す.

第 4 章

明示的な Oracle の実装：Grover's algorithm

第 4 章では、Grover's algorithm (GA) において Oracle(ブラックボックス回路) を不要とする手法を提案する。GA は、量子振幅増幅とも呼ばれ、量子計算の優位性を理論的に保証する重要かつ基盤的な量子アルゴリズム/サブルーチンとして知られている。GA を応用した例として、充足問題 [66], 量子機械学習 [67], 制約付き多項式バイナリ最適化 [68], 量子振幅推定 [47] などがある。その中でも本研究では、データベース検索問題に着目する [69]。データベース検索は、産業界でも広く使われる重要な問題であると同時に、GA が初めて提案された際のモチーフでもある。“検索問題”は“パターンマッチング問題”と捉えることもでき、以下の問題ような問題として示すことができる。まず、各インデックスが振られた多数の要素で構成されたデータベースを用意する。我々のタスクは「与えられたクエリ (query) データと最も近い要素をもつインデックスは何か？」を見付けることである。データベースがクエリデータを含んでおり、そのインデックスを調べることが“検索”と呼ばれる。1996 年に発表された Grover の原著論文 [69] では、インデックスのみに着目した単純化された問題が扱われ、それをモチーフにいわゆる教科書的な量子振幅増幅が示された。それによると「 N 個のデータがあった場合、 $O(\sqrt{N})$ 回の試行でクエリデータを検索する」ことができるということが理論的に証明できる。尚、古典手法では最も良い手法でも $O(N)$ 回の試行が必要となる。これが Grover's algorithm によって得られる 2 乗加速の効果である。

これを量子アルゴリズムに応用するために、GA の実装方法がこれまで数多く研究されている。しかし、我々の知る限り、真の意味で実用にたまるものは見当たらない。すなわち、現実的な問題設定において GA を実装した例は無いのである。本研究では、状態準備からパターンマッチング、読み出しに至る問題を一气通貫で解決するアルゴリズムを提案する。本章では、初めに従来の GA について説明した後に、その実装上の問題点について記す。その後、改めて問題設定を明確化した後に、提案アルゴリズムとそれを用いたデモンストレーションについて記載する。

4.1 従来の Grover's algorithm と問題点

Grover's algorithm とは

本項では、教科書的な GA [69] についてレビューする。まず、初期状態 $|\Psi\rangle$ とクエリ状態 $|q\rangle$ を以下のように定義する。

$$|\Psi\rangle \equiv \sin \theta |q\rangle + \cos \theta |q^\perp\rangle, \quad (4.1)$$

$$|q\rangle \equiv \frac{1}{N} P |\Psi\rangle, \quad (4.2)$$

$$|q^\perp\rangle \equiv \frac{1}{N'} (1 - P) |\Psi\rangle \quad (4.3)$$

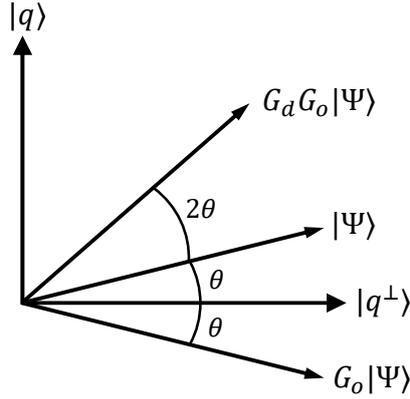


図 4.1: Grover's algorithm のイメージ

ここで $|q^\perp\rangle$ は $|q\rangle$ に直行する状態であり、これを用いることで $|\Psi\rangle$ は式 (4.1) と分解できる。尚、 $N = \sqrt{\langle\Psi|P|\Psi\rangle}$ と $N' = \sqrt{\langle\Psi|(\mathbb{1} - P)|\Psi\rangle}$ は正規化定数、 P は射影演算子、 $\theta = \langle q|\Psi\rangle$ である。目的はクエリ状態の振幅 $\sin\theta$ を増幅することである。これを実現するのが、Grover 演算子

$$G \equiv G_d G_o,$$

である。ここで G_o, G_d はそれぞれ、Oracle 演算子と Diffusion 演算子と呼ばれ、

$$\begin{aligned} G_o &\equiv \mathbb{1} - 2P, \\ G_d &\equiv 2|\Psi\rangle\langle\Psi| - \mathbb{1} \end{aligned}$$

と表すことができる。意味合いとしては、 G_o により増幅対象の状態の位相を反転させ（マーキング）、 G_d により $|\Psi\rangle$ を中心に位相を折り返すことに対応する（図 4.1）。実際、上記の方程式を組み合わせることで

$$G \begin{bmatrix} |q\rangle \\ |q^\perp\rangle \end{bmatrix} = \begin{bmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{bmatrix} \begin{bmatrix} |q\rangle \\ |q^\perp\rangle \end{bmatrix}$$

が得られる。この式の意味するところは「 G は、 $|q\rangle, |q^\perp\rangle$ で張られた平面内を角度 2θ だけ回転する」ということである。その結果、 G を t 回 $|\Psi\rangle$ に作用させた状態 $|\Psi^t\rangle \equiv G^t |\Psi\rangle$ は

$$|\Psi^t\rangle = \sin((2t+1)\theta)|q\rangle + \cos((2t+1)\theta)|q^\perp\rangle$$

と表される。従って、適切な t が分かれば $|q\rangle$ の振幅を 1 近くまで増幅することができる。これが最も基本的な GA の考え方である。

実装時の問題点 1: Oracle 演算子の構成

次に、これを明示的に実装することを考える。最も重要なのが、振幅増幅の対象となる状態をマーキングする G_o だが、これを構成するには P が必要となる。ところで P は式 (4.2) を通じて $|q\rangle, |\Psi\rangle$ によって表現される。これは、検索の問題設定では都合が悪い。なぜなら、検索結果に対応する状態（振幅増幅される対象）に対する情報が手元にないと、Grover 演算子が構成できないことを意味するからである。これが従来の Grover's algorithm を実用する際に直面する 1 つ目の問題点である。但し、“検索”以外の用途、具体的には何らかの量子計算の結果得られた状態を増幅して観測したい場合や量子振幅推定の文脈では、ancillary qubit の追加などにより、予め観測対象の状態を区別できるようにするなどの工夫により、GA の振幅増幅の恩恵を得ることができる。しかし、それらの設計はアルゴリズムごとに行う必要があり、経験的な側面が大きい。

実装時の問題点 2: 状態準備にかかる計算量

実際にアルゴリズムを使用するには、何らかの演算を行うことで初期状態 $|\Psi\rangle$ と $|q\rangle$ を準備、もしくは表現する必要がある。いま $|\Psi\rangle$ として n -qubit 系を仮定すると、その次元は 2^n である。この状態にナイーブに状態を埋め込むことを考えると、 2^n ($\equiv N$) 個の multi-qubit-controlled gate が必要になる（第3章で記した通り、様々な状態準備の手法が提案されているが、どれも $O(2^n)$ のリソースが必要）。従って、GA により試行回数 (query complexity) が $O(\sqrt{N})$ ですむとしても、トータルの計算量は n に対して指数的に増大する。これでは GA に依って得られた量子加速が大きく毀損されてしまう。これが2つ目の問題点である。

4.2 提案手法

本項では、前述した2つの問題を解決するアルゴリズムについて説明する。

4.2.1 問題設定：量子パターンマッチング

まず、手元に N_D 次元の実データベクトル $\mathbf{a}_k = [a_{0,k}, \dots, a_{N_D-1,k}]^T$ があるとする。ここで $k = 0, \dots, N_I - 1$ は各データに振られたインデックスを表す。すなわち、データベースに N_I 個の要素が格納されているという問題設定である。この状況下で、 N_D 次元の実データベクトルとしてクエリ (query) $\mathbf{b} = [b_0, \dots, b_{N_D-1}]^T$ が与えられる。尚、 \mathbf{a}_k, \mathbf{b} は共に量子状態として与えられるため、正規化条件 $\sum_j a_{j,k}^2 = 1, \sum_j b_j^2 = 1$ を満たす。我々の目標は、データベースに格納された要素の中で最も \mathbf{b} に近いもののインデックス (target index; $k_* = \operatorname{argmax}_k |\mathbf{a}_k^T \mathbf{b}|$) を求めることである。

4.2.2 提案手法：Black box less Grover's algorithm

Step1: 状態準備

前述したタスクを量子デバイスで行うためには、データを量子状態として埋め込む必要がある。具体的には、以下のようにデータベース状態 $|\text{database}\rangle$ とクエリ状態 $|\text{query}\rangle$ を用意すればよい。Amplitude encoding (AE) を用いると

$$|\text{database}\rangle = A|0\rangle^{\otimes(n_D+n_I)} \quad (4.4)$$

$$= \frac{1}{\sqrt{N_I}} \sum_{j=0}^{N_D-1} \sum_{k=0}^{N_I-1} a_{jk} |j\rangle_D |k\rangle_I \quad (4.5)$$

$$\equiv \frac{1}{\sqrt{N_I}} \sum_k |\text{data}(k)\rangle \otimes |k\rangle_I, \quad (4.6)$$

$$|\text{query}\rangle = B|0\rangle^{\otimes n_D} \quad (4.7)$$

$$= \sum_{j=0}^{N_D-1} b_j |j\rangle_D. \quad (4.8)$$

となる。ここで $\{|j\rangle_D\}$ はデータに対応するヒルベルト空間 \mathcal{H}_D 上で直交した計算基底を用いた表現であり、 $\{|k\rangle_I\}$ はインデックスに対応するヒルベルト空間 \mathcal{H}_I 上で直交した計算基底を用いた表現である。単純化のため、用いる qubit 数を n_D, n_I とし、各空間の次元数を $N_D = 2^{n_D}, N_I = 2^{n_I}$ とする。式 (4.5) では、 k 番目のデータベクトルに対応する量子状態 $|\text{data}(k)\rangle = \sum_{j=0}^{N_D-1} a_{jk} |j\rangle_D$ を導入した。

すなわち, $|\text{database}\rangle$ が全てのインデックスに対応するデータベクトルの重ね合わせ状態をとっていることに対応する. 式 (4.4), (4.7) では, 対応する量子状態を生成するユニタリー演算子として A, B を導入した. 但し, 4.1 節で指摘した通り, A, B を明示的に構成するためには指数個のゲート操作が必要となる [70, 71, 72, 73].

ちなみに, Basis encoding (BE) を用いると

$$|\text{database}\rangle = \frac{1}{\sqrt{N_I}} \sum_{j=0}^{N_D-1} \sum_{k=0}^{N_I-1} \hat{a}_{jk} |j\rangle_D |k\rangle_I, \quad (4.9)$$

$$|\text{query}\rangle = \frac{1}{\sqrt{N_{BE}}} \sum_{j=0}^{N_D-1} \hat{b}_j |j\rangle_D. \quad (4.10)$$

となる. ここで \hat{a}_{jk} と \hat{b}_j はバイナリ値 (i.e., 0 or 1) であり, 元々の状態 a_{jk} and b_j から決定される. $\hat{\mathbf{a}}$ は N_I 個の non-zero 要素を含み, N_{BE} はクエリを BE した際の正規化定数である. これを用いると, 例えば $n_I = n_D = 2$ の下で, クエリが $|1\rangle_D$ だとすると, $\hat{\mathbf{b}} = [\hat{b}_0, \hat{b}_1, \hat{b}_2, \hat{b}_3]^\top = [0, 1, 0, 0]^\top$ と表せる. 同様に, データベース内でデータ $|1\rangle_D$ がインデックス $|0\rangle_I$ に格納されているとすると, $\hat{\mathbf{a}} = [\hat{a}_{00}, \hat{a}_{10}, \hat{a}_{20}, \hat{a}_{30}, \dots, \hat{a}_{23}, \hat{a}_{33}]^\top = [0, 1, 0, 0, \dots, 0, 0]^\top$ と表せる.

Step2: 類似度計算

古典データベクトル \mathbf{a}_k と \mathbf{b} の内積は, 量子状態を用いた文脈では以下のように Fidelity として解釈できる.

$$\langle \text{query} | \text{data}(k) \rangle = \langle 0^{n_D} | B^\dagger | \text{data}(k) \rangle, \quad (4.11)$$

ここで $|0^n\rangle = |0\rangle^{\otimes n}$ という表現を導入した. ところで式 (4.11) の右辺は

$$|\text{query}'\rangle = |0\rangle^{\otimes n}$$

と $B^\dagger | \text{data}(k) \rangle$ の間の Fidelity と捉えることもできる. そこで本提案手法では, 式 (4.11) の評価に inversion test [74] を用いる. inversion test は 2 状態間の Fidelity を計算するナイーブな手法であり, 状態 $B^\dagger | \text{data}(k) \rangle$ を計算基底で測定した際に $|\text{query}'\rangle$ が得られる確率を推定することに対応する.

今, 我々の手元には各 $|\text{data}(k)\rangle$ ではなく $|\text{database}\rangle = A |0\rangle^{\otimes (n_D+n_I)}$ があることに注意すると, inversion test によって得られる状態を $|\Psi\rangle$ として

$$\begin{aligned} |\Psi\rangle &\equiv (B^\dagger \otimes \mathbb{1}_I) A |0\rangle^{\otimes (n_D+n_I)} \\ &= \frac{1}{\sqrt{N_I}} \sum_k B^\dagger | \text{data}(k) \rangle \otimes |k\rangle_I. \end{aligned}$$

と表すことができる. ここで $\mathbb{1}_I$ は \mathcal{H}_I における恒等演算子である. この状態に対して $|\Psi\rangle$ のうち最初の n_D -qubit を計算基底で測定し, それらの測定結果が全て 0 となる状態を post-select する. このような操作の全体像を示したのが図 4.2 である. 図中で G は後述する Grover 演算子である. 以上の操作の結果得られる終状態は

$$|R\rangle = \frac{1}{C_R} \sum_k r_k |k\rangle_I, \quad (4.12)$$

と表すことができる. 但し,

$$r_k = \langle \text{query} | \text{data}(k) \rangle, \quad C_R = \sqrt{\sum_k |\langle \text{query} | \text{data}(k) \rangle|^2}.$$

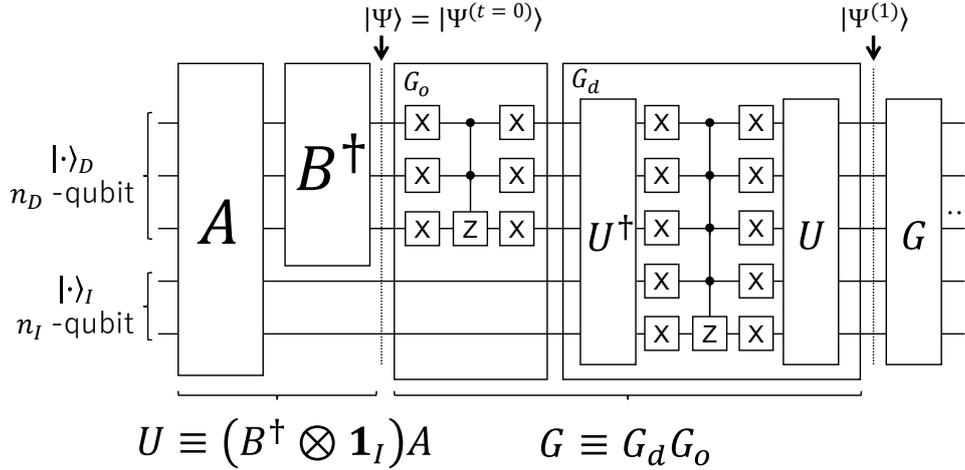


図 4.2: パターンマッチを行う量子回路の全体像. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

である。従って、式 (4.12) の状態に対して \mathcal{H}_I における計算基底で測定を行うと、 $|\langle \text{query} | \text{data}(k) \rangle|^2$ に比例した確率でインデックス k の確率分布が得られそうである。すなわち、我々が目的としていたパターンマッチングの結果に相当する情報が得られる。

しかし、ここで示したアプローチには致命的な問題点がある。それは target index k_* を推定するために行う最後の post-select の成功確率が、データベースのサイズ N_I が大きくなるにつれて、非常に小さくなってしまふという点である。具体的には、post-select が成功して結果的に target index k_* が同定できる確率は

$$\begin{aligned}
 P(\text{index} = k_*) &= |{}_I \langle k_* | \langle 0^{n_D} | \Psi \rangle|^2 \\
 &= \frac{1}{N_I} |\langle \text{query} | \text{data}(k_*) \rangle|^2 \\
 &\leq \frac{1}{N_I}.
 \end{aligned}$$

と表される。この結果は「target index を同定するために必要な計算量は $O(N_I)$ であり、古典手法と同一である」ことを意味している。そこで導入するのが Grover's algorithm である。本問題設定の場合、 $|\Psi\rangle$ のうち $|0\rangle^{\otimes n_D} |k\rangle_I$ に一致する振幅の増幅を行う。

Step3: 振幅増幅

はじめに本アルゴリズムのポイントを記す。本設定において振幅増幅の目的は、 $|\text{database}\rangle = A|0\rangle^{\otimes n_D+n_I}$ のうち $|\text{query}\rangle = B|0\rangle^{\otimes n_D}$ と重なりをもつ要素を増幅することである。これは $|\Psi\rangle = (B^\dagger \otimes \mathbf{1})A|0\rangle^{\otimes n_D+n_I}$ のうち $|\text{query}'\rangle = |0\rangle^{\otimes n_D}$ との重なりを増幅することと等価であり、後述するように、このシンプルなトリックこそが Oracle 演算子を明示的に構成するうえで重要となる。すなわち、Grover 演算子に含まれる Oracle 演算子の作用が、エンタングルしたクエリ状態 $|\text{query}\rangle$ ではなく、単純な計算基底 $|0\rangle^{\otimes n_D}$ への射影で済むからである。これはクエリの生成演算子 B を明示的に構成できさえすれば、単純な inversion test で実現できる。この時点で Oracle 演算子は、既に“Oracle(神託機械)”ではなく、求めたい解に関数する情報が手元になくても実装可能である。一旦、Grover 演算子が構成できてしまえば、データベース内の各要素とクエリの類似度（式 (4.12) における non-zero な r_k に対応）に比例するように振幅を増幅できる。本アルゴリズムは、従来の Grover's

algorithm の拡張的な概念となっており, target state が初期状態を構成する基底状態ではなく, 分布をもっているという問題設定 [75] と見ることが出来る. 次に, ここまでの理論を実際に適用してみる. まず, 初期状態 $|\Psi\rangle$ と target state $|q\rangle$ を以下のように定義する.

$$|\Psi\rangle = (B^\dagger \otimes \mathbb{1}_I)A|0\rangle^{\otimes(n_D+n_I)},$$

$$|q\rangle \equiv \frac{(|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I)|\Psi\rangle}{\sqrt{\langle\Psi|(|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I)|\Psi\rangle}}.$$

尚, $|0^n\rangle = |0\rangle^{\otimes n}$ である. ここで target state $|q\rangle$ は, $|\Psi\rangle$ のうち \mathcal{H}_D において $|\text{query}'\rangle = |0\rangle^{\otimes n_D}$ と重なりをもつ要素である. 具体的には, $|\Psi\rangle$ は

$$|\Psi\rangle = \sin\theta|q\rangle + \cos\theta|q^\perp\rangle,$$

と表される. ここで $|q^\perp\rangle$ は $|q\rangle$ と直交する状態であり, $\sin\theta = \langle q|\Psi\rangle$ である. また, 我々の問題設定におけるデータベース状態と target state は, シンプルに一つの基底状態で表される訳ではなく, 複数の状態に分布しているため, 以下のように計算基底を用いて表すのが便利である.

$$|\Psi\rangle = \sum_{x=0}^{N-1} \psi_x|x\rangle, \quad |q\rangle = \sum_{x=0}^{N-1} q_x|x\rangle, \quad (4.13)$$

尚, $N = N_D N_I$ はヒルベルト空間 $\mathcal{H}_D \otimes \mathcal{H}_I$ の次元数であり, $\{x\}$ は当該空間における基底の集合である. ところで $\{q_x\}$ は最大でも N_I の non-zero 成分しかもたず, 明示的に以下のように表現できる.

$$q_x = \begin{cases} \psi_x/\langle q|\psi\rangle & \text{if } x \in \mathcal{C}, \\ 0 & \text{if } x \notin \mathcal{C}, \end{cases}$$

ここで \mathcal{C} は

$$\mathcal{C} = \{x \mid \langle x|(|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I)|x\rangle \neq 0\}$$

で定義される集合である. 但し, $\langle q|\Psi\rangle = [\langle\Psi|(|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I)|\Psi\rangle]^{1/2}$ である.

振幅増幅の目的は Grover 演算子を作用させて係数 $\sin\theta$ を増幅することである. Grover 演算子は以下のように, Oracle 演算子 G_o と Diffusion 演算子 G_d に分解できる.

$$\begin{aligned} |\Psi^{(t+1/2)}\rangle &= G_o|\Psi^{(t)}\rangle \\ &= (\mathbb{1} - 2|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I)|\Psi^{(t)}\rangle, \\ |\Psi^{(t+1)}\rangle &= G_d|\Psi^{(t+1/2)}\rangle \\ &= U(2|0^{n_D+n_I}\rangle\langle 0^{n_D+n_I}| - \mathbb{1}_{DI})U^\dagger|\Psi^{(t+1/2)}\rangle \\ &= (2|\Psi\rangle\langle\Psi| - \mathbb{1}_{DI})|\Psi^{(t+1/2)}\rangle, \end{aligned}$$

尚, $U = (B^\dagger \otimes \mathbb{1}_I)A$ は初期状態 $|\Psi^{(0)}\rangle = |\Psi\rangle = U|0\rangle^{\otimes n_D+n_I}$ を生成するユニタリ演算子である. また, 振幅増幅された状態 $|\Psi^{(t)}\rangle$ は初期状態 $|\Psi\rangle$ に対して Grover 演算子 $G = G_d G_o$ を t 回作用させることで生成される. Oracle 演算子 G_o は target state $|q\rangle$ の位相を反転する. その際, $|q\rangle$ に直交する他の状態は変化させない. Diffusion 演算子 G_d は全状態の平均的な振幅に対して $|q\rangle$ の振幅を反転させる. 最も重要なのが, 従来 “Oracle” と呼ばれていた G_o は, 本アルゴリズムでは target state(解状態)に関するいかなる情報も必要とせず, 明示的に構成可能な点である. エンコーダーから振幅増幅に至る量子回路の全容を示したのが図 4.2 である.

4.1 節で示した従来の Grover’s algorithm と同様に, 本アルゴリズムでも Grover 演算子を t 回作用させた状態を明示的に以下のように表すことができる.

$$|\Psi^{(t)}\rangle = G^t|\Psi\rangle = \sin((2t+1)\theta)|q\rangle + \cos((2t+1)\theta)|q^\perp\rangle,$$

この式から分かることは、増幅された状態は t に対して周期的に振動するということである。従って、大きく増幅された $|q\rangle$ を観測するには適切な回数 t を選択する必要がある。そこで t について解析するために $|\Psi^{(t)}\rangle$ を式 (4.13) と同様に計算基底で展開して考える。

$$|\Psi^{(t)}\rangle = \sum_x \psi_x^{(t)} |x\rangle, \quad \psi_x^{(t)} = A_x \sin(\omega t + \delta_x),$$

ここで ω , A_x , δ_x (すなわち、周波数, 振幅, 初期状態に対する位相) は

$$\begin{aligned} \omega &= 2 \arcsin \langle q | \Psi \rangle, \\ A_x &= \sqrt{\frac{q_x^2 - 2\langle q | \Psi \rangle \psi_x q_x + \psi_x^2}{1 - \langle q | \Psi \rangle^2}} \\ &= \begin{cases} \psi_x / \langle q | \Psi \rangle & \text{if } x \in \mathcal{C}, \\ \psi_x / \sqrt{1 - \langle q | \Psi \rangle^2} & \text{if } x \notin \mathcal{C}, \end{cases} \\ \delta_x &= \arccos \left(\frac{q_x - \langle q | \Psi \rangle \psi_x}{\sqrt{q_x^2 - 2\langle q | \Psi \rangle \psi_x q_x + \psi_x^2}} \right) \\ &= \begin{cases} \arccos \sqrt{1 - \langle q | \Psi \rangle^2} & \text{if } x \in \mathcal{C}, \\ \arccos(-\langle q | \Psi \rangle) & \text{if } x \notin \mathcal{C}, \end{cases} \end{aligned}$$

と表すことができる。尚, $\langle q | \Psi \rangle = [\langle \Psi | (|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I) | \Psi \rangle]^{1/2}$ である。ちなみに, $x \in \mathcal{C}$ に対応する $\psi_x^{(t)}$ の位相と $x \notin \mathcal{C}$ に対応するものの位相は, ちょうど $\pi/2$ だけ互いにずれていることを指摘しておく。

この結果に基づくと, クエリに一致した振幅 ($x \in \mathcal{C}$) が増幅された結果を解析的に表現できる。 t 回の Grover 演算によって増幅された確率 P_x^t は

$$\begin{aligned} P_x^{(t)} &= \{\psi_x^{(t)}\}^2 \\ &= \frac{\psi_x^2}{2\langle q | \Psi \rangle^2} \left[1 - \cos 2(\omega t + \arccos \sqrt{1 - \langle q | \Psi \rangle^2}) \right] \\ &\leq \frac{\psi_x^2}{\langle q | \Psi \rangle^2}. \end{aligned} \tag{4.14}$$

となる。式 (4.14) に着目すると, $x \in \mathcal{C}$ に対して最も高い P_x が得られる最適な iteration 回数 t_* は

$$t_* = CI \left(\frac{\arccos \langle q | \Psi \rangle}{2 \arcsin \langle q | \Psi \rangle} \right), \tag{4.15}$$

と表される。尚, $CI(z)$ は z に最も近い実整数を返す (切り捨て)。

式 (4.15) が示しているのは, 最適な iteration 回数 t_* は初期状態における重なり $\langle q | \Psi \rangle = [\langle \Psi | (|0^{n_D}\rangle\langle 0^{n_D}| \otimes \mathbb{1}_I) | \Psi \rangle]^{1/2}$ に依存するということである, すなわち, 問題を高精度に扱うには $\langle q | \Psi \rangle$ を高精度に推定する必要がある。ナイーブには $|\Psi\rangle$ を $|0\rangle^{\otimes n_D}$ へ射影することで実行できるが, それでは非効率である。なぜなら, 仮に推定誤差を ϵ に抑えようとする, $O(1/\epsilon^2)$ 回の測定が必要になるからである。その代わりに洗練された振幅推定法を採用することができる [47, 76, 77]。これらを用いればパラメタ ω や $\langle q | \Psi \rangle$ が $O(1/\epsilon)$ 回の Grover 演算で推定できる。しかしながら, 実用的な場面を考えてみると, データベースのサイズ N_I が非常に大きなことが予想される。さらにデータベース内のヒットする要素の数も全体からすると極少数であろう。その場合, 大まかには $\omega \simeq \langle q | \Psi \rangle \simeq 1/\sqrt{N_I}$ と計算できる。すると最適な iteration 回数 $t_* \simeq \sqrt{N_I}$ となり, $\psi_x^2 \rightarrow 1/N_I$ の下で $P_x^{(t_*)} \simeq N_I \psi_x^2 \simeq 1$ となる。この結果は, 本アルゴリズムによるパターンマッチングに必要な試

行数が古典計算に比べて平方根で削減できることを示しており、従来の Grover's algorithm と同様に 2 乗加速が得られることを示している。

最後に、量子回路の実装について記す。本フレームワークでは、Oracle 演算子 G_o は非常にシンプルであり、一つの n_D -controlled Toffoli gate で実装できる。その結果、Oracle 部分の深さは $O(n_D^2)$ ですむ [78]。一方、Diffusion 演算子 G_d にはデータベース生成回路 A, B を含む U^\dagger, U が必要である。従って、これをナイーブな手法で厳密に実装するにはシステムサイズに対して指数関数的に多くのゲートが必要となる ($n^{n_D+n_I}$)。これは NISQ では到底実装できない深さであり、FTQC であっても量子加速を大きく毀損する程の困難を伴う。この問題点を解決するために第 3 章に記した AAE を導入する。

Step4: AAE の導入による効率化

第 3 章に記した AAE を用いることで、以下のようにデータベース状態 $|\text{database}\rangle$ とクエリ状態 $|\text{query}\rangle$ を近似的に生成することができる。

$$\begin{aligned}
 |\text{database}\rangle &\simeq A(\boldsymbol{\theta}_a)|0\rangle^{\otimes n_D+n_I} & (4.16) \\
 &= \frac{1}{\sqrt{N_I}} \sum_{j=0}^{N_D-1} \sum_{k=0}^{N_I-1} \tilde{a}_{jk}|j\rangle_D|k\rangle_I \\
 &= \frac{1}{\sqrt{N_I}} \sum_k |\text{data}(k)\rangle \otimes |k\rangle_I, \\
 |\text{query}\rangle &\simeq B(\boldsymbol{\theta}_b)|0\rangle^{\otimes n_D} & (4.17) \\
 &= \sum_{j=0}^{N_D-1} \tilde{b}_j|j\rangle_D.
 \end{aligned}$$

ここで $A(\boldsymbol{\theta}_a), B(\boldsymbol{\theta}_b)$ は、それぞれデータベース状態とクエリ状態の生成を行うパラメタ化されたユニタリ演算子である。 $\boldsymbol{\theta}_a, \boldsymbol{\theta}_b$ は、学習の結果得られた最適なパラメタベクトルである。そして、 $\tilde{a}_{jk}, \tilde{b}_j$ は、学習された PQC A, B によって生成される状態に対応する係数である。式 (4.16), (4.17) にある \simeq は、近似的な埋め込みであるところから生じる誤差を表している。これらの AAE に関する演算子の後に Grover 演算がなされる。ゲート構造を含む量子回路の詳細を図 4.3 に示す。

最後に、AAE は上述のような amplitude encoding のみならず、式 (4.9), (4.10) のような basis encoding にも適用できることを指摘しておく。論理的な構成を考えた際には、Toffoli gate を用いた実装が straightforward なため、特に [79] のような比較的効率的な手法を用いる場合もあるが、AAE を使うことでゲート数を劇的に削減することができる。AAE の導入には学習過程における有限の誤差が避けられないため近似的な埋め込みにはなってしまうが、状況に応じて使い分けることが重要であろう。

4.3 実装例：量子画像検索

提案アルゴリズムは一般の量子状態におけるパターンマッチングが行えることから、その応用可能性は非常に幅広い。本節では、例として本アルゴリズムの画像のパターンマッチング問題への適用した場合を示す。現在、産業の様々な領域で画像処理タスクが活用されているが、単純な操作であっても大量の画像データを扱う場合は、数値やテキストの処理に比べると計算負荷が高い。そのような状況を鑑みて、量子計算を使用してより効率的な処理を実現しようという試みがなされている。古典情報として表現された画像を量子計算で扱うためには、量子状態としてエンコードする必要があるが、

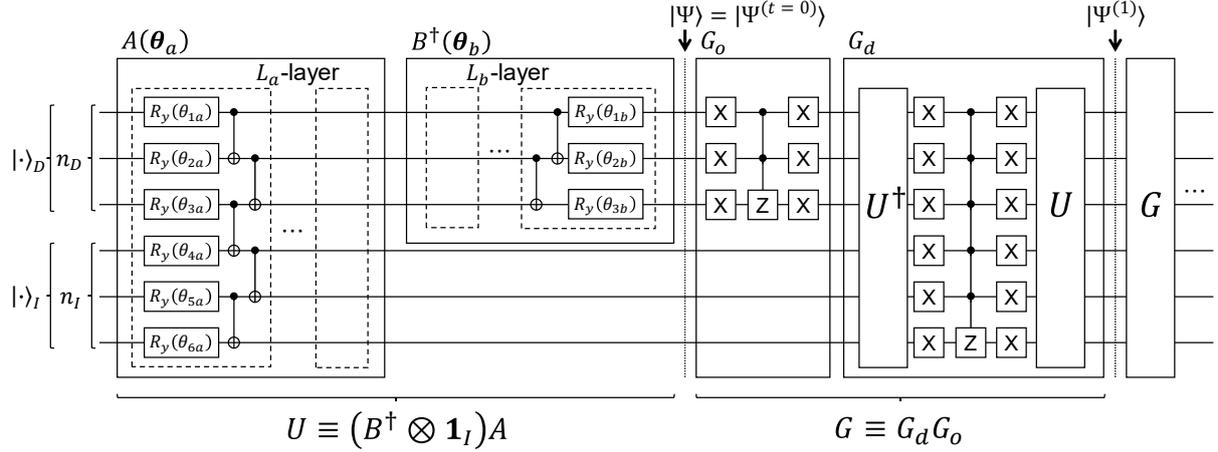


図 4.3: パターンマッチを行う量子回路の全体像の詳細 $(n_D, n_I) = (3, 3)$. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

これまで多くの表現方法が提案されている。以下では、まず量子画像に関する代表的なエンコード方法について簡潔に説明する。その後、パターンマッチング問題を一般的な表現で記述する。最後に、デモンストレーションとして、具体的なパターンマッチ問題を設定し、シミュレーションと実機を使って行った実験について記す。

4.3.1 量子状態を用いた画像の表現：Quantum image representation

量子状態を用いた画像の表現方法と画像処理の方法は [80] のような文献により、網羅的に記されている。特に表現方法については、行いたい画像処理タスクの内容・適用するアルゴリズムに応じて多くの種類が提案されているが、それらは大まかに amplitude encoding をベースにしたものと、basis encoding をベースにしたものに分類できる。前者の代表的なものとして flexible representation for quantum images (FRQI) が、後者の代表的なものとして novel enhanced quantum representation (NEQR) がある。本項では、両者について順に説明する。

Flexible representation for quantum images (FRQI)

FRQI は amplitude encoding をベースにしたエンコード方法であり、各画素 (pixel) に割り当てられた輝度 (色の強度情報) を重ね合わせを用いて表現する。FRQI が提案された原著論文 [81] ではグレースケール画像が用いられたが、その後カラー画像 [82] や動画等への応用 [83] も提案されている。ここでは簡単のためにグレースケール画像を例に取り、基本的な概念について説明する。いま手元に $N_p = 2^n \times 2^n$ pixel のグレースケール画像があるとし、それを FRQI を用いて表現すると

$$\begin{aligned}
 |I_{FRQI}\rangle &= \frac{1}{\sqrt{N_p}} \sum_{z=0}^{N_p-1} (\cos \theta_z |0\rangle + \sin \theta_z |1\rangle) |z\rangle \\
 &= \frac{1}{\sqrt{N_p}} \sum_{z=0}^{N_p-1} |f_z\rangle |z\rangle,
 \end{aligned} \tag{4.18}$$

と表現される。ここで $|0\rangle, |1\rangle$ は 1-qubit における計算基底を表し、 $\{|z\rangle\}$ は $\log N_p = 2n$ -qubit における計算基底の集合である。 $\theta = (\theta_0, \theta_1, \dots, \theta_{2^{2n}-1}), \theta_i \in [0, \pi/2]$ は輝度を表すベクトルであり、

$|f_z\rangle = \cos \theta_z |0\rangle + \sin \theta_z |1\rangle$ は、位置 z にある画素の輝度を表す。式 (4.18) にあるような $2^n \times 2^n$ pixel の画像は、 $(2n + 1)$ -qubit で表すことができる。

FRQI の利点は、各画素の輝度の絶対値を任意の精度で最小の qubit 数で表現できる点である。この点は各画素の輝度を直接 amplitude encoding により埋め込む手法 [84] よりも優れていると言える。実際、[84] で表現できるのはあくまで画素間の輝度の相対値であり、実用を想定すると大変使いにくい。例えば、1 箇所のみ輝度が最大でそれ以外の全ての pixel が輝度 0 であるケース、2 箇所が輝度最大のケースでそれ以外の全ての pixel が輝度 0 であるケースでは振幅が 2 倍も異なってしまうのである。

一方、欠点としては、振幅を精度良く推定するためには数多くの測定が必要であり、qubit 数の増加に伴い指数的に必要な測定回数が増加するという点がある。しかしながら、[85] でも指摘されているように、もしタスクの目的が厳密な画像情報を扱うものではなく、必要な精度がそこそこで構わないのであれば、この欠点は致命的ではない。例えば、本研究で着目しているパターンマッチングでは、各画像の厳密な輝度情報までは必要とせず、ユーザーが必要な精度に応じて測定回数を調整することができる。

Novel Enhanced Quantum Representation (NEQR)

NEQR は各画素の輝度をバイナリによりデジタル表現するエンコード方法である [86]。予め必要な有効精度に応じて qubit の数を決めておくことで、例えば q -qubit あれば 2^q 階調の輝度が表現できる。いま手元に $N_p = 2^n \times 2^n$ pixel のグレースケール画像があると、それを 2^q 階調の NEQR を用いて表現すると

$$\begin{aligned} |I_{NEQR}\rangle &= \frac{1}{\sqrt{N_p}} \sum_{z=0}^{N_p-1} \otimes_{i=0}^{q-1} |c_z^i\rangle |z\rangle \\ &= \frac{1}{\sqrt{N_p}} \sum_{z=0}^{N_p-1} |f(z)\rangle |z\rangle, \end{aligned}$$

となる。ここで $c_z^i \in \{0, 1\}$ は i 桁目のビット情報を表し、 $f(z) = \otimes_{i=0}^{q-1} c_z^i \in [0, 2^q - 1]$ は位置 z における画素の輝度に対応するビット列である。この表現からも、輝度の階調が使用するビット数に依存するために、高精度な表現を行うには FEQR に比べて多くのビット数が必要になることが分かる。一方で、各状態が直交していることから、 $O(N_p)$ 回の測定で N_p -pixel の量子画像を正確にデコードできる点は大きな利点である。

4.3.2 データベースとクエリの量子画像表現（一般系）

ここでは各画素の輝度の階調が $[0, N_C - 1]$ の範囲の整数で構成された N_P -pixel の画像を仮定する。これは 1 枚の画像を表現するデータベクトルが $N_D = N_P N_C$ 次元を持つことに対応する。この場合、前述した NEQR を想定するとスムーズに扱える。

まず、輝度の階調と画素の位置に対応する qubit 数を、それぞれ n_C, n_P とおくと $N_C = 2^{n_C}, N_P = 2^{n_P}$ である。従って、1 枚の画像を表現するのに必要な qubit 数は $n_D = n_C + n_P$ となる。いま、クエリの j 番目の画素の輝度をバイナリで表した整数値を g_j として、対応する量子状態を NEQR を用いて表示すると

$$|\text{query}\rangle = \frac{1}{\sqrt{N_P}} \sum_{j=0}^{N_P-1} |g_j\rangle_C |j\rangle_P. \quad (4.19)$$

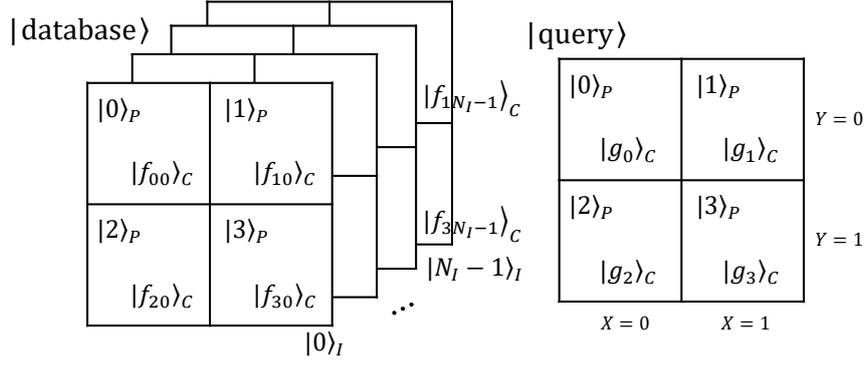


図 4.4: Database state (左) と Query state (右) . Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

となる. これは式 (4.10) において \hat{b}_j が $|j\rangle_D$ が式 (4.19) における $|g_j\rangle_C|j\rangle_P$ に一致したもののみ 1 である状態に対応する. 図 4.4 は $N_P = 4$ の場合を示したものである. データベース状態もクエリと同様の考え方で構成する. すなわち, データベース内の k 番目の画像の j 番目の画素の輝度を f_{jk} とすると, この画像に対応する量子状態は

$$|\text{data}(k)\rangle = \frac{1}{\sqrt{N_P}} \sum_{j=0}^{N_P-1} |f_{jk}\rangle_C |j\rangle_P.$$

と表される. 尚, $n_D = n_C + n_P$ である. データベース状態 $|\text{database}\rangle$ は, 式 (4.6) のように $|\text{data}(k)\rangle$ の重ね合わせとして表現される. 図 4.4 ではそのような $|\text{database}\rangle$ のイメージを示している. 尚, $\sum_{j=0}^{N_P-1} |f_{jk}\rangle_C$ と $\sum_{j=0}^{N_P-1} |g_j\rangle_C$ は正規化されていることに注意されたい. 我々の目的は, $|\text{database}\rangle$ の要素のうち $|\text{query}\rangle$ と最も高い類似度をもつインデックスを推定することである. 振幅増幅を行わない場合にインデックス k が観測される確率が

$$\begin{aligned} P(\text{index} = k) &= |{}_I\langle k | \langle 0^{n_D} | \Psi \rangle|^2 \\ &= \frac{1}{N_I} |\langle \text{query} | \text{data}(k) \rangle|^2 \\ &= \frac{1}{N_I} \left[\sum_{j=0}^{N_P-1} \frac{\langle g_j | f_{jk} \rangle}{N_P} \right]^2 \\ &\leq \frac{1}{N_I}. \end{aligned} \quad (4.20)$$

と表されることに注意する.

4.3.3 問題設定

ここではクエリデータが 4-pixel のバイナリ画像の場合を例にとって具体的なタスクを考える. 尚, この設定は $n_C = 1, n_P = 2$ に対応し, NEQR でも FRQI でも同等の表現になる. 4-pixel のバイナリ画像に対応する量子状態を全て列挙すると $\{|0h\rangle, |1h\rangle, \dots, |Fh\rangle\}$ の全 16 通りがある (図 4.5). 尚, $\cdot h$ は 16 進表現を表す. このうち, 例として 8 枚の画像がそれぞれのインデックスに登録された

($n_I = 3$ に対応) 以下のデータベースを考えることにする.

$$\begin{aligned} & \{|data(0)\rangle, |data(1)\rangle, |data(2)\rangle, |data(3)\rangle, |data(4)\rangle, |data(5)\rangle, |data(6)\rangle, |data(7)\rangle\} \\ & = \{|0h\rangle, |2h\rangle, |4h\rangle, |6h\rangle, |8h\rangle, |Ah\rangle, |Ch\rangle, |Eh\rangle\}. \end{aligned}$$

従って, データベース状態 $|database\rangle$ は 6-qubit の量子状態となる. 内訳は, 輝度 $n_C = 1$, 画素位置 $n_P = 2$ ($X \in \{0, 1\}, Y \in \{0, 1\}$), インデックス $n_I = 3$ である. このデータベースに対して, 様々なクエリを与え, どのような観測結果が得られるかを確認した.

4.3.4 実験の詳細

本実験は, 以下の 3 つの設定 (ケース) で行った. (i) 状態生成 (AAE) とパターンマッチングの両方とも QASM シミュレータ (Qiskit [87]), (ii) 状態生成 (AAE) は QASM シミュレータ, パターンマッチングは実機, (iii) 状態生成 (Toffoli gate による厳密埋込み) とパターンマッチングの両方 QASM シミュレータ. 尚, 実機には超電導量子コンピュータ IBM Quantum (ibm.kawasaki) を使用した.

続いて, AAE により画像をエンコードした際の設定を以下に記す. データベース状態を生成する PQC $A(\theta_a)$ には, 6 層の hardware efficient ansatz (HEA) (図 4.3) を採用し, クエリ状態を生成する PQC $B(\theta_b)$ には, 3 層の HEA を採用した. 各層は parameterized single-qubit Y -rotation gate $R_y(\theta_r) = \exp(-i\theta_r\sigma_y/2)$ と CNOT gate で構成されている. 尚, θ_r は r 番目のパラメータを表し, σ_y は Pauli Y operator を表している. すなわち, $A(\theta_a)$ と $B(\theta_b)$ は実行列である. θ_r は, 各学習過程の初めにランダムに初期化される. MMD を計算する際のカーネル関数には, $\kappa(x, y) = \exp(-64(x - y)^2)$ を用いた. コスト関数の r 番目のパラメータに関する勾配を推定するにあたって生成したサンプル数は, $B(\theta_b)$ の学習には, q_θ^+ , q_θ^- , $q_{\theta_r}^{H+}$, $q_{\theta_r}^{H-}$ それぞれに対し 400 サンプルを用いた. 同様に $A(\theta_a)$ の学習には, それぞれに 10000 サンプルを用いた. 最急降下法のための optimizer には, Adam [88] を使用した. 学習率は学習の過程により変化させており, 最初の 100iteration には 0.1 を, それ以降は 0.01 とした. 学習に必要な iteration 数, すなわちパラメータ更新の回数は, $B(\theta_b)$ の学習には 300, $A(\theta_a)$ の学習には 500 とした.

AAE の学習の後, パターンマッチングアルゴリズムを実行した. 確率分布 $P(\text{index} = k)$ を, 式 (4.20) において $k \in \{0, 1, \dots, 7\}$ に相当する全ての状態に対して求めることで, クエリとの類似度が最も高いインデックスを推定した. これを Grover's algorithm による振幅増幅を適用した場合と適用しない場合に対して行い, 両者を比較した. サンプル数 (ショット数) は必要な精度に応じて調整することができるが, 図示したものは全てサンプル数が 512 の結果である.

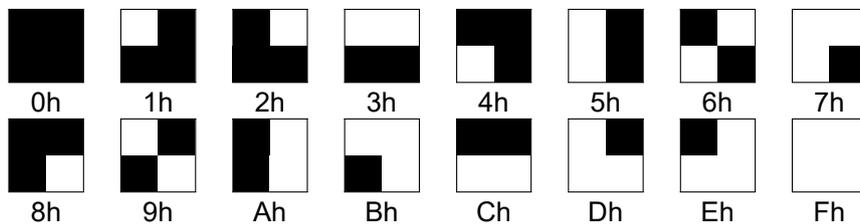


図 4.5: 4 pixel-binary data (“h”は 16 進数の意). Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

4.3.5 結果と考察

振幅増幅前

まずはじめに、振幅増幅を行わない場合 ($t = 0$) のインデックスの確率分布を見てみる (図 4.6) . 横軸に示されているのは、式 (4.20) における $P(\text{index} = k) = |\langle k | \langle 0^{n_D} | \Psi \rangle|^2$ で与えられる経験的確率をサンプリングに依って計算した値であり、縦軸に示されているのは、データベースのインデックスと格納されているデータである (index: 0h,1,2,3,4,5,6,7)=(data: 0h,2h,4h,6h,8h,Ah,Ch,Eh). ここで縦軸のうち、“Other”と示されているものは、 $|\Psi\rangle$ を $|0\rangle^{\otimes n_D}$ 以外の状態へ射影した際の確率の総和、すなわち post-selection で棄却された失敗確率である. 棒グラフの意味するところは、青色 (上)、灰色 (中)、橙色 (下) がそれぞれ、ケース (i), (ii), (iii) に対応している. グラフ中に挿入されている小画像は、入力したクエリデータに対応している.

これらの結果から、AAE とパターンマッチアルゴリズムが、両方ともよく機能している様子が見てとれる. すなわち、(i) で示されるシミュレーションの結果、(ii) で示される実機を用いた結果が、(iii) の厳密埋め込みを用いた理論的な結果によく一致している. 特に、クエリと一致したデータベース中の要素に対応するインデックスが最も高い確率で観測される状況 (図中では “match” と表現) が、期待通りに得られている. これは式 (4.20) で示されるように、データベース要素とクエリの重なり (Fidelity; 最大 $1/N_T$) が類似度として適切に算出されていることを示している. 特に本問題設定においては、観測される確率値が 2 つの画像の Hamming 距離に対応していることを指摘しておく. 従って、実用的な場面を想定すると、ユーザーは予め画像同士が “一致” と判断するための閾値を Hamming 距離に基づいて設定をしておく. その値を用いて、本アルゴリズムの結果として得られた値を選別すればよい. 例えば、図 4.6 において、クエリとして “2h” を入力した場合を考えると、解の候補として (0h, 2h, 6h, Ah) が得られるので、その中から最終的な確認を行うというフローを持ちれば、アルゴリズムに必要な精度を落とすこともでき、サンプル数の観点から効率性を向上させることもできる. 更に、仮にデータベースがクエリと一致するデータを含んでいなかった (図中では “match” が無い) としても、Hamming 距離に応じて解の候補になり得る点を指摘しておく (これは図中では “closest” と示している). また、上述の通り、(ii) 実機に依る実験結果と (iii) 理論からの結果は一致しているが、図 4.6 の様子を見ると、実デバイスにおけるデコヒーレンスの影響が (ii) には現れていることが分かる. 特に (ii) では (i) に比べると射影の結果 $|0\rangle^{\otimes n_D}$ となったものの確率、すなわち対応したインデックスに関する情報量を持ったものが低くなりがちである点も見て取れる. すなわち、デコヒーレンスにより情報量を失い、結果として “others” の状態として観測されたと解釈できる. これらは量子状態の振幅が一様に劣化する、脱分極ノイズ (depolarization noise) によるものである.

振幅増幅後

次に、同じ設定のものに対して振幅増幅を適用した結果に目を向ける. 図 4.7 に示したものが、Grover 演算を $t = 5$ 回作用させた結果である. 尚、グラフの見方は図 4.6 と同様である. 但し、ノイズによる影響が大きく有意な情報が取れなかったため、実デバイスの結果は載せていない. 結果を見ると、失敗確率である “others” の値が図 4.6 に比べて大きく減少し、post-select の成功確率が向上しており、アルゴリズムが適切に作用していることが分かる. 結果として、式 (4.14) で表されるように、 ψ_x に対応する各係数に $1/\langle q | \Psi \rangle$ が加わった相対的關係を維持したまま、成功確率が全体的に増幅された. 全体として、AAE によってエンコードされたもの (青色の上段のバー) と Toffoli gate によって厳密埋め込みされたもの (橙色の下段のバー) もよく一致しているものの、一部では違いが現

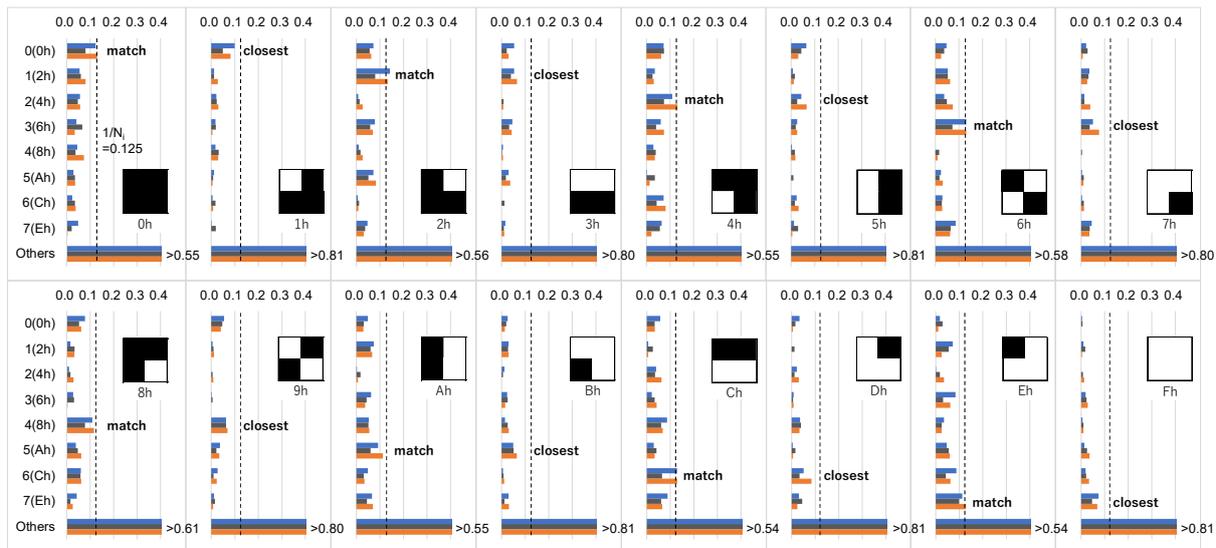


図 4.6: 振幅増幅前のパターンマッチの結果. 縦軸は, データベースの index (index 0,1,2,3,4,5,6,7)=(data 0h,2h,4h,6h,8h,Ah,Ch,Eh). 棒グラフは, 観測確率 $P(\text{index} = k) = |I(k)|\langle 0^{nD} | \Psi \rangle|^2$ を表す. 色は, 青 (上) がケース (i), 灰色 (中) がケース (ii), 橙 (下) がケース (iii) を表す. “Others” は, 失敗確率, すなわち $|0\rangle^{\otimes nD}$ 以外の状態に $|\Psi\rangle$ が射影された確率の合計を表す. “match” は, データベースの中で Query と一致しているものの index を示し, “closest” は, 一致はしていないが Query に最も近いものに対応する index を表す. 図中のインセットは Query データを表す. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

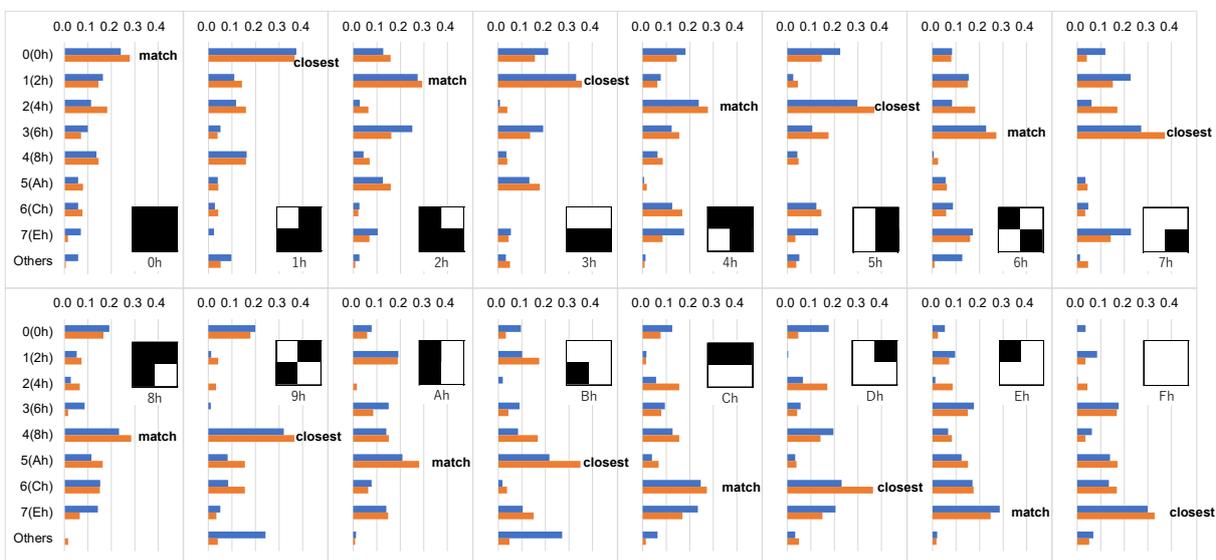


図 4.7: 振幅増幅後のパターンマッチの結果 (iteration 数 $t = 5$). 色は, 青 (上) がケース (i), 橙 (下) がケース (iii) を表す. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

れている. これは AAE が機械学習的な手法によってエンコーダに対応するパラメタを探索している

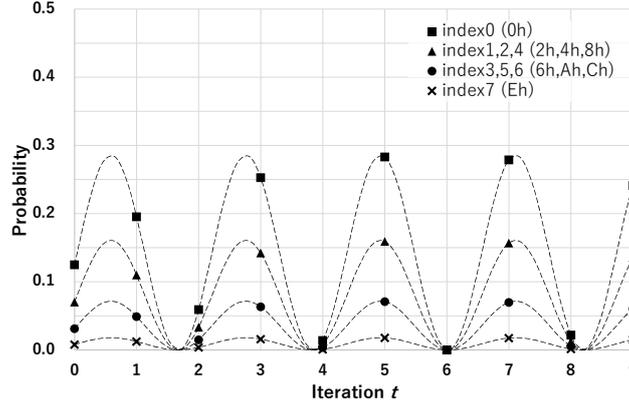


図 4.8: 振幅増幅後の観測確率と iteration 数の関係. プロットは数値実験の結果を表し, 破線は式 (4.14) による理論曲線を表す. データベースは図 4.6, Query データは “0h”である. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

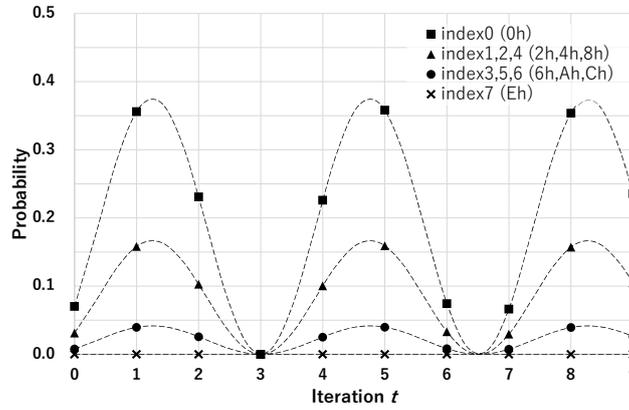


図 4.9: 振幅増幅後の観測各地ると iteration の関係. Query データが “1h”であり, それ以外の設定は Fig. 4.8 と同様. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

ことに起因しており, 有限の誤差が生じている様子を示している. このような誤差が最終的なタスクにどの程度影響するかは, 4.3.7 項で例を出して検討する.

続いて, Grover 演算の回数 t と式 (4.14) で表されるインデックス k に対応する状態が観測する確率の関係について議論する. 例として 0h と 1h に関する結果を図 4.8, 4.9 に示した. 2つの結果を見比べて分かるのが, 観測確率の振動の振幅と周波数が異なる点である. これは式 (4.14) 中の $\langle q|\Psi\rangle$ を反映している. 4.2.2 項で指摘した通り, 最適なイテレーション数 t_* を正確に推定するには, $\langle q|\Psi\rangle$ を正確に推定する必要がある. このタスクは, 式 (4.14) で与えられる振動の周波数 $\omega = 2 \arcsin \langle q|\Psi\rangle$ が高い場合には困難である. しかし, N_I が大きくなるような実用的な問題設定では, ω は小さな値となり, 式 (4.14) で表される確率は t に対して単調増加する. この結果, t としてラフな値を選択しても十分な振幅増幅の効果が得られることが期待できる. 別の特徴として, 8つのインデックスが4つのグループに分かれて振動していることが見て取れる. これはクエリとデータベースの各要素の間の Hamming 距離 (HD) を反映している. すなわち, それぞれ確率の高いグループから順に, 図 4.8 では $\text{HD}=0,1,2,3$, 図 4.9 では $\text{HD}=1,2,3,4$ に対応している. この結果は, 入力がバイナリ画像であることにも起因するが, パターンマッチングの文脈からは至って自然である.

4.3.6 回路深さと multi qubit gate の数の比較

本項では、提案アルゴリズムを実装する際の量子回路のサイズについて議論する。まず、AAE を用いることで従来のデータ埋め込み手法に比べて劇的に回路深さと multi qubit gate の数を削減できる点を思い出す。前述のデモンストレーションで用いた量子回路はデータベース状態を 6-qubit で構成したが、生成回路は 6 層の HEA とし合計 30 個の CNOT gate を用いた。これは従来の厳密埋め込みの手法と比較すると歴然とした差がある。6-qubit 系に Toffoli gate を用いて状態埋め込みを行うと、32 個の 6-qubit Toffoli gate が必要となる。これは 128 個の CNOT gate と 96 個の auxiliary qubit に相当する [89]。別のより効率的な手法 [79] によれば、長距離相互作用も用いれば、任意の n -qubit gate は $(2^n - n - 1)$ 個の CNOT gate により $|0 \cdots 0\rangle$ から生成できるとされている。すなわち、 $n = 6$ の場合は、57 CNOT gate になる。しかし、超電導量子デバイスを想定すると、CNOT gate が作用できるのは隣接 qubit のみである。従って、合計の CNOT gate の数を [89] に示されている

$$\frac{10}{3}2^n + 2n^2 - 12n + \begin{cases} 14/3, & n \text{ even,} \\ 10/3, & n \text{ odd,} \end{cases}$$

に基づいて換算すると、 $n = 6$ では 218 CNOT gate となる。以上から、厳密埋め込みは qubit 数に対して指数的に多くの CNOT gate を必要とするため、大規模系でデータロードをするには大きな実装上の困難を伴うことが分かる。

一方、データロードに PQC を用いる場合は、精度保証が厳密には困難な点を受け入れれば、深さと multi qubit gate の数ともそれほど巨大にはならない。具体的には、 n -qubit 系に対し、深さ $O(\text{poly}(n))$ 、multi qubit gate 数 $O(n)$ 程度である。従って、今回は実装性に重きを置き、データベース状態の生成にそれほど精度を要求されない問題設定を扱った。実際、パターンマッチ問題はクエリとデータベースに格納されたデータの相対的な類似度が比較できれば構わないため、シミュレーション結果が示した通り、浅い回路でも十分動作が確認できた。次に PQC の構成についてももう少し考えてみる。本研究ではシンプルな HEA を取り扱ったが、実用を想定すると、 $O(\text{poly}(n))$ の深さの PQC で必要な精度の埋め込みが可能な回路を設計するための指針が必要である。例えば、[90, 91] に示されているような、データベース構造をふまえた ansatz 設計方法などが重要になる。また、機械学習によりエンコーダを生成することから、変分法を用いた学習過程における勾配消失問題（いわゆる barren plateau 問題 [92]）を回避することも必要である。尚、これには既にいくつかの回避方法が提案されてきている。例として、回路の初期化を工夫する方法 [93]、特別な ansatz 構造を用いること [94]、パラメタの埋め込み方を工夫する方法 [95] などがある。

最後に Grover 演算子について記す。本研究ではナイーブな Grover 演算子の構成を採用したが、特に Diffusion 演算子 G_d の実装方法については、実装時の gate 数や深さを抑えるための、様々な変形例が提案されている。例えば、[96, 97, 98] や最近提案されたものとしては [99, 100, 101] がある。これらを活用することで振幅増幅まで含めた実機デバイスをを用いたデモンストレーションが可能となるだろう。

4.3.7 エンコーディング誤差の影響確認

提案アルゴリズムでは、機械学習的な学習過程を経て PQC で構成されたエンコーダのパラメタを探索することを想定する。この場合、学習の最終状態でも有限の誤差が生じる可能性がある。また、実機による実行を想定すると、NISQ を使用する限りには、有限のサンプル数に起因するショットノイズやデコヒーレンスに起因する脱分極ノイズなど、多くの誤差要因が存在する。一般に、ある情報を符号化・エンコードすることを考えると、どのような符号化を施すかによって、耐ノイズ性能

が大きく異なる。本研究では、量子画像をモチーフとしているが、量子画像のエンコード方法として (i)FRQI (Amplitude encoding ベース), (ii)NEQR (Basis encoding ベース) の2つを紹介した。本項では、これらの2つのエンコード方法におけるノイズへの感度について、数値シミュレーションを用いて確認する。

実験設定

ノイズに対する感度を調べるための入力として“手書き数字のデータ”[102]を用いた。具体的には0から7までの8種類の1桁の数字が、 8×8 -pixelの領域に16階調の輝度情報の集合として与えられる状況を考える。Amplitude encoding と Basis encoding を行うことを考えると、各画素位置 x の輝度情報 u_x は、 $u_x^{AE} \in [0, 1]$, もしくは $u_x^{BE} \in \{0, 1\}^{\otimes 4}$ のように与えられる。画像一枚の量子状態は、全画素の輝度の重ね合わせて表現される。各エンコード方法に対応する量子状態は4.3.1項に示したが、各エンコード方法を用いた場合にインデックス x が観測される確率分布 $P(\text{index} = x)$ を、後の解析のために明示的に記しておく。 N_I はデータベースに格納されている画像の枚数であり、 N_P は画素数である。従って、実験設定では $N_I = 8$, $N_P = 64$ である。

(i)FRQI (Amplitude encoding ベース) を用いると、

$$P_{FRQI}(\text{index} = x) = \frac{1}{N_I} \left(\frac{N_p^{FRQI}(x)}{N_p} \right)^2,$$

$$N_p^{FRQI}(x) = \sum_{j=0}^{N_p-1} \left(\cos \theta_j \cos \theta_{jx} + \sin \theta_j \sin \theta_{jx} \right),$$

と表される。尚、 θ_j はクエリ画像の j 番目の pixel の輝度に対応し、 θ_{jx} はデータベース内の x 番目の画像の j 番目の pixel の輝度に対応する。

(ii)NEQR (Basis encoding ベース) を用いると、

$$P_{NEQR}(\text{index} = x) = \frac{1}{N_I} \left(\frac{N_p^{NEQR}(x)}{N_p} \right)^2,$$

$$N_p^{NEQR}(x) = \sum_{j=0}^{N_p-1} \otimes_{i=0}^{q-1} \langle c_j^i | c_{jx}^i \rangle,$$

表される。尚、 c_j^i はクエリ画像の j 番目の pixel の輝度を表すビット列のうち i 番目の値に対応し、 c_{jx}^i はデータベース内の x 番目の画像の j 番目の pixel の輝度を表すビット列のうち i 番目の値に対応する。

今、各 u_x に対してノイズ ϵ_x が加わる状況を考える。但し、ノイズの分布は正規分布 $N(0, \sigma)$, $\sigma = \max(\{u_x\})\sigma_0$ を仮定する。この下でノイズが印加された状態の振幅を正規化するには、ノイズ印加後の振幅 $\{u_x + \epsilon_x | x \in \text{all basis}\}$ を $\sqrt{\sum_x (u_x + \epsilon_x)^2}$ で割ればよい。以上の操作をデータベース要素とクエリの全てに対して行った。その後、異なるノイズの大きさ $\sigma_0 \in \{0.05, 0.1, 0.3, 0.5\}$ について、各インデックスが観測される確率分布を計算することでノイズの影響を確認した。但し、簡単のため、振幅増幅前の状態を調べた。振幅増幅後の状態を考えるには、各振幅の大きさの相対関係を保ったまま各振幅を $1/\langle q | \Psi \rangle$ 倍すればよい。

結果と考察

得られた結果を図4.10, 4.11に示す。FRQIとNEQR共に正しいインデックスが高確率で観測されており、適切に機能していることがわかる。識別性の観点からは、NEQRの方が優れており、match

したインデックスとそれ以外のインデックスの違いがはっきりと見て取れる。しかしながら、類似度、すなわちクエリと格納されているデータがどの程度似ているかというアナログ的な情報を読み取るという観点からは、適しているとは言えない。言い換えると、FRQI はファジーなマッチングを見るのに適しており、NEQR は完全一致の判別が目的の場合に適している。例えば、図 4.10 のスコアに基づくと “3” と “5” はよく似ていると判断することができる。（クエリ “3” におけるインデックス “5” と、クエリ “5” におけるインデックス “3” の比較）

次にエラー耐性について考えると、NEQR の方が優れていると言える。FRQI では、ノイズの大きさが 10% 以下ではスコアに大きな影響は見られない。一方で NEQR では、5% 程度のノイズで無視できない影響が見られる。さらに 30% 程のノイズ下ではスコアが劇的に小さくなってしまっている。これらの挙動は各フォーマットの特徴から理解できる。FRQI では、輝度情報を階調数によらず 1-qubit で表現 ($\cos \theta|0\rangle + \sin \theta|1\rangle$) しており、一般的な画像では多くの画素が何らかの輝度情報を持っている（白飛びや黒つぶれは稀）、すなわち殆どの基底が non-zero の振幅を持っている。いま、エンコード時の誤差がランダムに生じると仮定するならば、それらの効果は互いに干渉し、補償し合うと考えられる。しかしながら、一方の NEQR では、振幅情報は非常にスパースである。すなわち、例えば輝度の階調が 16 階調だった場合を考えると、輝度を表現する基底は 16 個あるが、non-zero になるのはわずか 1 つであり、残りの 15 個は 0 となる。このようにほぼ one-hot ベクトルとなっているもの同士の内積から類似度を計算するため、比較的エンコード時の誤差への感度は高いのである。

4.4 結論

第 4 章では、有名かつサブルーチンとして重要な量子アルゴリズムである Grover’s algorithm に着目し、実装を考える際に障害となる Oracle（ブラックボックス回路）を明示的に実装する手法について記した。合わせて、量子優位性を得る上でボトルネックとなる初期状態準備に必要な time complexity を、AAE の応用により指数関数的に改善した。これにより、NISQ における実現に一步近付いただけでなく、FTQC の時代における有用性を拡張した。

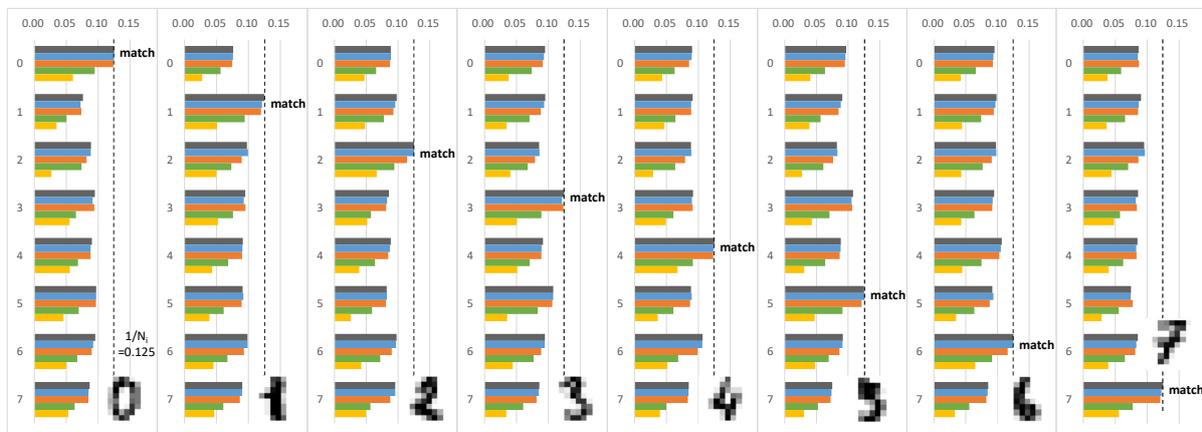


図 4.10: エンコード時にノイズを与えてパターンマッチを行った結果 (FRQI) . 縦軸はクエリ, 横軸は類似度, インセットは正解の数字を表し, 色違いはノイズの大きさの違いに対応している. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

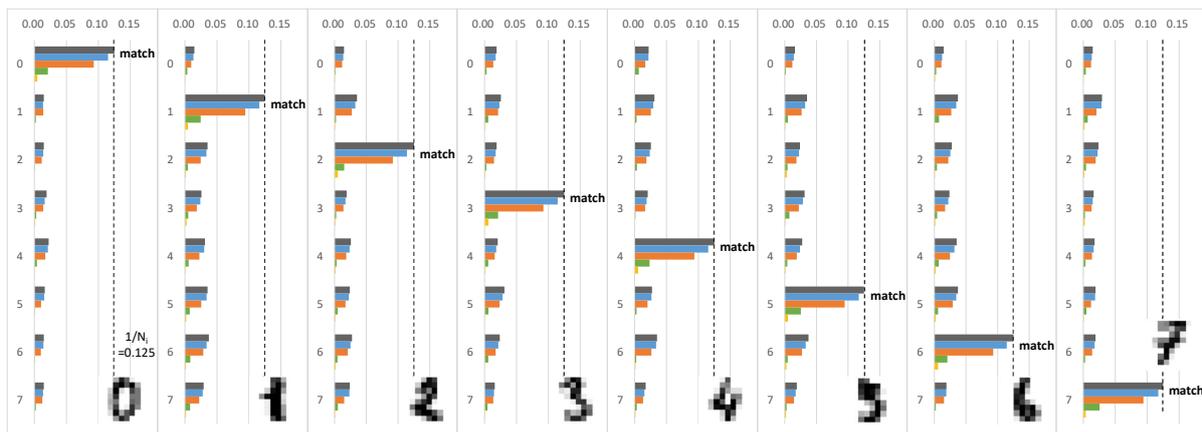


図 4.11: エンコード時にノイズを与えてパターンマッチを行った結果 (NEQR) . 縦軸はクエリ, 横軸は類似度, インセットは正解の数字を表し, 色違いはノイズの大きさの違いに対応している. Reprinted figure from [44]. Copyright ©2022 by American Physical Society. All rights reserved.

第5章

学習の効率化：NTK を利用した量子-古典ハイブリッドニューラルネットワーク

第5章では、量子-古典ハイブリッドニューラルネットワーク (Quantum classical neural network; qcNN) において、効率的に学習を行うフレームワークを提案する [45]. 大規模な FTQC の実現まで依然として時間がかかるということを鑑みると、NISQ を用いて如何に高度な処理ができるかを考える必要がある. その際に欠かせない視点が、量子回路・量子計算に加えて如何に古典計算機を活用するかである. 本研究の着眼点は、目下様々なタスクにおいて非常に強力な性能を示している古典ニューラルネットワーク (Classical neural network; cNN), 特に深層学習の技術・ノウハウを最大限に活用しつつ、量子ニューラルネットワーク (Quantum neural network; qNN) のもつ潜在的な優位性活かす、という点である. 後述するように、既存の研究でも量子と古典を組み合わせるアプローチは存在しているものの、解決すべき問題点がある. 本研究では、それらの問題点を解決しつつ、理論と数値実験の両面から新たなフレームワークの有用性を示す.

5.1 先行研究とそれらの問題点

これまで、qNN や qcNN を用いた研究が盛んになされ、様々なタスクで古典手法よりも高い性能を示す可能性が示唆されてきた [103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 45, 115, 116]. しかしながら、以下の2つの課題がある. 1つ目の課題が、既存の qNN や qcNN は、学習過程の結果として“良い”解へ収束する理論的な保証が無いという点である. 特に深刻なのが勾配消失 (barren plateau) 問題 [117] である. これは qubit 数が増加するにつれて、指数関数的に学習時の勾配ベクトルの大きさが小さくなるというものである. これに対し、部分的な解決策は様々なものが提案されているが [118, 93, 119, 120, 121, 94, 122, 123, 124, 125], あらゆる状態に対応できる一般的な解決策は未だない. 2つ目の課題が、qNN のもつ表現力の高さに由来する潜在的な優位性が、古典手法に比べて良い解を示すことに対応するという点に理論保証が無い点である. この点に関しては、最近の研究 [126] により、量子カーネル法の文脈において、広いクラスの古典手法に比べて高い性能を示す条件が明らかになってきた. それに基づき、projected quantum kernel (PQK) というアイデアが生まれ、優位性を示す条件に合致している可能性があることが分かった. ここで注意すべきは、量子カーネル法は既に様々な研究 [127, 128, 129, 130, 131, 132, 133, 134] が理論と実験の両面からなされているにも関わらず、PQK も含めて、効果的な量子カーネルの設計指針が大いに非自明である点である. また、量子・古典を含めカーネル法の計算には N_D 個のデータを扱うには $O(N_D^2)$ の computational cost がかかる点も、実用を踏まえると大きな問題点である. 一方の cNN は、学習コストが N_D に対してスケールしないと仮定すれば、 $O(N_D)$ ですむ. 従って、前述した量子カーネル

法の優位性を示すような qNN・qcNN を、学習が容易な形で実現できれば、実用上大変好ましい。

5.2 アプローチ：NTK 理論の導入と提案モデルのアウトライン

cNN, 特に深い cNN において, 学習過程の収束特性を含む, 基本的な性質を解析する強力なアプローチに Neural tangent kernel (NTK) 理論 [135] がある. 端的に言えば, NTK は cNN の学習プロセスにおける出力関数のダイナミクスを技術する方程式を支配するような, 関数空間におけるカーネルである. NTK は次の特筆すべき性質をもつ. まず, cNN においてノード数が大きい極限 (幅が無間大に相当) をとると時不変 (time-invariant) になる. さらには, パラメタの初期値をランダムに選択することで正定になる. その結果, 例えば最小二乗誤差に基づく回帰問題を考えると, 学習プロセスは線形の微分 (もしくは差分) 方程式で記述できる. そして, 学習過程の解析は時間不変の正定行列のスペクトルの解析に還元することができる.

本研究では, NTK 理論で直接解析できるような qcNN のクラスに着目しつつ, 以下のようなフレームワークを提案する. まず, 古典データを量子状態として qNN にエンコードする. 次に qNN の出力状態へ適切な乱択測定を行うことで, 量子カーネルによって特徴量空間へ射影された特徴情報を, 改めて古典データとして抽出する. 得られた古典ベクトルを今度は cNN へ入力し, 適切な処理をする. すなわち, タスクに合わせたコスト関数を定義, 評価し, 最小化することで, この cNN を目的のタスクに合わせて学習させる. ポイントは量子の部分は固定されており, 勾配消失の心配が無い点である. 結果として, 前項で挙げた 2 つの課題を解決しうるフレームワークが得られる.

参考までに本研究に関連のある文献を以下に挙げておく. ガウシアンプロセスとの関係 [136], NTK スペクトルと cNN の収束特性の関係 [137], そして, 分類問題の場合の NTK [138, 139, 140, 141].

5.3 関連研究

量子 NTK 理論に関する最近の研究について, 以下に簡単に解説する. それらは, qNN の出力状態に関するコスト関数を NTK として定義するものである. まず [142] では, NTK が時間に対して大きな変化がないという仮定のもとで得られるコストに関して, 線型微分方程式 (後述する式 (5.5) に相当) の特性を詳細に調べている. これは後続研究 [143] にも繋がるコンセプトである. 論文では, 理論解析と数値実験を用いて, パラメタの数が非常に多い時, すなわち系が over-parameterization regime にある場合, コスト関数が指数関数的に減少するというダイナミクスが示されている. この挙動は [144] でも数値解析の結果として示された. また, [145] では, NTK と勾配消失の関係について論じている. すなわち, 上記のような NTK の収束性を実現するためには, qubit 数 n に対して, $O(4^n)$ のパラメタが必要となり, 勾配消失と同じような現象が起こると主張している. それに対し [146] では, そのような大きな要求に対しての処方箋を示した. 具体的には, 2^n の次元をもつヒルベルト空間があったとしても, その有効次元が d_{eff} であれば, $O(d_{\text{eff}}^2)$ のパラメタで学習過程のダイナミクスが指数関数的な収束を示すことが保証された. これら全ての研究は, パラメタ化された量子系に関する議論である. 一方で, 本研究では量子-古典ハイブリッドの系を扱うものであり, チューニングパラメタを古典回路部分に限定している. すなわち, 取り出される NTK もそれらのパラメタに対するものである. その違いは得られる結果に対して大きな影響を与える. 本研究における NTK は時不変であり (定理 5.5), 出力関数は over-parameterization regime において Gaussian になる (定理 5.3, 5.4). このような系の性質は上述の関連研究では触れられていない. 特に, NTK が時不変であるという性質は出力関数の指数関数的な収束性を保証するためには重要である. 上述の研究ではこの点を “仮定” として与えた上で議論を行っている. 本研究における NTK は, ともすると純粋な古典オブジェクトに聞こえるかもしれないが, 決してそうではない. 5.1 節で触れた projected quantum kernel によ

て構成される非線形関数の形で、量子回路部分の特徴量が NTK に現れるのである。その上、勾配消失も起こらないという点を付言しておく。

5.4 NTK 理論

はじめに NTK 理論が提案されたのは [135] である。これは最急降下法ベースの学習過程における、無限幅の cNN のダイナミクスを解析する手法である。特に、NTK 理論は「なぜパラメタ数がデータよりも大幅に多い領域、すなわち over-parameterization regime において、cNN がよく機能するのか。具体的には、様々な機械学習タスクにおいて学習誤差の観点で優れているのか。」を説明することができる。5.4 節では、NTK 理論についてレビューする。合わせて、重要な観点として、どのような cNN では NTK 理論が破綻するかについても記す。本研究で導入するモデルは、NTK の観点から cNN の破綻点を議論した際の処方箋として提案するものである。

5.4.1 問題設定

NTK 理論では教師あり学習に着目する。ここでは N_D 個の教師データとして (\mathbf{x}^a, y^a) ($a = 1, 2, \dots, N_D$) を考える。尚、 \mathbf{x}^a は入力ベクトル、 y^a は出力値である。ここでは簡単のために出力値 y^a としてスカラーを仮定しているが、元々の NTK 理論ではベクトル出力の場合を扱っている。仮定として、データセットには以下の（隠れた）関係性があるものとする。

$$y^a = f_{\text{goal}}(\mathbf{x}^a), \quad \forall a.$$

本問題の目的はモデル $f_{\theta(t)}$ (cNN の出力に対応) が何らかの指標の下で f_{goal} に近づくように学習を行うことである。尚、 $\theta(t)$ はパラメタ更新の iteration が t の時の学習対象のパラメタである。 $f_{\theta(t)}$ と f_{goal} の違いを測る指標としては、主に回帰問題に用いられる平均二乗誤差 (Mean squared error; MSE)

$$\mathcal{L}_t^C = \frac{1}{2} \sum_{a=1}^{N_D} (f_{\theta(t)}(\mathbf{x}^a) - f_{\text{goal}}(\mathbf{x}^a))^2 = \frac{1}{2} \sum_{a=1}^{N_D} (f_{\theta(t)}(\mathbf{x}^a) - y^a)^2, \quad (5.1)$$

や、主に分類問題に用いられるバイナリクロスエントロピー (Binary cross entropy; BCE)

$$\mathcal{L}_t^C = - \sum_{a=1}^{N_D} (y^a \log \sigma_s(f_{\theta(t)}(\mathbf{x}^a)) + (1 - y_a) \log \sigma_s(f_{\theta(t)}(\mathbf{x}^a))), \quad (5.2)$$

がある。尚、 σ_s はシグモイド関数、 y^a は 0 か 1 をとるバイナリラベルである。

いま、関数 $f_{\theta(t)}$ は L 層の全結合ネットワークで構成されているとし、 n_ℓ を第 ℓ 層目のノード数 (幅) とする。(但し、 $\ell = 0, L$ はそれぞれ、入力層と出力層とする。) すると、入力 \mathbf{x}^a は以下に示す関係に従って、出力 $f_{\theta(t)}(\mathbf{x}^a)$ に変換される。

$$\begin{aligned} \boldsymbol{\alpha}^{(0)}(\mathbf{x}^a) &= \mathbf{x}^a, \\ \boldsymbol{\alpha}^{(\ell)}(\mathbf{x}^a) &= \sigma(\tilde{\boldsymbol{\alpha}}^{(\ell)}(\mathbf{x}^a)), \\ \tilde{\boldsymbol{\alpha}}^{(\ell+1)}(\mathbf{x}^a) &= \frac{1}{\sqrt{n_\ell}} W^{(\ell)} \boldsymbol{\alpha}^{(\ell)}(\mathbf{x}^a) + \xi b^{(\ell)}, \\ f_{\theta(t)}(\mathbf{x}^a) &= \tilde{\boldsymbol{\alpha}}^{(L)}(\mathbf{x}^a), \end{aligned}$$

尚、 $W^{(\ell)} \in \mathbf{R}^{n_\ell \times n_{\ell-1}}$ は重み行列、 ξ はバイアスの調整パラメタ、 $b^{(\ell)} \in \mathbf{R}^{n_\ell}$ は第 ℓ 層目のバイアスベクトル、 σ は微分可能な活性化関数である。ここで、調整パラメタベクトル $\theta(t)$ が $\{W_{jk}^{(\ell)}\}$ と $b^{(\ell)}$ の

全要素で構成されている点に注意する。最急降下法を用いてパラメタ更新を行うと、

$$\frac{\partial \theta_j(t)}{\partial t} = -\eta \frac{\partial \mathcal{L}_t^C}{\partial \theta_j} = -\eta \sum_a \frac{\partial f_{\theta(t)}(\mathbf{x}^a)}{\partial \theta_j} \frac{\partial \mathcal{L}_t^C}{\partial f_{\theta(t)}(\mathbf{x}^a)},$$

と表される。尚、簡単のために連続時間を仮定した。また、 η は学習率、 θ_j は j 番目のパラメタである。全てのパラメタ $\{W_{jk}^{(\ell)}\}$ と $b^{(\ell)}$ は、互いに独立したガウス分布からサンプルされた値で初期化されるとする。

5.4.2 NTK の定義

NTK は出力関数 $f_{\theta(t)}$ のダイナミクスの中に以下のように現れる。 $f_{\theta(t)}$ の時間微分を計算すると、

$$\begin{aligned} \frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial t} &= \sum_j \frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial \theta_j} \frac{\partial \theta_j}{\partial t} \\ &= -\eta \sum_{j,b} \frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial \theta_j} \frac{\partial f_{\theta(t)}(\mathbf{x}^b)}{\partial \theta_j} \frac{\partial \mathcal{L}_t^C}{\partial f_{\theta(t)}(\mathbf{x}^b)} \\ &= -\eta \sum_b K^{(L)}(\mathbf{x}, \mathbf{x}^b, t) \frac{\partial \mathcal{L}_t^C}{\partial f_{\theta(t)}(\mathbf{x}^b)}, \end{aligned} \quad (5.3)$$

となる。ここで、 $K^{(L)}(\mathbf{x}, \mathbf{x}', t)$ を

$$K^{(L)}(\mathbf{x}, \mathbf{x}', t) = \sum_j \frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial \theta_j} \frac{\partial f_{\theta(t)}(\mathbf{x}')}{\partial \theta_j}.$$

と定義した。この $K^{(L)}(\mathbf{x}, \mathbf{x}', t)$ が NTK と呼ばれる。以下では、 $f_{\theta(t)}$ の軌跡が無限幅の極限 ($n_1, n_2, \dots, n_{\ell-1} \rightarrow \infty$) において、NTK を用いて解析的に計算できることを見ていく。

5.4.3 定理

NTK の最も重要な特徴が、NTK が幅無限大の極限において、時不変かつ半正定値な関数 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ に収束するという点である。以下でこの性質について記すが、その前に $f_{\theta(0)}$ に関する補題を記す。

補題 5.1 (Proposition 1 in [135]) σ を Lipschitz な非線形関数としたとき、幅無限大の極限 $n_\ell \rightarrow \infty$ ($1 \leq \ell \leq L-1$) において、初期化直後の出力 $f_{\theta(0)}$ は、共分散行列 $\Sigma^{(L)}(\mathbf{x}, \mathbf{x}')$ が以下の再帰的な関係式で与えられるような Centered Gaussian process に従う。

$$\begin{aligned} \Sigma^{(1)}(\mathbf{x}, \mathbf{x}') &= \frac{1}{n_0} \mathbf{x}^T \mathbf{x}' + \xi^2, \\ \Sigma^{(\ell+1)}(\mathbf{x}, \mathbf{x}') &= \mathbf{E}_{h \sim \mathcal{N}(0, \Sigma^{(\ell)})} [\sigma(h(\mathbf{x})) \sigma(h(\mathbf{x}'))] + \xi^2, \end{aligned} \quad (5.4)$$

但し、期待値は共分散 $\Sigma^{(\ell)}$ をもつ Centered Gaussian process h について計算する。

証明は [135] の Appendix A.1 に記載されているが、注目すべきは任意の関数 $z(h(\mathbf{x}), h(\mathbf{x}'))$ の期待値が、

$$\mathbf{E}_{h \sim \mathcal{N}(0, \Sigma^{(\ell)})} [z(h(\mathbf{x}), h(\mathbf{x}'))] = \frac{1}{2\pi \sqrt{|\tilde{\Sigma}^{(\ell)}|}} \int d\mathbf{h}(\mathbf{x}) d\mathbf{h}(\mathbf{x}') \exp\left(-\frac{1}{2} \mathbf{h}^T \left(\tilde{\Sigma}^{(\ell)}\right)^{-1} \mathbf{h}\right) z(h(\mathbf{x}), h(\mathbf{x}')),$$

と計算できる点である。但し、 $\tilde{\Sigma}^{(\ell)}$ は 2×2 の行列であり、

$$\tilde{\Sigma}^{(\ell)} = \begin{pmatrix} \Sigma^{(\ell)}(\mathbf{x}, \mathbf{x}) & \Sigma^{(\ell)}(\mathbf{x}, \mathbf{x}') \\ \Sigma^{(\ell)}(\mathbf{x}', \mathbf{x}) & \Sigma^{(\ell)}(\mathbf{x}', \mathbf{x}') \end{pmatrix},$$

また、ベクトル \mathbf{h} は $\mathbf{h} = (h(\mathbf{x}), h(\mathbf{x}'))^T$ を表し、 $|\tilde{\Sigma}^{(\ell)}|$ は行列 $\tilde{\Sigma}^{(\ell)}$ の行列式である。

そして、補題 5.1 から NTK に関する以下の定理が導かれる。

定理 5.1 (Theorem 1 in [135]) σ を Lipschitz な非線形関数としたとき、幅無限大の極限 $n_\ell \rightarrow \infty$ ($1 \leq \ell \leq L-1$) において、Neural tangent kernel $K^{(L)}(\mathbf{x}, \mathbf{x}', t)$ は、以下の再帰的な関係式に従って時不変な関数 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ に収束する。

$$\begin{aligned} \Theta^{(1)}(\mathbf{x}, \mathbf{x}') &= \Sigma^{(1)}(\mathbf{x}, \mathbf{x}') = \frac{1}{n_0} \mathbf{x}^T \mathbf{x}' + \xi^2, \\ \Theta^{(\ell+1)}(\mathbf{x}, \mathbf{x}') &= \Theta^{(\ell)}(\mathbf{x}, \mathbf{x}') \dot{\Sigma}^{(\ell)}(\mathbf{x}, \mathbf{x}') + \Sigma^{(\ell+1)}(\mathbf{x}, \mathbf{x}'), \end{aligned}$$

但し、 $\dot{\Sigma}^{(\ell)}(\mathbf{x}, \mathbf{x}') = \mathbf{E}_{h \sim \mathcal{N}(0, \Sigma^{(\ell)})} [\dot{\sigma}(h(\mathbf{x})) \dot{\sigma}(h(\mathbf{x}'))]$ 、 $\dot{\sigma}$ は σ の微分である。

ここで、定義から行列 $(\Theta^{(L)}(\mathbf{x}^a, \mathbf{x}^b))$ は対称かつ半正定であることに注意すると、以下が言える。

定理 5.2 (Proposition 2 in [135]) σ を Lipschitz な非線形関数としたとき、 $L \geq 2$ かつ入力ベクトル \mathbf{x} が $\mathbf{x}^T \mathbf{x} = 1$ と正規化されていればカーネル $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ は正定である。

幅が無限大の NTK に関するこれらの定理を用いて $f_{\theta(t)}$ の軌跡を解析できることを、次節で示す。

定理 5.1, 5.2 からの帰結

定理 5.1, 5.2 を用いると、幅が無限大の極限では、微分方程式 (5.3) は、

$$\frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial t} = -\eta \sum_b \Theta^{(L)}(\mathbf{x}, \mathbf{x}^b) \frac{\partial \mathcal{L}_t^C}{\partial f_{\theta(t)}(\mathbf{x}^b)}. \quad (5.5)$$

と書き直せる。この方程式の解は \mathcal{L}_t^C の表式によって異なるものになる。特に重要なケースはコスト関数 \mathcal{L}_t^C が平均二乗誤差 (MSE) の場合だろう。本論文の表式 (5.1) を用いれば、MSE の場合に対応する微分方程式は

$$\frac{\partial \mathcal{L}_t^C}{\partial f_{\theta(t)}(\mathbf{x}^b)} = f_{\theta(t)}(\mathbf{x}^b) - y^b, \quad (5.6)$$

と表される。これに式 (5.5) を代入することで常微分方程式が得られる。この方程式は [147] にあるような一般的な手法を用いることで、以下のように解析的に解くことができる。

$$f_{\theta(t)}(\mathbf{x}^a) = \sum_{j,b} V_{aj}^T (V_{jb} f_{\theta(0)}(\mathbf{x}^b) - V_{jb} y^b) e^{-\eta \lambda_j t} + y^a, \quad (5.7)$$

但し、 $V = (V_{jb})$ は直交行列であり、 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ を

$$\sum_{a=1}^{N_D} \sum_{b=1}^{N_D} V_{ja} \Theta^{(L)}(\mathbf{x}^a, \mathbf{x}^b) V_{bk}^T = \lambda_j \delta_{jk}. \quad (5.8)$$

と対角化するものである。また、 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ が半正定であることから固有値 λ_j は非負である。

定理 5.2 の条件が満たされた時、 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ は正定となり、 $\lambda_j > 0$ が全ての j に対して成立する。従って、 $t \rightarrow \infty$ の下では、解 (5.7) は全ての a に対して、 $f_{\theta(t)}(\mathbf{x}^a) = y^a$ が成立する。換言すると、コ

コスト関数 \mathcal{L}_t^C の値は global minimum $\mathcal{L}_t = 0$ に到達する。この global minimum への収束性により、over-parameterized cNN が適切に学習できるという事実が説明できる。

さらに一般の \mathbf{x} に対する有用な公式が導かれる。まず、式 (5.5),(5.6),(5.7) を用いると、幅が無限大の極限において、

$$\begin{aligned}\frac{\partial f_{\boldsymbol{\theta}(t)}(\mathbf{x})}{\partial t} &= -\eta \sum_b \boldsymbol{\Theta}^{(L)}(\mathbf{x}, \mathbf{x}^b) (f_{\boldsymbol{\theta}(t)}(\mathbf{x}^b) - y^b) \\ &= -\eta \sum_{b,c,j} \boldsymbol{\Theta}^{(L)}(\mathbf{x}, \mathbf{x}^b) V_{bj}^T (V_{jc} f_{\boldsymbol{\theta}(0)}(\mathbf{x}^c) - V_{jc} y^c) e^{-\eta \lambda_j t}.\end{aligned}$$

となる。これを解くと

$$f_{\boldsymbol{\theta}(t)}(\mathbf{x}) = - \sum_{b,c,j} \boldsymbol{\Theta}^{(L)}(\mathbf{x}, \mathbf{x}^b) V_{bj}^T D_j (V_{jc} f_{\boldsymbol{\theta}(0)}(\mathbf{x}^c) - V_{jc} y^c), \quad (5.9)$$

が得られる。但し、

$$D_j = \begin{cases} (1 - e^{-\eta \lambda_j t}) / \lambda_j & (\lambda_j > 0) \\ \eta t & (\lambda_j = 0) \end{cases}.$$

である。ここで、初期パラメタ $\boldsymbol{\theta}(0)$ を Gauss 分布からランダムにサンプルし、初期パラメタに関して $f_{\boldsymbol{\theta}(t)}(\mathbf{x})$ の平均とると、

$$\langle f_{\boldsymbol{\theta}(t)}(\mathbf{x}) \rangle = \sum_{b,c,j} \boldsymbol{\Theta}^{(L)}(\mathbf{x}, \mathbf{x}^b) V_{bj}^T D_j V_{jc} y^c.$$

が得られる。

公式 (5.9) は未知のデータに対する出力の予測に用いることができる。しかし、NTK を対角化して V を求めるには $O(N_D^3)$ の計算量が必要である。一方、cNN の場合、学習にかかる計算量は N_P をパラメタ数として $O(N_D N_P)$ 程度である。従って、 N_D が非常に大きく $O(N_D^3)$ もの古典計算量が実現困難な場合には、式 (5.9) を直接解くのではなく、 $N_P \leq O(N_D)$ となるような有限幅の cNN を用いればよい。そのような場合、NTK 理論は cNN の挙動を解決するための理論的なツールとして用いることができる。

最後に、コスト関数がバイナリクロスエントロピー (BCE) で与えられる場合を考える。式 (5.2) からコストの関数微分は

$$\begin{aligned}\frac{\partial \mathcal{L}_t^C}{\partial f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a)} &= -y^a \frac{\dot{\sigma}_s(f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a))}{\sigma_s(f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a))} - (1 - y^a) \frac{-\dot{\sigma}_s(f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a))}{1 - \sigma_s(f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a))} \\ &= -y^a + \sigma(f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a)),\end{aligned} \quad (5.10)$$

で与えられる。尚、最後の行ではシグモイド関数の微分公式

$$\dot{\sigma}_s(q) = (1 - \sigma_s(q)) \sigma_s(q).$$

を用いた。式 (5.10) に式 (5.5) を代入することで

$$f_{\boldsymbol{\theta}(t)}(\mathbf{x}^a) = -\eta \int_0^t dt' \sum_b \boldsymbol{\Theta}^{(L)}(\mathbf{x}^a, \mathbf{x}^b) (-y^b + \sigma(f_{\boldsymbol{\theta}(t')}(\mathbf{x}^a))),$$

が得られ、同様に一般の入力 \mathbf{x} に対して、

$$f_{\boldsymbol{\theta}(t)}(\mathbf{x}) = -\eta \int_0^t dt' \sum_b \boldsymbol{\Theta}^{(L)}(\mathbf{x}, \mathbf{x}^b) (-y^b + \sigma(f_{\boldsymbol{\theta}(t')}(\mathbf{x}^a))).$$

が得られる。これらは線型微分方程式ではないため、MSE の場合とは異なり解析的には解けない。しかし、[147] のような一般的な教科書にある手法を用いて数値的に解くことができる。

5.4.4 cNN が適切に学習できないケース（本研究の動機）

NTK 理論によれば、定理 5.2 が成立する限りにおいては、コスト関数は $t \rightarrow \infty$ の極限で global minimum に収束する。しかし、現実的には有限の時間 $t = \tau$ で学習過程を終了する必要がある。従って、cNN の挙動を解析する上では、収束速度が重要な点になる。本節では、収束速度の観点から cNN が適切に学習できないケースについて議論する。

収束速度は、固有値 $\{\lambda_j\}_{j=1}^{N_D}$ に依存することに注意する。もし最小固有値 λ_{\min} が十分に 0 より大きければ、 $O(1/\lambda_{\min})$ 程度の iteration で速やかに global minimum へ収束する。しかし、そうでない場合は、収束速度は固有値スペクトルだけでは決まらず式 (5.7) も考慮する必要がある。実際、妥当な仮定のもと行われた数値実験がそうであることが [137] で示されている。

まず、式 (5.7) を各固有値 λ_j 毎の寄与に切り出し、以下のように書き直す。

$$w_j(t) = (w_j(0) - g_j) e^{-\eta\lambda_j t} + g_j,$$

尚、 $w_j(t) = \sum_a V_{ja} f_{\theta(t)}(x^a)$, $g_j = \sum_a V_{ja} y^a$ とおいた。ここで学習を $t = \tau < O(1/\lambda_{\min})$ で終了することを考える。 $S_{\eta\tau} = \{j | \lambda_j < 1/\eta\tau\}$ の下で、指数関数部分を

$$e^{-\eta\lambda_j \tau} \simeq \begin{cases} 1 & \text{if } j \in S_{\eta\tau} \\ 0 & \text{otherwise} \end{cases},$$

と近似すると、

$$w_j(\tau) \simeq \begin{cases} w_j(0) & \text{if } j \in S_{\eta\tau} \\ g_j & \text{otherwise} \end{cases}. \quad (5.11)$$

が得られる。

同様の近似を iteration step τ におけるコスト関数 \mathcal{L}_τ^C に適用すると、以下のように計算される。

$$\begin{aligned} \mathcal{L}_\tau^C &\equiv \frac{1}{N_D} \sum_{a=1}^{N_D} (f_{\theta(\tau)}(\mathbf{x}^a) - y^a)^2 = \frac{1}{N_D} \sum_{a=1}^{N_D} \left[\sum_{j=1}^{N_D} V_{aj}^T (w_j(\tau) - g_j) \right]^2 \\ &\simeq \frac{1}{N_D} \sum_{a=1}^{N_D} \left(\sum_{j \in S_{\eta\tau}} V_{aj}^T (w_j(0) - g_j) \right)^2 \\ &= \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} w_j(0)^2 + \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} g_j^2 - \frac{2}{N_D} \sum_{j \in S_{\eta\tau}} w_j(0) g_j. \end{aligned}$$

ここで $w_j(0)$ はガウス分布に従う変数の和のため、 $w_j(0)$ もまた以下の共相関行列に従う。

$$\begin{aligned} \langle w_j(0) w_k(0) \rangle &= \sum_{a,b} V_{ja} V_{kb} \langle f_{\theta(0)}(\mathbf{x}^a) f_{\theta(0)}(\mathbf{x}^b) \rangle \\ &= \sum_{a,b} V_{ja} \Sigma^{(L)}(\mathbf{x}^a, \mathbf{x}^b) V_{bk}^T. \end{aligned}$$

従って、

$$\langle \mathcal{L}_\tau^C \rangle \simeq \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} \sum_{b,c} V_{jb} \Sigma^{(L)}(\mathbf{x}^b, \mathbf{x}^c) V_{cj}^T + \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} g_j^2. \quad (5.12)$$

が得られる。また、共相関行列は直交行列 V' によって

$$V'_{jb} \Sigma^{(L)}(\mathbf{x}^b, \mathbf{x}^c) V_{ck}^T = \lambda'_j \delta_{jk},$$

と対角化されるので、式 (5.12) の第一項は

$$\frac{1}{N_D} \sum_{j \in S_{\eta\tau}} \sum_{b,c} V_{jb} \boldsymbol{\Sigma}^{(L)}(\mathbf{x}^b, \mathbf{x}^c) V_{cj}^T = \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} \sum_{k=1}^{N_D} \lambda'_k (\mathbf{v}'_k \cdot \mathbf{v}_j)^2,$$

と書き換えられる。但し、 $\mathbf{v}_j = \{V_{ja}\}_{a=1}^{N_D}$ 、 $\mathbf{v}'_j = \{V'_{ja}\}_{a=1}^{N_D}$ である。

また、式 (5.12) の第二項は

$$\frac{1}{N_D} \sum_{j \in S_{\eta\tau}} g_j^2 = \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} (\mathbf{y} \cdot \mathbf{v}_j)^2,$$

と書き換えられる。但し、 \mathbf{y} は $\mathbf{y} = \{y^a\}_{a=1}^{N_D}$ で定義されるラベルベクトルである。最終的に

$$\langle \mathcal{L}_\tau^C \rangle \simeq \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} \sum_{k=1}^{N_D} \lambda'_k (\mathbf{v}'_k \cdot \mathbf{v}_j)^2 + \frac{1}{N_D} \sum_{j \in S_{\eta\tau}} (\mathbf{y} \cdot \mathbf{v}_j)^2.$$

が得られる。

コスト \mathcal{L}_τ^C が大きくなる場合、第一項と第二項の大きさに依って、以下の2つで特徴付けられる。(i) 大きな固有値に対応する固有ベクトル $\boldsymbol{\Sigma}^{(L)}(\mathbf{x}^b, \mathbf{x}^c)$ と小さな固有値に対応する固有ベクトル $\boldsymbol{\Theta}^{(L)}(\mathbf{x}^b, \mathbf{x}^c)$ が揃っている場合、第一項目が大きくなる。(ii) ラベルベクトルが小さい固有値に対応する固有ベクトル $\boldsymbol{\Theta}^{(L)}(\mathbf{x}^b, \mathbf{x}^c)$ と揃っている場合、第二項目が大きくなる。特に重要なのが、2つの目の条件(ii)が成立する場合である。要するに、ラベルベクトルが小さい固有値に対応する固有ベクトル $\boldsymbol{\Theta}^{(L)}(\mathbf{x}^b, \mathbf{x}^c)$ に揃っている場合には、実用的な時間内に cNN を最適化できないのである。もしもそのようなデータセットが与えられた場合、cNN よりも高い性能を示す別の手法が求められる。これこそが、本研究の提案手法を導入する大きな同期である。

注釈 1: 学習データのラベルにノイズが加わる場合には、コスト関数を厳密に0まで下げる必要はない。例えば、ノイズベクトル ϵ が $\mathbf{y} = \tilde{\mathbf{y}} + \epsilon$ というかたちで正しいラベル $\tilde{\mathbf{y}}$ に印加される場合、過学習を避けるためにも、学習過程は $\sum_{j \in S_{\eta\tau}} (\epsilon \cdot \mathbf{v})^2$ が小さくなる前のタイミング $t = \tau$ で終了することが好ましい。実際、NTK 理論の原著論文 [135] では、過学習を避けるためのアイデアとして早期終了 (early stopping) に言及している。この場合、汎化性能を担保するためにも、 $\sum_{j \in S_{\eta\tau}} (\mathbf{y} \cdot \mathbf{v})^2$ の代わりに $\sum_{j \in S_{\eta\tau}} (\tilde{\mathbf{y}} \cdot \mathbf{v})^2$ の値を下げるよう学習を進めるべきである。

5.5 提案手法

本節では、NTK 理論によって解析可能な qcNN モデルを用いた教師あり学習について説明する。詳細に入る前に、この qcNN の特筆すべき点の概要を記す。この qcNN は図 5.1 に示すように、前半の量子回路部分と後半の cNN が接続された構成をとる。5.4 節で示したのと同じように、cNN の幅無限大の極限では時不変の NTK が得られる。これによりこの系の挙動を理論的に解析することができる。重要な点は、本モデルの NTK には、量子回路のデータエンコード部分を反映したある種の量子カーネルが現れる点である。これは qcNN の出力が、量子特徴量空間（ヒルベルト空間）で定義される量子状態の関数として表現されることを意味している。従って、量子エンコーダーを適切に設計できれば、本モデルは純粋な古典系の性能を凌駕する可能性がある。以降では、5.5.1 項から 5.5.2 項でモデルの詳細を議論し、5.5.4 項で期待される優位性について議論する。

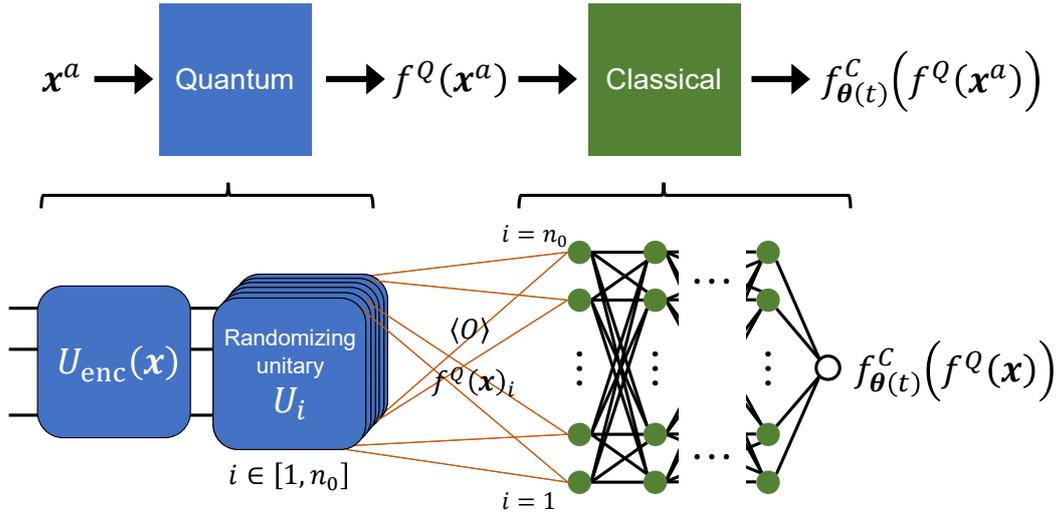


図 5.1: 提案 qcNN モデルの概要. 前半の量子パートはデータ \mathbf{x}^a をエンコードするためのユニタリ演算子 $U_{\text{enc}}(\mathbf{x}^a)$ と、その後続くランダムユニタリ演算子 U_i , そして Observable O による測定で構成され、量子状態 $f^Q(\mathbf{x}^a)_i$ の特徴量を抽出する. n_0 個の異なる量子回路により特徴量ベクトル $\mathbf{f}^Q(\mathbf{x}^a) = (f^Q(\mathbf{x}^a)_1, f^Q(\mathbf{x}^a)_2, \dots, f^Q(\mathbf{x}^a)_{n_0})$ を構成し、後半の古典パートに入力する. 古典パートは n_0 ノードの multi-layered NN で構成される. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

5.5.1 提案手法：量子-古典ハイブリッドニューラルネットワーク

問題設定として、5.4 節と同じものを考える. すなわち、 N_D 個の教師データとして (\mathbf{x}^a, y^a) ($a = 1, 2, \dots, N_D$) を考える. 尚、 \mathbf{x}^a は入力ベクトル、 y^a は出力値である. 仮定として、データセットには以下の（隠れた）関係性があるものとする.

$$y^a = f_{\text{goal}}(\mathbf{x}^a), \quad \forall a.$$

本問題の目的はモデル $f_{\theta(t)}$ が何らかの指標の下で f_{goal} に近づくように学習を行うことである. 尚、 $\theta(t)$ はパラメタ更新の iteration が t の時の学習対象のパラメタである. 本 qcNN $f_{\theta(t)}$ は、量子パート \mathbf{f}^Q と古典パート $f_{\theta(t)}^C$ を以下のように接続されたものとして構成される.

$$f_{\theta(t)} = f_{\theta(t)}^C \circ \mathbf{f}^Q.$$

後述するように、本モデルでは学習パラメタは古典パートのみにあることに注意されたい. 従って、 $\theta(t)$ の添字は古典パートのみにある.

n -qubit で構成される量子パートでは、まずユニタリ演算子 U_{enc} により古典入力ベクトル \mathbf{x}^a を量子状態へエンコードする. すなわち $|\psi(\mathbf{x}^a)\rangle = U_{\text{enc}}(\mathbf{x}^a)|0\rangle^{\otimes n}$ なる量子状態が生成される. その後、ランダムユニタリ演算子 U_i を作用させ、オブザバブル O で測定することで、量子パートの出力の期待値は

$$f^Q(\mathbf{x}^a)_i = \langle \psi(\mathbf{x}^a) | U_i^\dagger O U_i | \psi(\mathbf{x}^a) \rangle = \langle 0 |^{\otimes n} U_{\text{enc}}(\mathbf{x}^a)^\dagger U_i^\dagger O U_i U_{\text{enc}}(\mathbf{x}^a) | 0 \rangle^{\otimes n}.$$

と表される. ここでこの手順を $i = 1, \dots, n_0$ まで繰り返し、 n_0 次元のベクトル $\mathbf{f}^Q(\mathbf{x}^a) = (f^Q(\mathbf{x}^a)_1, f^Q(\mathbf{x}^a)_2, \dots, f^Q(\mathbf{x}^a)_{n_0})$ を構成する. これが最終的な量子パートの出力となる. この

乱択プロセスにより, Classical shadow tomography [38, 148] を用いた機械学習と同じように, $|\psi(\mathbf{x}^a)\rangle$ から特徴量を抽出することができる. 但し, 本手法では snapshot と呼ばれる tomography density matrix を構成する必要はなく, 直接特徴量ベクトルを構成し, 後続の古典パートで処理を行う. この際, 後述するように, n_0 の値は NTK が時不変になる程度に大きい有限の値を用いており, ダイナミクスを解析的に扱うことができる. 一見すると n_0 次元の巨大なベクトル $\mathbf{f}^Q(\mathbf{x}^a)$ を構成する過程は非効率に見えるかもしれないが, 5.6 節の数値実験の結果からも分かるように実用的な大きさの n_0 でも十分に機能する.

本節では, 各要素として以下の設定をとる. 古典入力データ \mathbf{x}^a は, エンコーダ U_{enc} を介して n -qubit の量子状態として埋め込まれる. 理想的には, [127, 149] で指摘されているように, 学習データの持つ対称性などの隠れた構造を反映しつつエンコード回路 U_{enc} を設計すべきである. ランダムユニタリ演算子 U_i に関しては, tensor product で表されるものを考える. 具体的には

$$U_i = U_i^1 \otimes U_i^2 \otimes \cdots \otimes U_i^{n_Q},$$

なるものを用いる. 但し, m は “locality” と呼ばれる整数であり, $n_Q = n/m$ も整数とする. U_i^k ($k = 1, 2, \dots, n_Q$) は unitary 2-design からサンプルされた, 互いに独立なユニタリ演算子とし, 学習過程では同じものを用いる. 尚, unitary 2-design を量子回路で実装するには $O(m^2)$ の gate が必要であることが示されている [150]. 最後に, オブザバブル O は n_Q 個の local operator の和で構成されるとする. すなわち,

$$O = \sum_{k=1}^{n_Q} I_{(k-1)m} \otimes O \otimes I_{(n_Q-k)m},$$

であり, I_u は 2^u 次元の恒等演算子, O は 2^m -次元の traceless な演算子とする.

次に古典パート $f_{\theta(t)}^C$ を記す. これはベクトル $\mathbf{f}^Q(\mathbf{x}^a)$ を入力とし, 出力として $f_{\theta(t)}^C(\mathbf{f}^Q)$ を返す cNN であり, $f_{\theta(t)}(\mathbf{x}^a) = f_{\theta(t)}^C(\mathbf{f}^Q(\mathbf{x}^a))$ と表せる. いま, $f_{\theta(t)}^C$ を L 層の全結合 cNN として実装すると, 5.4 節で導入したのと同様に

$$\begin{aligned} \alpha^{(0)}(\mathbf{x}^a) &= \mathbf{f}^Q(\mathbf{x}^a), \\ \alpha^{(\ell)}(\mathbf{x}^a) &= \sigma(\tilde{\alpha}^{(\ell)}(\mathbf{x}^a)), \\ \tilde{\alpha}^{(\ell+1)}(\mathbf{x}^a) &= \frac{1}{\sqrt{n_\ell}} W^{(\ell)} \alpha^{(\ell)}(\mathbf{x}^a) + \xi b^{(\ell)}, \\ f_{\theta(t)}^C(\mathbf{f}^Q(\mathbf{x}^a)) &= \tilde{\alpha}^{(L)}(\mathbf{x}^a), \end{aligned} \tag{5.13}$$

が成り立つ. 尚, $\ell = 0, 1, \dots, L-1$ である. また, 5.4 節の場合と同様に, $W^{(\ell)}$ は $n_{\ell+1} \times n_\ell$ の重み行列であり, $b^{(\ell)}$ は n_ℓ 次元のバイアスベクトルである. W と $b^{(\ell)}$ の各要素は互いに独立なガウス分布からサンプルされた初期化パラメタである.

$\theta(t)$ は最急降下法

$$\frac{\partial \theta_p(t)}{\partial t} = -\eta \frac{\partial \mathcal{L}_t^Q}{\partial \theta_p(t)}, \tag{5.14}$$

で更新される. ここで \mathcal{L}_t^Q は $f_{\theta(t)}$ と f_{goal} の距離を反映するコスト関数である. また, η は学習率, $\theta_p(t)$ ($p = 1, 2, \dots, P$) は $\theta(t)$ の p 番目の要素であり, $W^{(1)}, W^{(2)}, \dots, W^{(L-1)}$ と $b^{(1)}, b^{(2)}, \dots, b^{(L-1)}$ に対応する. パラメタを更新するタスクは古典パートにしか現れないため, N_D 個の学習データ $\{(\mathbf{x}^a, y^a)\}$ ($a = 1, 2, \dots, N_D$), cNN $f_{\theta(t)}^C$ と量子パートからの出力を用いて, 古典機械学習で開発された洗練されたソルバーを適用することができる.

5.5.2 Quantum neural tangent kernel

5.4 節で示されたように、最急降下法 (5.14) を用いたパラメタ更新の際には、出力関数 $f_{\theta(t)}$ の時間変化は以下に従う。

$$\frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial t} = -\eta \sum_{a=1}^{N_D} K_Q(\mathbf{x}, \mathbf{x}^a, t) \frac{\partial \mathcal{L}_t^Q}{\partial f_{\theta(t)}(\mathbf{x}^a)}. \quad (5.15)$$

但し、 $K_Q(\mathbf{x}, \mathbf{x}', t)$ は “Quantum neural tangent kernel (QNTK)” であり、

$$K_Q(\mathbf{x}, \mathbf{x}', t) = \sum_{p=1}^P \frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial \theta_p(t)} \frac{\partial f_{\theta(t)}(\mathbf{x}')}{\partial \theta_p(t)}.$$

で定義される。これが半正定値であることは素直に示すことができる。次節では、なぜこれが “Quantum neural tangent kernel” と呼べるかを説明する。

5.5.3 QNTK に関する定理

まず初めに $L = 1$ の場合の出力 $f_{\theta(0)}$ の確率分布に関する定理を記す。

定理 5.3 σ を Lipschitz な非線形関数としたとき、 $L = 1$ の下で $n_0 \rightarrow \infty$ の極限では、出力 $f_{\theta(0)}$ は、共分散行列 $\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}')$ が以下の式で与えられるような Centered Gaussian process に従う。

$$\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}') = \frac{\text{Tr}(\mathcal{O}^2)}{2^{2m} - 1} \sum_{k=1}^{n_Q} \left(\text{Tr}(\rho_{\mathbf{x}}^k \rho_{\mathbf{x}'}^k) - \frac{1}{2^m} \right) + \xi^2.$$

但し、 $\rho_{\mathbf{x}}^k$ は以下のように定義される reduced density matrix である。

$$\rho_{\mathbf{x}}^k = \text{Tr}_k (U_{\text{enc}}(\mathbf{x})|0\rangle^{\otimes n} \langle 0|^{\otimes n} U_{\text{enc}}(\mathbf{x})^\dagger), \quad (5.16)$$

尚、 Tr_k は $(km - m)$ 番目の qubit から $(km - 1)$ 番目の qubit を除いたヒルベルト空間全体に対する部分トレースである。

本定理の証明は [45] の Appendix A を参照されたい。ここで $\sum_{k=1}^{n_Q} \text{Tr}(\rho_{\mathbf{x}}^k \rho_{\mathbf{x}'}^k)$ の部分が、[126] で導入された projected quantum kernel (PQK) に相当することに注意されたい。これを用いたのは以下の理由による。それは、qubit 数が大きな場合、あるクラスの特徴量マップの下 [126, 134, 151] では、すなわちヒルベルト空間の次元が大きな場合、純粋状態 $\text{Tr}(\rho_{\mathbf{x}} \rho_{\mathbf{x}'}^k) = |\langle \psi(\mathbf{x}) | \psi(\mathbf{x}') \rangle|^2$ の内積によって構成されるグラム行列が恒等行列に近付いていくという現象が起こる。換言すると、そのようなカーネルを用いると量子優位性が得られないことを意味している。PQK はそのような問題を回避しようとされている。PQK は、高次元ヒルベルト空間内の密度行列を低次元空間に射影 (5.16) することにより、射影された密度行列間内積によって定義されるカーネルのグラム行列は、恒等行列とは大きく異なる何らかの量子系特有の構造をもちうる。

共相関行列 $\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}')$ は PQK の性質を反映するため、以下の系が導かれる。

系 5.1 定理 5.3 から得られる共相関行列 $\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}')$ は

$$\xi = \sqrt{\frac{n_Q \text{Tr}(\mathcal{O}^2)}{(2^{2m} - 1) 2^m}}.$$

とおけば

$$\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}') = \frac{\text{Tr}(\mathcal{O}^2)}{2^{2m} - 1} \sum_{k=1}^{n_Q} \text{Tr}(\rho_{\mathbf{x}}^k \rho_{\mathbf{x}'}^k),$$

と表せる.

すなわち, 式 (5.13) で与えられるバイアスペクトルの係数を適切に選択できれば, $\Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}')$ は定数項を除いて正確に PQK になる.

$L = 1$ の場合の結果に基づくと, 次の定理が導かれる. まず, $L = 1$ の時の $f_{\theta(0)}$ の分布は, 再帰的に以下のように計算できる.

定理 5.4 σ を Lipschitz な非線形関数としたとき, $L = 1$ の下で $n_0, n_1, \dots, n_{L-1} \rightarrow \infty$ の極限では, 出力 $f_{\theta(0)}$ は, 共分散行列 $\Sigma_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ が再帰的に以下の式で与えられるような Centered Gaussian process に従う.

$$\begin{aligned} \Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}') &= \frac{\text{Tr}(\mathcal{O}^2)}{2^{2m} - 1} \sum_{k=1}^{n_Q} \left(\text{Tr}(\rho_{\mathbf{x}}^k \rho_{\mathbf{x}'}^k) - \frac{1}{2^m} \right) + \xi^2, \\ \Sigma_Q^{(\ell+1)}(\mathbf{x}, \mathbf{x}') &= \mathbf{E}_{h \sim \mathcal{N}(0, \Sigma_Q^{(\ell)})} [\sigma(h(\mathbf{x}))\sigma(h(\mathbf{x}'))] + \xi^2, \end{aligned} \quad (5.17)$$

但し, 期待値は共分散行列 $\Sigma_Q^{(\ell)}$ の centered Gaussian process について計算するものとする.

本定理の証明は [45] の Appendix B を参照されたい. 尚, 量子の場合 (5.17) と古典の場合 (5.4) の違いは, 第一層目に対応する共相関行列のみである.

幅無限大の極限における QNTK は, 定理 5.1 と同様に考えると以下のように表せる.

定理 5.5 σ を Lipschitz な非線形関数としたとき, $n_0, n_1, \dots, n_{L-1} \rightarrow \infty$ の極限では, QNTK $K_Q(\mathbf{x}, \mathbf{x}', t)$ は, 以下の再帰的な関係式に従って時不変な関数 $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ に収束する.

$$\begin{aligned} \Theta_Q^{(1)}(\mathbf{x}, \mathbf{x}') &= \Sigma_Q^{(1)}(\mathbf{x}, \mathbf{x}') = \frac{\text{Tr}(\mathcal{O}^2)}{2^{2m} - 1} \sum_{k=1}^{n_Q} \left(\text{Tr}(\rho_{\mathbf{x}}^k \rho_{\mathbf{x}'}^k) - \frac{1}{2^m} \right) + \xi^2, \\ \Theta_Q^{(\ell+1)}(\mathbf{x}, \mathbf{x}') &= \Theta_Q^{(\ell)}(\mathbf{x}, \mathbf{x}') \dot{\Sigma}_Q^{(\ell)}(\mathbf{x}, \mathbf{x}') + \Sigma_Q^{(\ell+1)}(\mathbf{x}, \mathbf{x}'), \end{aligned}$$

但し, $\dot{\Sigma}_Q^{(\ell)}(\mathbf{x}, \mathbf{x}') = \mathbf{E}_{h \sim \mathcal{N}(0, \Sigma_Q^{(\ell)})} [\dot{\sigma}(h(\mathbf{x}))\dot{\sigma}(h(\mathbf{x}'))]$, $\dot{\sigma}$ は σ の微分である.

本定理の証明は [45] の Appendix C を参照されたい. 尚, 上記の 2 つの定理は [135] とほぼ同様の証明過程を用いている.

$L = 1$ の場合, QNTK は量子カーネルの構造をそのまま引き継ぐ. これが $K_Q(\mathbf{x}, \mathbf{x}', t)$ を “Quantum” NTK と呼んでいる理由である. さらに, このような第一層目の構造は $L > 1$ の場合にも引き継がれていき, その結果得られるカーネルは PQK の非線形関数となる. 有効な量子カーネルの設計方法が大いに非自明であることを鑑みるに, 本手法で $L > 1$ の場合に現れるように, 非線形カーネルが自動的に生成できることは大変有用である. また, 活性化関数として ReLU を用いた場合の $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ の解析的な再帰的計算方法は [45] の Appendix D を参照されたい.

古典の場合, 定理 5.5 は qcNN の学習過程の挙動を解析的に調べるための重要な特性である. 特に定理 5.2 と式 (5.8) の下の議論, すなわちカーネル $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ の半正定値性が重要性である. 実際, 定理 5.2 と同じような結果を得ることができる.

定理 5.6 σ を Lipschitz な非線形関数としたとき, QNTK $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ は以下のような $\{c_a\}_{a=1}^{N_D}$ が存在しない限りは正定である. (i) $\sum_a c_a \rho_{\mathbf{x}^a}^k = \mathbf{0}$ ($\forall k$) かつ $\sum_a c_a = 0$ かつ $c_a \neq 0$ ($\exists a$). (ii) $\xi = 0$ かつ $\sum_a c_a \rho_{\mathbf{x}^a}^k = I_m / 2^m$ ($\forall k$) かつ $\sum_a c_a = 1$.

本定理の証明は [45] の Appendix E を参照されたい. (i) の条件は, データを埋め込んだ reduced density matrices が線型従属である場合と解釈できる. これは冗長なデータを除くことで回避できる. 一方, (ii) の条件には直感的な解釈を与えることは難しいが, ξ を 0 より大きな値に選ぶことで容易に回避できる.

上記の定理に基づくと, 学習過程や最終的な性能を理論的に解析できる. cNN 部分の幅無限大の極限では, 式 (5.15) で与えられる出力 $f_{\theta(t)}(\mathbf{x})$ のダイナミクスは

$$\frac{\partial f_{\theta(t)}(\mathbf{x})}{\partial t} = -\eta \sum_b \Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}^b) \frac{\partial \mathcal{L}_t^Q}{\partial f_{\theta(t)}(\mathbf{x}^b)}. \quad (5.18)$$

で与えられる. この方程式の量子の場合 (5.18) と古典の場合 (5.5) の違いは NTK のかたちのみであるため, 5.4 節の議論をそのまま適用できる. 特に, コスト関数 \mathcal{L}_t^Q が平均二乗誤差 (MSE) (5.1) の場合, 式 (5.18) の解は

$$f_{\theta(t)}(\mathbf{x}^a) = \sum_j V_{aj}^{QT} \left(V_{jb}^Q f_{\theta(0)}(\mathbf{x}^b) - V_{jb}^Q y^b \right) e^{-\eta \lambda_j t} + y^a, \quad (5.19)$$

で与えられる. 但し, 直交行列 V^Q は, $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ を

$$\sum_{a=1}^{N_D} \sum_{b=1}^{N_D} V_{ja}^Q \Theta_Q^{(L)}(\mathbf{x}^a, \mathbf{x}^b) V_{bk}^{QT} = \lambda_j^Q \delta_{jk}.$$

と対角化するものである. また, $\{\lambda_j^Q\}$ は $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ の固有値であり, 一般に半正定である. もし定理 5.6 が成立すれば $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ は正定であり, $\{\lambda_j^Q\}$ が全て正の値になることと等価である. その場合, 式 (5.19) から $t \rightarrow \infty$ で $f_{\theta(t)}(\mathbf{x}^a) \rightarrow y^a$ となり, 学習過程が完結する. また, コスト関数が binary cross entropy (5.2) の場合は

$$f_{\theta(t)}(\mathbf{x}^a) = -\eta \int_0^t dt' \sum_b \Theta_Q^{(L)}(\mathbf{x}^a, \mathbf{x}^b) (-y^b + \sigma(f_{\theta(t')}(\mathbf{x}^a))). \quad (5.20)$$

となる.

5.5.4 提案モデルの優位性

本節では, 提案 qcNN が他のモデルに対して有する優位性について 2 つのシナリオについて記す.

古典モデルに対する優位性

まず初めに古典モデルに対する qcNN の優位性について論じる. そのために, 本研究における QNTK が PQK の非線形関数のかたちで量子状態に関する特徴量を含んでいることを思い出す (定理 5.5). 従って, PQK が古典計算では計算が困難なこと [126] から, この QNTK も古典計算では表現が困難であると言える. その結果, 出力 (5.19), (5.20) は, 古典手法では到達不可能な程度に学習誤差や汎化誤差を抑えられる可能性がある. ここで有効な量子カーネルの設計が一般には非自明である

ことを考慮すると、与えられたタスクに対して古典手法を凌駕する非線形カーネル関数をニューラルネットワークベースで生成できると有用である。

上述の点を掘り下げるために、学習誤差の観点での優位性を考えてみる。具体的には、

$$\min_{\sigma \in F, L} \langle \mathcal{L}_\tau^C \rangle > \min_{\sigma \in F, L, U_{\text{enc}}} \langle \mathcal{L}_\tau^Q \rangle, \quad (5.21)$$

が成立する状況を考える。ここでコストの変化がしなくなる程度に長い時間 τ を仮定する。また、 F を微分可能な Lipschitz 関数として、初期パラメタ全体についての平均を考える。もし式 (5.21) が成立するならば、この qcNN モデルは学習誤差の観点で古典モデルよりも優れていると言える。条件 (5.21) を解析的に解釈するために、コストが MSE であることも合わせて仮定する。すると条件 (5.21) は式 (5.11) を用いて以下のように近似できる。

$$\min_{\sigma \in F, L} \left\{ \sum_{j \in S_{\eta\tau}^C} \sum_{k=1}^{N_D} \lambda_k^{C'} (\mathbf{v}_k^{C'} \cdot \mathbf{v}_j^C)^2 + \sum_{j \in S_{\eta\tau}^C} (\mathbf{y} \cdot \mathbf{v}_j^C)^2 \right\} > \min_{\sigma \in F, L, U_{\text{enc}}} \left\{ \sum_{j \in S_{\eta\tau}^Q} \sum_{k=1}^{N_D} \lambda_k^{Q'} (\mathbf{v}_k^{Q'} \cdot \mathbf{v}_j^Q)^2 + \sum_{j \in S_{\eta\tau}^Q} (\mathbf{y} \cdot \mathbf{v}_j^Q)^2 \right\}, \quad (5.22)$$

但し、 $(\{\lambda_k^C\}_{k=1}^{N_D}, \{\mathbf{v}_k^C\}_{k=1}^{N_D})$, $(\{\lambda_k^Q\}_{k=1}^{N_D}, \{\mathbf{v}_k^Q\}_{k=1}^{N_D})$, $(\{\lambda_k^{C'}\}_{k=1}^{N_D}, \{\mathbf{v}_k^{C'}\}_{k=1}^{N_D})$, $(\{\lambda_k^{Q'}\}_{k=1}^{N_D}, \{\mathbf{v}_k^{Q'}\}_{k=1}^{N_D})$ はそれぞれ $\Sigma^{(L)}(\mathbf{x}, \mathbf{x}')$, $\Sigma_Q^{(L)}(\mathbf{x}, \mathbf{x}')$, $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$, $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ の固有値と固有ベクトルのペアである。また、 $S_{\eta\tau}^C$ と $S_{\eta\tau}^Q$ は、それぞれ $\lambda_j^C < 1/\eta\tau$ と $\lambda_j^Q < 1/\eta\tau$ に対応するインデックスである。以下では、 $S_{\eta\tau}^C$ や $S_{\eta\tau}^Q$ のインデックスに対応する固有値ベクトルを “bottom eigenvectors” と呼ぶことにする。いま、条件 (5.21) は条件 (5.22) に変換され、これらは共相関行列と NTK の固有ベクトルで表されている。特に重要なのは、両辺の第二項目である。この項は、 $\Theta^{(L)}(\mathbf{x}, \mathbf{x}')$ や $\Theta_Q^{(L)}(\mathbf{x}, \mathbf{x}')$ の bottom eigenvectors がラベルベクトル \mathbf{y} に、どれほど沿っているかにより依存している。従って、もしも古典計算不能な QNTK の bottom eigenvectors が \mathbf{y} に全く沿っていなければ、古典の対応物が \mathbf{y} に沿っていたとしても、式 (5.22) は満たされそうである。すなわち、本 qcNN が古典手法に対して優位性を持ち得るということである。この議論はまた、量子優位性を考える上でデータセットの構造が重要であることを示唆している ([126] の Supplemental Materials Section 7 も参照のこと)。本問題設定の場合、 \mathbf{y} を $\sum_{j \in S_{\eta\tau}^C} (\mathbf{y} \cdot \mathbf{v}_j^C)^2 \gg \sum_{j \in S_{\eta\tau}^Q} (\mathbf{y} \cdot \mathbf{v}_j^Q)^2$ となるように操作すれば、qcNN モデルが優位性を有するデータセットを得られる。量子優位性を示す実用的なデータセットを明らかにし、それに対応するエンコーダを得るための包括的な研究は、実に重要であるが、今後の研究テーマとしておく。

量子カーネル法についての補足

提案 qcNN モデルは、学習過程における計算量的複雑性の観点から、量子カーネル法に比べて優位性がある。[132] にあるように、量子カーネルは representer theorem [152] を用いることで、一般に標準的な変分法によって解を探す場合に比べて、学習誤差の観点から優れていることが多い。尚、ここで “標準的な” と指したのは、エンコーダーを一度しか用いないタイプの手法であり、エンコーダーを繰り返し使用する、いわゆる “re-uploading” タイプの手法は含まない。しかしながら量子カーネル法は、古典カーネル法と同様に、スケーラビリティが乏しい。すなわち、量子カーネルを計算するには、データ数 N_D に対して、 $O(N_D^2)$ の計算量が必要である。それに対して、提案 qcNN は古典パートが幅無限大の極限のカーネル法に一致し、iteration 数を T とすれば、近似器の学習に必要な計算量は $O(N_D T)$ ですむ。従って、iteration 数が $T \ll N_D$ の下では、提案 qcNN モデルはスケーラブルな量子カーネル法であると言える。

提案モデルが純量子, 純古典モデルを凌駕する設定

次に, 提案モデルが以下に記す量子状態の特徴量予測問題において, 学習誤差の観点で他のモデルよりも優れた性能を示す可能性について論じる. まず, 学習データとして $\{\rho(\mathbf{x}^a), y_a\}$ が手元にあるとする. 但し, $\rho(\mathbf{x}^a)$ は \mathbf{x}^a なる入力ラベル (例えば温度のようなパラメタ) で特徴づけられる不明な量子状態, y_a はオブザバブル (例えば全磁化) の平均出力値とする. 問題は, この学習データを元に, 新たなラベル \mathbf{x} , もしくは $\rho(\mathbf{x})$ に対応する y の予測器を構成することである. いま, 提案モデルは直接 $\rho(\mathbf{x}^a)$ にアクセスできるとする. この場合, $\{\mathbf{x}^a, y_a\}$ しかアクセスできない古典モデルに比べると, 明らかに良い予測器を得ることができる. また, 定理 5.5 の下に記載の通り, 提案モデルは量子カーネルの非線形関数を表せることから, $\rho(\mathbf{x}^a)$ に直接アクセスできるとは言え, $\rho(\mathbf{x}^a)$ の線型関数しか表現できない純量子のモデルよりも良く予測できるだろう. これらの優位性は 5.6.3 項の数値実験でも示される. 更に, [148] では $\rho(\mathbf{x}^a)$ を近似するための classical shadow を生成するために乱択測定を行うモデルを提案しており, 新たな入力 $\rho(\mathbf{x})$ に対する予測値 y を予測するための shadow の関数を構成している. これに対し, 本手法では classical shadow を用いずに乱択測定の結果から直接近似器を構成するため, [148] で提案されたクラスを含む. 従って, 前者は後者を凌駕する可能性がある. 重要な点は, [148] で示されている彼らのモデルで効率的に解くことができる問題のクラスは, 原理的に提案モデルでも解くことができる点である. 最後に, [153] では同様の特徴量予測の問題において, 定数個の学習データで量子モデルでは解けるのに対して, 古典モデルでは指数関数個の学習データが必要となるようなクラスを同定していることを指摘しておく. この後の数値計算では, 提案モデルが量子優位性を示すような設定を同定することを目指す.

5.6 数値実験: QNN, CNN との比較

本章の目的は以下の質問に数値的に答えることである.

- QNTK の収束速度がどの程度なのか. また, 理論的に示された幅無限大の qcNN と比べて, 現実的な有限幅の qcNN の学習ダイナミクスがどの程度なのか.
- Locality m が qcNN の学習にどの程度影響するのか. 尚, m は埋め込まれたデータの特徴量抽出の際の乱択ユニタリのサイズである.
- 提案 qcNN モデルは, 純古典モデルや純量子モデルに対して明らかな優位性があるのか.

これらの問題を確かめるため, 以下の 3 つの数値実験を行った. まず, 最初の質問について確認するために, 5.6.1 項では, 幅有限 qcNN と幅無限大の qcNN を用いて回帰/分類タスクを行った. 特に, 様々な量子データエンコーダを用いて結果を比較した. 次に二つ目の質問について確認するために, 5.6.2 項では, ある回帰タスクを行った. 最後に三つ目の質問について確認するために, 5.6.3 項では, 有限幅の qcNN, 純量子 NN(qNN) と純古典 NN(cNN) を用いて, 回帰/分類タスクの性能を比較した. 問題設定としては, ある量子プロセスから生成されるデータセットを仮定した. これらの数値実験では, 量子回路シミュレータとして qulacs [154] を使用した.

5.6.1 幅有限 qcNN と幅無限大 qcNN の比較

本節では, 幅有限 qcNN と幅無限大 qcNN に対して, 様々なタイプの量子エンコーダを用いて, 回帰タスクと分類タスクを行った.

問題設定

量子回路の選択 量子データエンコーダとして、表 5.1, 図 5.2 に示されるような、4 種類の量子回路 $U_{\text{enc}}(\mathbf{x})$ を用いた。5 種類とも全て n -qubit で構成されており、初めに Hadamard gate をそれぞれの qubit へ作用させる。その後、RZ-gate を用いてデータ $x_i \in [1, 2, \dots, n]$ を量子状態へ埋め込む。尚、 $\text{RZ}(x_i) = \exp(-2\pi i x_i)$ である。ここで、データベクトルは $\mathbf{x} = [x_1, x_2, \dots, x_n]$ とし、データベクトルの次元は qubit 数と等しいとする。その後の回路を便宜上、以下のように type-A と type-B に分類して考える。まず、type-A のエンコーダには Ansatz-A, Ansatz-A4, Ansatz-A4ne (Ansatz-A4 は Ansatz-A を 4 回繰り返したもの) を割り当てる。これらはデータ間のクロスターム $x_i x_j$ ($i, j \in [1, 2, \dots, n]$) を RZ-gate によって埋め込む。一方、type-B のエンコーダには Ansatz-B, Ansatz-Bne を割り当てる。これらは type-A のようなクロスタームはなく、上述の RZ-gate でサンプルにデータを埋め込んだものである。従って、type-A の方が type-B に比べて高い非線形を備えている。エンコーダ間のもう一つの違いが、CNOT-gate の存在である。Ansatz-A, Ansatz-A4, Ansatz-B には CNOT-gate があるが、Ansatz-A4ne, Ansatz-Bne にはない。CNOT-gate は qubit 間にエンタングルメントをもたらすため、量子性の有無が現れる (ne は “non-entangled” の意)。一般に、CNOT-gate の多い巨大な量子回路は古典シミュレートが計算量的に困難とされるため、CNOT-gate を含む Ansatz-A, Ansatz-A4, Ansatz-B は、それ以外に比べて特定のタスクでは良い性能を示すことが期待される。量子回路に続く古典 NN の構成は次の項で記す。

表 5.1: $U_{\text{enc}}(\mathbf{x})$ の比較

Circuit type	Cross-term	CNOT	Depth
Ansatz-A	Yes	Yes	$\times 1$
Ansatz-A4	Yes	Yes	$\times 4$
Ansatz-A4ne	Yes	No	$\times 4$
Ansatz-B	No	Yes	$\times 1$
Ansatz-Bne	No	No	$\times 1$

古典 NN の学習方法 提案フレームワークでは、学習パラメタは古典パート (cNN) にしか含まれず、標準的な最適化手法によりパラメタ更新を行う。まず、量子回路の出力 $f^Q(\mathbf{x}^a)_i = \langle \psi((\mathbf{x}^a)) | U_i^\dagger O U_i | \psi((\mathbf{x}^a)) \rangle$, $i \in [1, 2, \dots, n_0]$ を、全学習データ $\{(\mathbf{x}^a, y^a)\}$, $a \in [1, 2, \dots, N_D]$ に対して計算する。出力は、locality $m = 1$ の unitary 2-designs [155] からサンプルされた n_0 個のランダムユニタリー $\{U_1, U_2, \dots, U_{n_0}\}$ を介して生成される。ここで期待値 $U_i^\dagger O U_i$ をサンプリングではなく、state vector simulator を使って計算した。尚、サンプリングに伴う shot noise の影響は 5.6.3 項で確認する。また、これらの値は cNN へ受け渡される (cNN の第一層目の幅が n_0 であることに注意)。cNN の学習は標準的な最急降下法で行われ、学習率等のハイパーパラメタは後述するように適切に設定される。 $t = 0$ に対応する初期パラメタは正規分布 $\mathcal{N}(0, \sqrt{2/N_{\text{param}}})$ からランダムにサンプリングされる。但し、 N_{param} は各層のパラメタ数であり、 $\mathcal{N}(\mu, \sigma)$ は平均 μ , 標準偏差 σ の正規分布を表す。

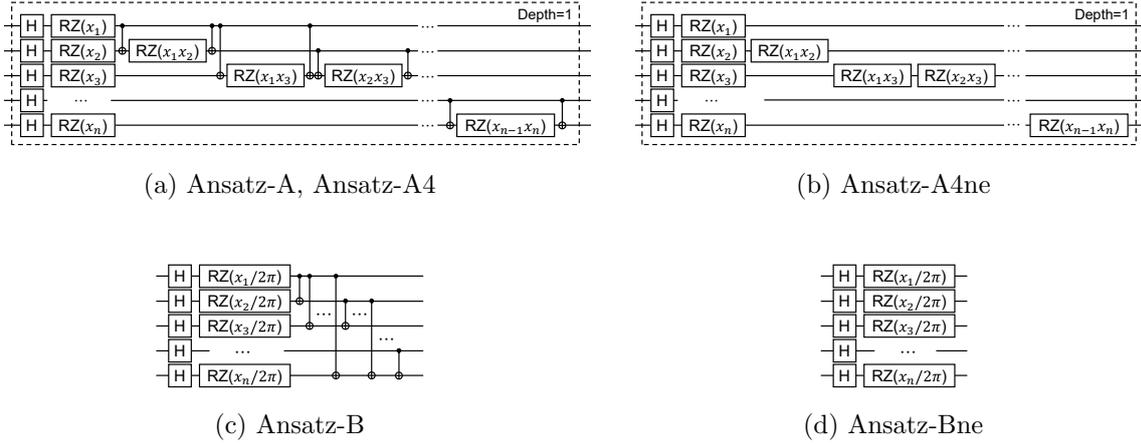


図 5.2: $U_{\text{enc}}(\mathbf{x})$ の構成. 各 qubit に Hadamard gate を作用させた後, RZ-gates の回転角として正規化されたデータ値 x_i ($i = 1, \dots, n$) をエンコードする. さらに, (a) と (c) には entangler として CNOT-gates を作用させる. また, (a) と (b) には RZ-gates の回転角として 2 つのデータ値の積をエンコードする (表 5.1 中の “Cross-term”). 尚, RZ(x) の回転角に入力する際の係数が (a) と (b) では $2\pi x$ である点, Ansatz-A4 と Ansatz-A4ne では破線で囲った回路 (shown as “Depth=1”) を 4 回繰り返す点にも注意されたい. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

結果

回帰タスク 回帰タスクには, 1次元のターゲット関数として $f_{\text{goal}}(x) = \sin(x) + \epsilon$ を用いた. 但し, ϵ は正規分布 $\mathcal{N}(0, 0.05)$ からサンプルされた独立同分布 (independent and identically distributed; i.i.d.) のノイズである. 学習データセット $\{x^a, f_{\text{goal}}(x^a)\}, a = 1, \dots, N_D$ は, $x \in U(-1, 1)$ からサンプルされる. 但し, $U(u_1, u_2)$ は $[u_1, u_2]$ の間の一様分布である. 量子回路への入力には, 1次元の入力 x を 4次元ベクトル $\mathbf{x} = [x_1, x_2, x_3, x_4] = [x, x^2, x^3, x^4]$ として与えた. ここで学習データ点は $N_D = 100$ とし, qubit 数は $n = 4$ とした. コスト関数には最小二乗誤差 (MSE) を用い, optimizer には学習率 10^{-4} の確率的勾配降下法 (stochastic gradient descent; SGD) を用いた. cNN は, ノード数 1000 の隠れ層 1層 ($L = 1, n_0 = 10^3$) を設定した.

学習過程におけるコスト関数の時間発展の比較を図 5.3 に示した. 左がシミュレーション ($n_0 = 10^3$) の結果, 右が cNN の幅無限大の下で計算した理論曲線である. シミュレーション結果は, $\{U_i\}$ と cNN のパラメタをそれぞれランダムに初期化した 100 回の試行のうち最も良い結果を記してある. 注目すべき点は, シミュレーションの結果得られた収束値が, 理論解析の結果の値と非常に良く一致している点である. これは提案手法における qcNN の性能が, 量子回路の形に依らず, 解析的に予測できることを示している. すなわち, 手元にデータセットと qcNN の構成があれば, 実際にシミュレーションを行うことなく, 性能が計算できるのである. これは有効な qcNN の設計を行う上では非常に優れた点である.

次に個々の曲線の振る舞いに着目する. まず全体的に, type-B のエンコーダが type-A よりも優れた性能を表している. これは恐らくターゲット関数に対して, type-A のエンコーダのもつ表現力が高すぎるからであると考えられる. 量子回路の表現力の解析については [156, 157] で系統的に調べられている. すなわち, 量子回路の繰り返し回数が, その回路の表現可能な関数のフーリエ係数の分布を決定付けていることが示されている. つまり, モデル関数が高い周波数成分を含んでいれ

ば、豊かな表現力を持っているといえる。この観点から考えれば、今回の問題設定では、エンコーダを1層のみ持っている type-B が、エンコーダを4層持っている type-A よりも性能が高いのは理にかなっている。また、この傾向は、文中には示していないが、Ansatz-A4 を用いてターゲット関数 $f_{\text{goal}}(x) = (x - 0.2)^2 \sin(12x)$ という複雑な関数に対して最も高い性能を示したことから確認できた。以上の結果より、エンコーダは qcNN の性能に直結することが分かる。特に表現力のチューニングにはよく注意して設計を行う必要がある。

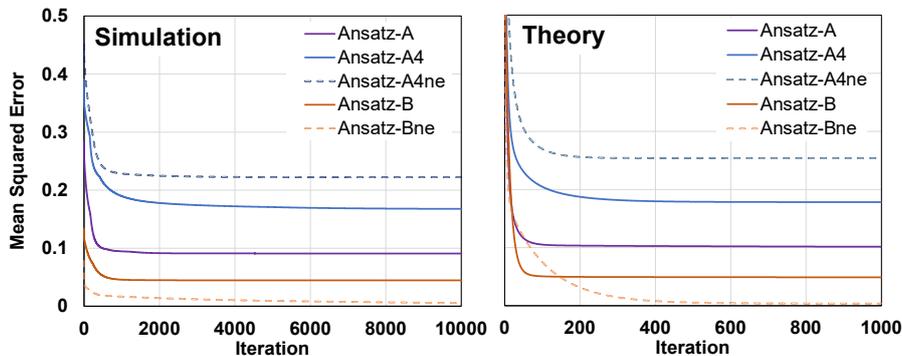


図 5.3: 回帰タスクにおける学習曲線。コスト関数の時間発展は、数値シミュレーションは $n_0 = 10^3$ 、理論曲線は $n_0 \rightarrow \infty$ で得られたもの。Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

分類タスク 分類タスクでは、量子サポートベクトルマシン (quantum support vector machine; QSVM) が量子優位性を示した際 [74] に用いた人工データセット [158] を用いた。各入力ベクトル \mathbf{x} は 2 次元であり、量子回路の qubit 数は $n = 2$ である。cNN への入力数は、cNN の幅と等しく $n_0 = 10^3$ (デフォルト) であるが、Ansatz-A4ne については $n_0 = 10^2$ と $n_0 = 10^4$ も行った。また、cNN の幅については $L = 1$ と $L = 2$ について調べた。cNN の活性化関数には以下を用いた。出力層については、 $L = 1$ と $L = 2$ の両方でシグモイド関数 $\sigma(q) = 1/(1 + e^{-q})$ を用いた。一方で、 $L = 2$ の入力層には ReLU $\sigma(q) = \max(0, q)$ を用いた。ノード数は、 $L = 1$ は $n_0 = 10^3$ とし、 $L = 2$ は $n_0 = n_1 = 10^3$ とした。出力値のラベル数 y は 2 クラスとし、以下のルールに従って対応するラベルを付与した。すなわち、もし $f_{\theta(t)}^C(f^Q(\mathbf{x}^a))$ が 0.5 よりも大きければ”1”、そうでなければ”0”とした。学習データ数は、各クラス $N_D = 50$ 。コスト関数は Binary cross entropy (BCE) (5.2) とし、optimizer には学習率 10^{-3} の Adam [88] を用いた。

学習過程におけるコスト関数の時間発展の比較を図 5.4 に示した。左がシミュレーションの結果、右が cNN の幅無限大で計算した理論曲線である。シミュレーション結果は、 $\{U_i\}$ と cNN のパラメータをそれぞれランダムに初期化した 100 回の試行のうち最も良い結果を記してある。特に $L = 1$ について、同じ ansatz 同士で比べると、数値実験と理論計算の軌跡がとてもよく一致している様子が見て取れる。一方で注目すべきは Ansatz-A4 と Ansatz-A4ne の間の大きな違いである。すなわち、理論曲線の方では前者は後者と比べて低い最終値へ到達しているのに比べ、数値実験では逆転している。ここで Ansatz-A4 は CNOT を含んでいるために、古典模倣が困難な状態を誘起することに注意されたい。この意味で (b) $L = 1$ Theory のみであるにせよ、量子特有の構造をもつ Ansatz-A4 が、それをもたない Ansatz-A4ne の性能を凌駕している点は興味深い。

さらに、cNN の幅の違いに注目する。すなわち、(a) $L = 1$ Simulation において、量子回路の部分を Ansatz-A4ne と固定した上で cNN の幅を $n_0 = 10^2, 10^3, 10^4$ と変化させた結果について比べてみる。(ここで、理論曲線は幅が無限大の極限 $n_0 \rightarrow \infty$ であることに注意されたい。) その結果、 n_0 が

大きいほど、収束速度が速く、最終的な収束値が小さくなる様子が見て取れる。これは NTK 理論の明らかな効果である。

$L = 2$ の結果を示した (c, d) では、type-A の最終的なコスト値がシミュレーションと理論でよく一致している一方で、type-B では大きな差が見える。これは学習曲線における変化率の減少がシミュレーションでは著しく小さく、Ansatz-B と Ansatz-Bne では 10000 iteration でも定常状態まで到達しなかったためである。さらに長時間のシミュレーションができれば、両者は一致するはずである。

最後に、汎化性能について確認するために、学習済みのモデルに 100 個のテストデータを与えた。図 5.5 に ansatz 毎の汎化性能の観点での誤答率を示した。[159] では古典カーネル法を用いた際の誤答率が 45% であると示されており、Ansatz-A4, -A4ne はこれを凌駕している。これは qcNN が古典手法と比べて十分な表現力をもっていることを示している。もう一つの重要な点として、汎化性能が学習性能に連動している点が挙げられる。すなわち、学習誤差の小さな ansatz は汎化誤差も小さくなっている。この傾向は、機械学習の一般的な特徴であるところの「高い表現力は過学習を引き起こし、機械学習の性能を低下させてしまう」とは異なっている。しかし、提案モデルでは projected quantum kernel (PQK) を用いていることに注意されたい。これは [126] において良い汎化性能を持ち得ることが示唆されたものであり、その結果として高い汎化性能を示していると解釈することができる。提案モデルと、純量子モデルや純古典モデルとのさらなる性能比較は 5.6.3 項で示される。

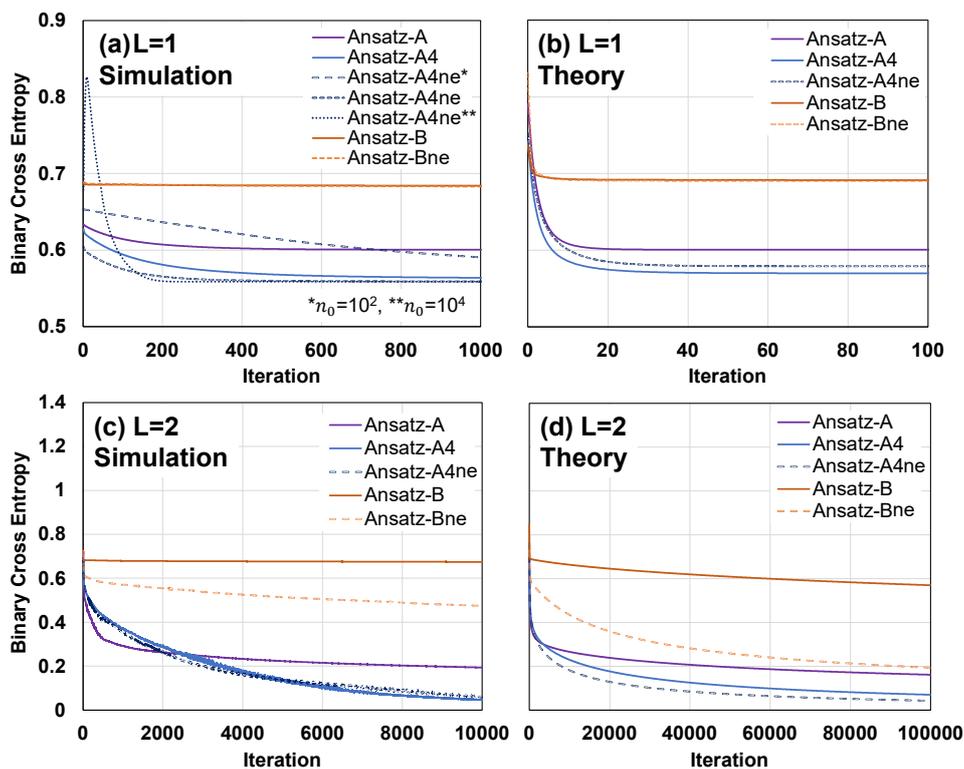


図 5.4: 分類タスクの学習曲線。それぞれの ansatz に対して同じデータセットを用いた結果。Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

5.6.2 Locality の機械学習性能への影響

本節では、提案モデルにおける randomizing unitary gate のサイズである Locality m に注目する。この値は提案モデルにおける reduced Hilbert space の次元 2^m を決定するハイパーパラメタである。

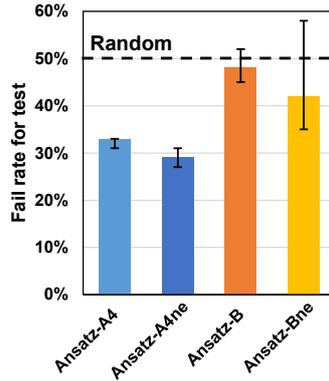


図 5.5: テストデータに対する誤答率 ($L = 1$, $n_0 = 10^3$). 棒グラフは 100 試行の median を表し, エラーバーの上端 (下端) は最高 (最低) 値を表す. 破線は 2 クラス分類における random guess の誤答率 50% を示す. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

[126] で指摘されているように, システムサイズが大きすぎると機械学習性能が低下するため, m の選択には注意が必要である. また, m は QNTK の固有値分布に影響を与えることで, 学習過程における収束速度に大きく影響する. そのような直接的な観点に加え, 量子回路中のランダム回路が量子効果を与えることを考慮し, 本節ではある系を想定して機械学習タスクを実行することで, locality m が収束速度と結果的な性能に与える影響を調べた. 設定したのは Heart Disease dataset [160] を用いた分類タスクである. このデータセットは特徴量が 12 あり, これを各 qubit に 1 つずつエンコードするために 12-qubit 系を用いた. タスクの目的は, 学習データを用いて患者が心疾患を患っているかを予測するモデルを構築することである. 学習データ数は 100 とし, 半が心疾患を患っている患者に関するデータである. 使用するモデルには, 提案 qcNN の $L = 1$ を用い, 同じデータセットに対して $m = 1, 2, 3, 4, 6$ を比較した. コスト関数を含むその他の設定は 5.6.1 項における分類タスクと同様である.

理論式としては, 式 (5.19) で与えられる学習過程の時間発展を用いた. すなわち, qcNN の幅は無限大に相当する. 得られた学習曲線を図 5.6 に記す. 期待通り, 収束速度とコスト関数の収束値は m に大きく依存している. この結果を解釈するために, まず学習曲線が QNTK $\Theta_Q^{(1)}(\mathbf{x}, \mathbf{x}')$ の固有値で与えられることを思い出す. 但し, \mathbf{x} は入力ベクトルである. 具体的には, 式 (5.19) で明示したように, 大きな固有値 λ_j に対応するインデックス j の要素は速く収束し, 一方の固有値が小さいものは遅く収束する. その結果, QNTK の固有値分布が学習過程の収束の挙動を支配するのである. 特に, 小さい固有値の割合が収束速度を特徴づける. 本研究のシミュレーションでは, 大きな m ほど固有値全体の値が小さくなる様子が観察された. この効果により, m が大きなものほどゆっくりと収束している様子が図 5.6 から見て取れる. 一方で, m が大きなほど固有値分布の分散は小さくなる様子も観察された. この結果, $m = 2$ の時の最小固有値が $m = 1$ の時のものよりも大きくなった. すなわち, 全体としては $m = 2$ の方が $m = 1$ よりもコストを最小化する点では優れている. これは一般に, m が大きいほど必ずしも高い性能を示す訳でなく, データ構造に応じた適した m が存在することを示している. 巨大なランダム量子回路を古典シミュレートすることが困難であることを考慮すると, これは提案 qcNN の量子パートの制限とも見える.

最後に, 最終的なコスト値に着目する. この値は未知データの分類性能を決める値であり, ansatz のタイプに大きく依存する. 特に注目すべきは Ansatz-A4ne と Ansatz-Bne が全ての m において最良の性能を示している点である. これらの ansatz は CNOT-gate を含んでおらず, 効率的な古典模倣が可能な状態である. これはすなわち, Heart Disease dataset は qcNN の分類性能を高めるには,

entanglement を含む量子的な特性が効果的ではなかったことを示している。この結果は [126] の主張とも合致している。すなわち、全ての量子機械学習システムが特定のデータセットに対する性能を向上させる訳ではないのである。次節では、CNOT-gate を含む qcNN が高い性能を示す別の問題設定について記す。

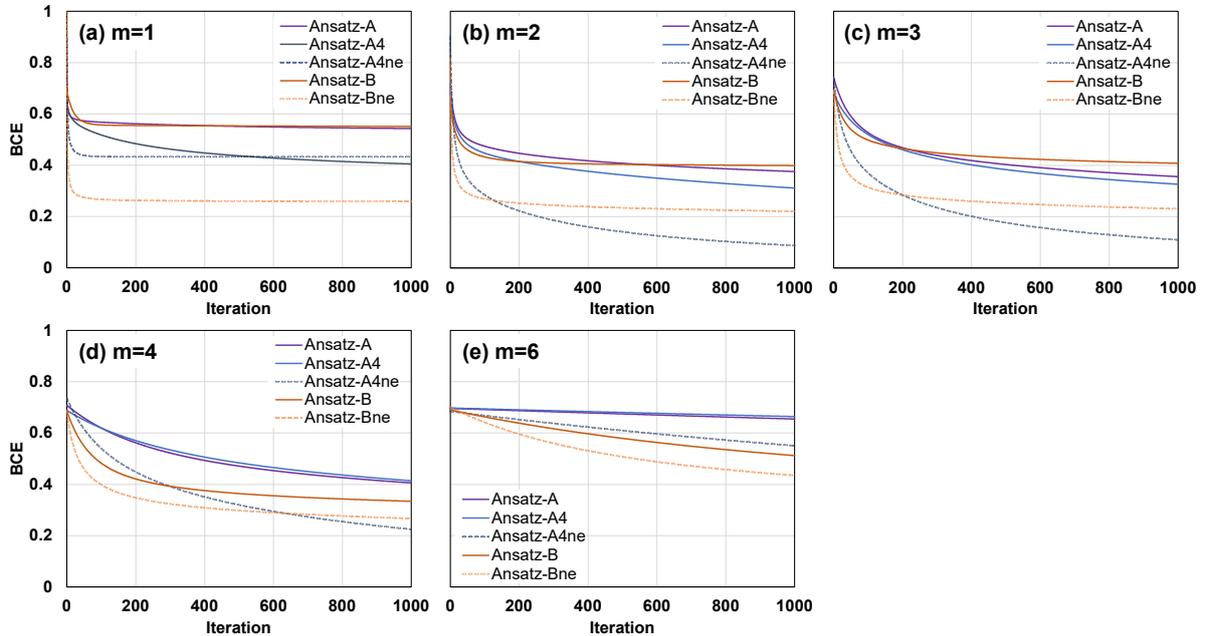


図 5.6: Heart Disease Data Set の分類タスクの学習曲線 (qcNN の理論曲線)。量子回路は 12-qubit, qcNN の古典パートの層数は $L = 1$ 。他の設定は、コスト関数を含め 5.6.1 項と同様。Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

5.6.3 純古典モデルと純量子モデルに対する qcNN の優位性

ここでは 5.4 で記した qcNN モデルの優位性を見るために、量子プロセスから生成されるデータセットの回帰、分類タスクについて調べる。本問題設定は、データセットの性質についてよく考慮した上で構成された量子機械学習モデルであれば、高い学習性能と汎化性能を両立できることを示し得ることを意図して構成されたものである。特に、例えば [127] のように、量子プロセスを介して生成されるデータセットに対する量子優位性の議論は既にいくつかなされている。提案 qcNN モデルがそのような好ましい特性をもち、実際に高い性能を発揮することを示す。しかも、純量子、純古典モデルに比べて、少ないパラメタ数もしくは小さい学習コストで量子データの生成プロセスを学習することを示す。

問題設定

データ生成プロセス まず、ここでいう量子プロセスから生成されるデータの意味を説明する。尚、以降は単に“量子データ”と呼ぶことにする。一般に、量子データはハミルトニアン $H(\mathbf{x}^a)$ 、例えば $\rho(\mathbf{x}^a) = e^{-iH(\mathbf{x}^a)}\rho_0e^{iH(\mathbf{x}^a)}$ でドライブされる量子系の出力状態と見ることができる。但し、入力 \mathbf{x}^a は、例えば温度のような、何らかの系の特徴量を表すものとする。状態 $\rho(\mathbf{x}^a)$ は、例えば実際の測定プロセスのような、詳細が不明な時間発展を経て、手元のデータになる。すなわち、出力

y^a はなんらかのオブザバブルによって得られる. 従って, y^a はプロセスや $\rho(\mathbf{x}^a)$ そのものの特徴を示している. このような学習データセット $\{\mathbf{x}^a, y^a\}_{a=1}^{N_D}$ があったとして, 解くべきタスクはタスクは, 未知の入力に対して良い汎化性能を示すような入出力のマッピングを近似する関数を構成することである. この問題は, 固体物理 [161] における the general quantum phase recognition (QPR) problem [162, 163, 114] と関係しており, 古典的には本質的に扱うことが難しく, いくつかの量子機械学習的手法では効率的に扱えることが示されている. [164, 165, 166]

ここで上記の問題を, 以下のようなデータセット $\{\mathbf{x}^a, y^a\}_{a=1}^{N_D}$ として扱う. まず, 入力データ $\{\mathbf{x}^a\}_{a=1}^{N_D}$ は, n 次元の一様分布 $[0, 2\pi]^n$ からサンプルされるものとする. ここで $\rho(\mathbf{x}^a)$ を詳細不明の量子プロセス $U_{\text{enc}}(\mathbf{x}) = e^{-iH(\mathbf{x})}$ から生成されるとしよう. シミュレーションでは, 図 5.7(a) で表される量子回路を用いた. すなわち, RX -gate の後にランダムな multi-qubit ユニタリ演算子 U_{random} を作用させたものである (詳細は付録 A を参照されたい). 出力 y^a はタスクに依って異なる. まず, 回帰タスク向けには, $y^a = cg(\mathbf{x}^a) + \epsilon^a$ とした. 但し, $g(\mathbf{x}) = \text{Tr}[\rho(\mathbf{x})O]$, ϵ は $\text{Var}[\epsilon] = 10^{-4}$ のガウシアンノイズである. 測定プロセスはいくらかの誤差が交じると仮定し, 不明なオブザバブル O が作用するとした. また, c は正規化定数であり $\text{Var}[g(\mathbf{x})] = 1$ を満たす. 次に, 分類タスク向けには, $g(\mathbf{x}^a) \geq n/2$ であれば $y^a = 1$, そうでなければ $y^a = 0$ とした. 但し, $g(\mathbf{x}) = \text{Tr}[\rho(\mathbf{x})O]$ である. シミュレーションでは, $O = \bigotimes_{i=1}^n (\sigma_z^{(i)} + \mathbf{1}^{(i)})/2$ とした. 但し, $\sigma_z^{(i)}$ は Pauli z 演算子, $\mathbf{1}^{(i)}$ は i -qubit の恒等演算子である. 学習データ数は, 回帰タスク向けには $N_D = 1000$, 分類タスク向けには $N_D = 3000$ とした. さらに, 汎化性能を検証するためのテストデータ数は, 回帰/分類ともに $N_{\text{test}} = 100$ とした.

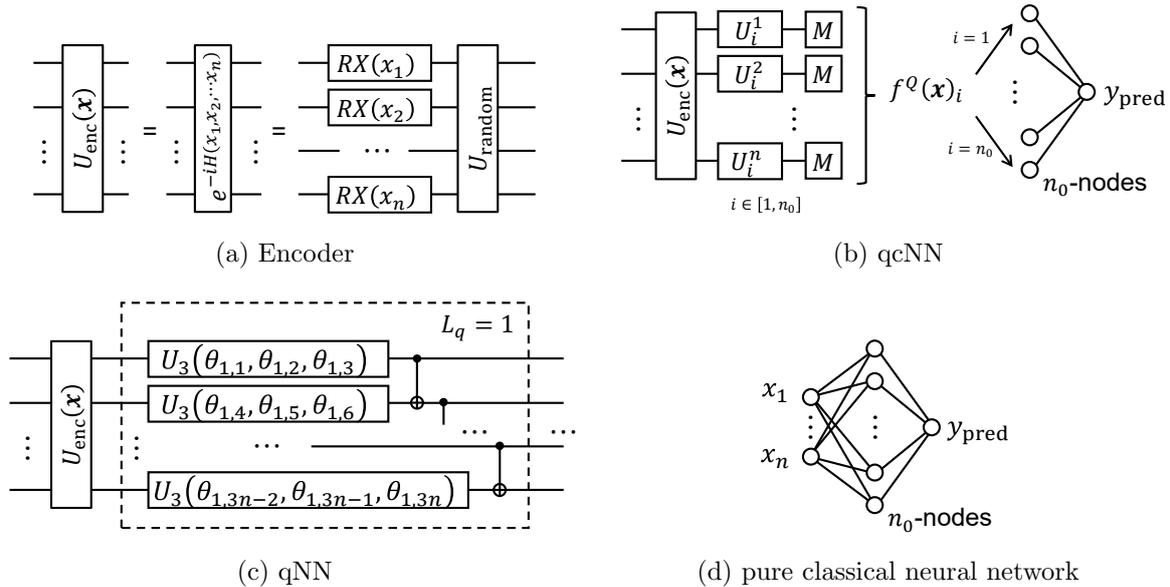


図 5.7: 5.6.1 項で用いたモデル. (a) 量子データを生成する量子回路. (b) 量子-古典ハイブリッドニューラルネットワーク (qcNN), (c) 量子ニューラルネットワーク (qNN), (d) 古典ニューラルネットワーク (cNN). (c) は $L_q = 10$ とした (10 層の qNN). 図中の M という箱は測定を表し, $U_3(\alpha, \beta, \gamma)$ は 3 つの Euler 角に関する single-qubit 回転ゲートである. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

学習モデル 学習モデルとしては図 5.7 に示した 3 種を用いた. まず, (b) に示したのが提案 qcNN モデルである. ポイントは, 学習データの生成源と同じエンコーダ $U_{\text{enc}}(\mathbf{x})$ を含んでいる点

である。これは qcNN が量子データ $\rho(\mathbf{x}^a)$ に直接アクセスできることに対応している。エンコーダの後には 1-qubit gate, すなわち $m = 1$ のランダムユニタリ演算子を作用させ、オブザバブル $O = \bigotimes_{i=1}^n (\sigma_z^{(i)} + \mathbf{1}^{(i)})/2 = \bigotimes_{i=1}^n M^{(i)}$ により測定を行ったものが量子パートの出力になる。ここでオブザバブルはデータ生成源と同じものを用いていることに注意されたい。しかしながら、ランダムユニタリ演算子を作用させていることから、オブザバブルの観点で一般性を損なわない。測定結果の期待値は、1 層の cNN ($L=1$) に入力される。出力ノードの活性化関数はタスク毎に以下のように異なる。すなわち、回帰タスクには恒等関数を、分類タスクにはシグモイド関数 $\sigma(q) = 1/(1 + e^{-q})$ を選択した。最後に cNN は、回帰タスクでは出力値 y_{pred} はそのままの値を用い、分類タスクでは閾値 0.5 よりも大きいものを $y_{\text{pred}} = 1$ とし、それ以外を $y_{\text{pred}} = 0$ とした。ここで [148] との違いに注意されたい。[148] では $\rho(\mathbf{x}^a)$ を近似するための classical shadow の生成にランダム測定を行い、shadow によって構成された古典データを用いた古典機械学習によって未知の $\rho(\mathbf{x})$ に対する y を予測している。一方、本研究では、ランダム測定から classical shadow を経ずに機械学習モデルを構成しており、明らかな計算優位性がある。

二つ目のモデルは、図 5.7 (c) に示した quantum neural network (qNN) である。このモデルは、qcNN と同様に、量子状態 $\rho(\mathbf{x}^a)$ を直接入力する。すなわち、シミュレーションでは、データベクトル \mathbf{x}^a は図 (a) と同じエンコーダ $U_{\text{enc}}(\mathbf{x})$ によって入力される。その後、図中の点線で表されるパラメタ付き量子回路 (PQC) が L_q 回作用される。qNN の出力は $y_{\text{pred}} = w \text{Tr}[\rho'(\mathbf{x})O']$ と計算される。但し、 $\rho'(\mathbf{x})$ は量子回路からの出力状態であり、 O' は $O' = \sigma_z^{(1)}$ である。そして、 w は scale factor である。すなわち、他の回路パラメタ $\{\theta_{i,j}\}$ と同様に最適化され、回帰タスクにおける出力範囲を調整する。尚、分類タスクでは $w = 1$ とした。一つ目のモデルと同様に、回帰モデルでは y_{pred} の値をそのまま用い、分類タスクにはバイナリ化した値を用いた。

三つ目のモデルは、図 5.7 (d) に示した 3 層の cNN である。すなわち、 n ノードの入力層、 n_0 ノードの隠れ層、1 ノードの出力層をもつ。入力層と隠れ層は全結合しており、出力層も隠れ層と全結合している。入力データ \mathbf{x}^a は、 i 番目の要素 x_i^a が第 i 番目のノードに入力される。活性化関数は層とタスクに応じて異なる。回帰タスクの場合、隠れ層にはシグモイド関数 $\sigma(q) = 1/(1 + e^{-q})$ 、出力層には恒等関数を用い、分類タスクの場合、隠れ層には ReLU $\sigma(q) = \max\{0, q\}$ 、出力層にはシグモイド関数を用いた。つまり、純古典モデルは量子データ $\rho(\mathbf{x}^a)$ もオブザバブル O も知らない状態である。

上記 3 モデル全てにおいて、4 種類の qubits 数 $n = \{2, 3, 4, 5\}$ で実験を行った。また、qcNN の cNN 部分のノード数、純古典 cNN のノード数として $n_0 = 10^3$ を設定した。オブザバブルの期待値は statevector を用いて計算し、パラメタ更新には optimizer として Adam [88] を用いた。

結果

回帰問題の結果 図 5.8 に回帰問題における学習過程とテストの結果を示した。縦軸には、予測値 y_{pred} と真値との差を二乗平均平方根誤差 (RMSE) を、横軸には qubit 数 n を取り、5 回の試行の平均と RMSE の標準偏差をプロットしてある。ここで、3 つのモデルはパラメタ数やオブザバブルの測定回数の点で、それぞれ異なるということに注意されたい。詳細は付録 A に記してある。特に qcNN のパラメタ数は、 n_0 の幅が同じにも関わらず、cNN よりも大幅に少ない。また、qcNN で必要なオブザバブルの測定回数は qNN に比べると大幅に少なくてすむ。なぜなら qcNN は最適化の過程で繰り返し量子回路を測定する必要が無いからである。

以上から、qcNN は他の 2 つに比べてコンパクトなモデルであると言える。それにも関わらず、図 5.8 から分かるように、qcNN は実験中の全ての qubit のケースで学習、テストの両方で高い性能を示した。これは主にいわゆる “inductive bias” [127] に依るところが大きい。すなわち、qcNN モデルは入力データを直接扱えるという点で、潜在的な優位性を秘めていたと言える。しかし、このバイアス

は qNN モデルにも同様のことが言えるが、こちらは qubit 数が増えるにつれて学習データ、テストデータ共に近似ができなくなっている。これは Hilbert space における表現力が十分ではないことによるものと考えられる。しかし、もし qNN モデルが十分な表現力を有していたとしても、その他の問題、例えば, barren plateau や測定数の増加など, qubit 数が大きい領域での学習には困難がある。また, qcNN や qNN では学習時とテスト時ではほぼ同じ性能を示している点は注目に値する。これはターゲットデータに対して過学習を起こしていないことを示している。一方で, cNN モデルではテスト時に性能が低下している。この差は恐らく, 量子モデルはデータセットの量子状態に直接アクセスできることが要因であろう。実際, inductive bias を活用することで高い汎化性能を獲得することが [127] でも示されている。

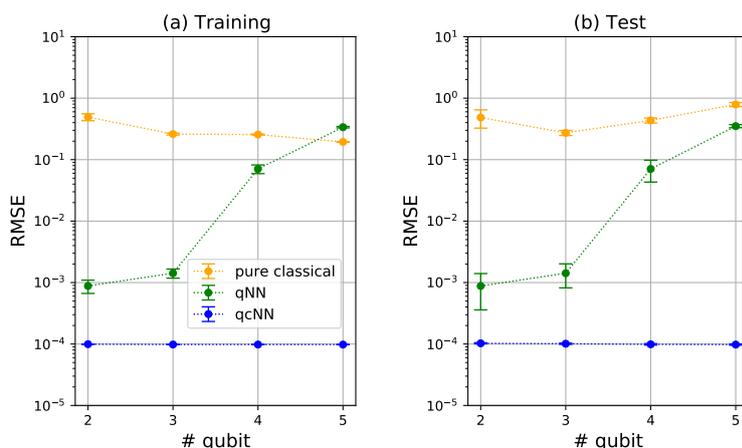


図 5.8: 回帰タスクの結果. (a) 学習データ, (b) テストデータ. 縦軸は root mean squared errors (RMSE), 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

分類問題の結果 図 5.9 に分類問題における結果を示した。縦軸には正解率を、横軸には qubit 数 n を取り、5 回の試行の平均と正解率の標準偏差をプロットしてある。このタスクでも回帰と同様の結果が観察された。すなわち, qcNN は実験中の全ての qubit のケースで学習、テストの両方で高い性能を示した。これも前述した要因と同じく, qcNN モデルの inductive bias と qNN モデルの表現力の不足であろう。また, 参考のために qcNN については, $N_{\text{ite}} = 3000$ かつ $n_0 = 3000$ の結果も示した (図 5.9 の ‘qvNN(tuned)’). ちなみに, 他のものは全て $N_{\text{ite}} = 1000$ かつ $n_0 = 1000$ である。この結果は古典パートの部分のチューニングにより性能を向上させることができることを示している。

shot noise の影響 最後に qcNN モデルを用いた回帰/分類タスクにおいて, shot 数 (測定回数) を変えたときにどう性能が変化するかを調べた。尚, 図 5.8, 5.9 に示した実験では statevector, すなわち shot 数は無限大であったことに注意されたい。問題設定は, データセットや学習モデルを含めて, 前の実験と同じである。得られた結果を図 5.10 に示した。(a) は回帰タスクの結果を RMSE でプロットしたもの, (b) は分類タスクの結果を正解率でプロットしたものである。両者とも qubit 数 n を横軸に取り, qcNN の古典パートの層数は $L = 1$ である。但し, $n = 5$ については $L = 2$ を取り, 第一層目と第二層目の幅は $n_0 = n_1 = 10^3$ とした。

回帰タスクの場合, cNN が二層あるケース ‘ $n = 5(L2)$ ’ を除く全てのケースで, RMSE ϵ と shot 数 N_{shot} の間に明らかな統計的な傾向 $\epsilon \sim O(1/\sqrt{N_{\text{shot}}})$ が見られた。図からは 10^4 -shot あれば statevector による理想状態に十分近い精度が得られる様子が見て取れる。また, cNN が二層のケー

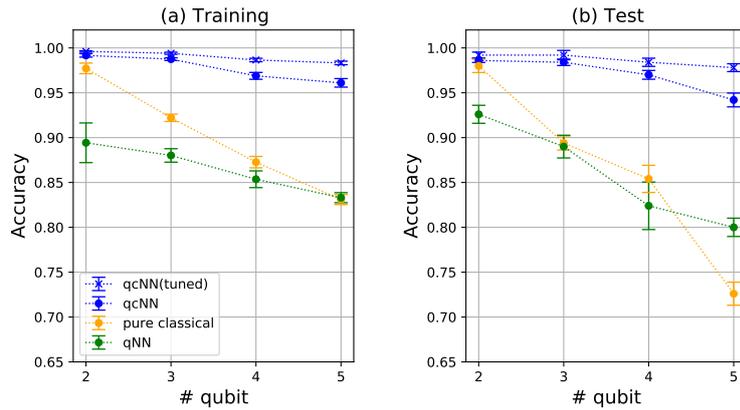


図 5.9: 分類タスクの結果. (a) 学習データ, (b) テストデータ. 縦軸は正解率, 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

スで, 特に qubit 数が小さい領域で RMSE が抑えられていることに注目されたい. これは cNN の高い非線形性により shot noise を相殺しているように見える. 図 5.10 の (b) に示した分類タスクの場合, 回帰タスクに比べて必要な shot 数が大幅に小さいことが見て取れる. 具体的には, 10^2 -shot で statevector と同等の正解率を示している. これは shot noise, もしくは cNN パートに入力されるデータに含まれる noise は, 分類タスクにはそれほど大きな影響を与えないことを意味している.

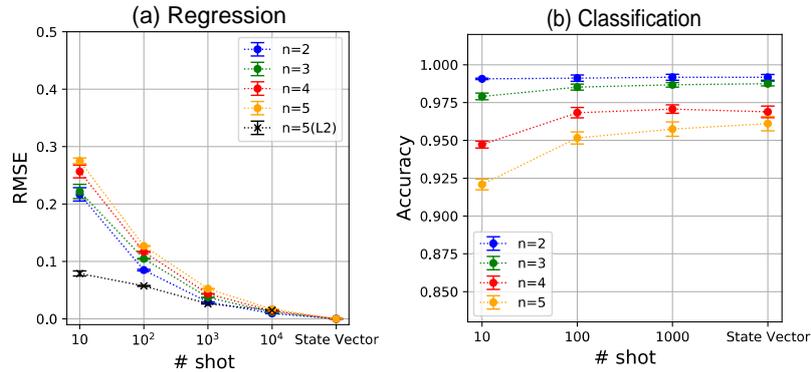


図 5.10: (a) 回帰タスクにおける shot noise の影響 (学習データ). 縦軸は root mean squared errors (RMSE), 横軸は qubit 数. (b) 分類タスクにおける shot noise の影響 (学習データ). 縦軸は正解率, 横軸は qubit 数. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

5.7 結論

本研究では, 量子データエンコーダの出力を古典ニューラルネットワーク (cNN) に入力することで, 古典の NTK 理論が直接適用できるフレームワークを提案し, 量子-古典ハイブリッドニューラルネットワーク (qcNN) について調べた. そして, 量子, 古典の両部分をランダムに初期化し, cNN のノード数 (幅) を無限大の極限に取ることで, 系全体に対して量子 NTK が定義でき, それが time-invariant になるため, 学習過程のダイナミクスを明示的に解析できることを示した. さらに, qcNN の最終的な出力が projected quantum kernel の非線形関数になることを示した. すなわち, 提

案 qcNN を用いることで、従来の量子カーネル法に比べて少ない computational complexity で回帰タスクや分類タスクを行える非自明な量子カーネルを実装できる。また、学習過程が解析的に扱えることから、qcNN が古典手法に比べて高い性能を示すデータセットが満たすべき条件を得た。さらに、量子プロセスから生成されたデータに対する学習タスクにおいて、提案 qcNN が cNN や qNN に比べて明確に高い性能を示すことを、数値シミュレーションで確認した。

量子カーネル法を用いたこれまでの研究や 5.6 節の結果が示すように、機械学習タスクの性能はデータエンコーダの設計やデータセットの構造に強く依存する。従って、ユーザーは与えられたデータセットに応じて、エンコーダを量子優位性が得られるように注意して設計する必要がある。実直なアプローチとしては、本研究では固定していた量子データエンコーダ部分を学習可能な形に置き換え、cNN の学習時に一緒に最適化するという方向が考えられる。実際、一般に深層学習ではデータエンコーダ（特徴量抽出）部とデータ処理部を合わせて学習している。つまり、そのような qNN-cNN ハイブリッドシステムは深層学習と同様の機能性を持ちうる。すなわち、さらなる予測性能の向上が期待でき、何らかの量子優位性が得られる可能性がある。しかしながら、qNN パートを学習する場合、スケーリング時には一般に勾配消失の問題が起きるため、比較的小さい qNN や適切な ansatz、コスト関数の選択を行う必要がある。これらは今後の課題とする。

第6章

効率的な量子異常検知の実装：Quantum generative model with optimal transport

本章では、量子状態を扱う生成モデルを提案する [46, 167]. 生成モデルは、画像処理や自然言語処理、異常検知など古典の機械学習の領域で様々なタスクで高い性能を発揮している。これらのモデルは、学習時に用いるサンプルデータが属する何らかの“クラス”をモデルが学習し、モデルの使用時にはこのクラスに属しているがサンプルデータセットには含まれない、新たなデータが生成可能である。これは量子計算の文脈では特に重要な意味を持ちうる。すなわち、一般に高次元系を扱う量子計算では、所望の状態の生成にかかる計算コストが重い。ここに生成モデルを用いてアプローチする。例えば、ある量子計算の解の集合（量子状態の集合）に対応するクラスを学習する。データセットに含まれるサンプル数は有限であるため、手元にあるのは離散的な量子状態である。しかし、一旦サンプルデータセットが属するクラスを適切に学習できれば、理想的には所望の状態に対応する量子状態を学習済みのモデルから容易に生成できるようになる。

これまで、いくつかの先行研究において量子状態を生成するアルゴリズムが提案されてきた。しかし、それらは“純粋状態の集合”の学習には対応していない。すなわち、混合状態として平均的な状態を扱っていたため、前述のような使い方はできない。本研究では、コスト関数として前述の用途に適したものの、純粋状態のアンサンブルを直接扱うことのできる最適輸送距離を提案する。合わせて、スケールリングに向けてコスト関数の local 化などを導入する。古典の生成モデルにおいて重要な用途として知られているものの一つに、異常検知タスクがある。そこで本章では、量子生成モデルのデモンストレーションとして量子状態の異常検知タスクを示す。これは今後、量子コンピューティングの発展に伴って量子デバイスが普及した際のヘルスチェックなどに用いることができるなど、応用上も重要な例である。

6.1 導入

6.1.1 Implicit Generative Model

生成モデルは与えられた学習データを生成する未知の確率分布の近似器として用いられる。生成モデルを設計する基本的な戦略は、以下の通りである。 \mathcal{X} をランダムな変数の空間とし、与えられた学習データ $\{\mathbf{x}_i\}_{i=1}^M \in \mathcal{X}^M$ の背後にある確率分布を $\alpha(\mathbf{x})$ と仮定する。この下で、パラメタ化された確率分布 $\beta_\theta(\mathbf{x})$ を用意し、学習データに対して定義される損失関数の近似値を最小化するようなパラメタ θ を学習する。

一般に、生成モデルは prescribed models と implicit models の2つに分類される。前者は、パラメ

タ θ により陽に $\beta_\theta(\mathbf{x})$ を表す. すると $\beta_\theta(\mathbf{x}_i)$ の対数尤度関数が計算できるため, 最尤推定により最適なパラメタを決定することができる. これは Kullback-Leibler divergence $KL(\alpha||\beta_\theta)$ を最小化することに対応する. 後者は, $\beta_\theta(\mathbf{x})$ の陽な形では与えられないモデルを用いる. このモデルの重要な特徴として, 隠れた低次元多様体上のランダム変数の確率分布により, 簡便に表現することができる点がある. また, データ生成プロセスが潜在変数からデータへの物理的なプロセスとして解釈できる [168]. implicit generative model の例としては, Generative adversarial networks (GAN) [169] などがある. 本研究では, この implicit generative model に着目する.

ここで, implicit generative model について, より具体的に考えてみる. implicit generative model はランダムな潜在変数 \mathbf{z} から \mathcal{X} への写像として表現される. 尚, \mathbf{z} は, サンプル空間の次元 N_x に比べて非常に小さな次元 N_z をもつ潜在変数空間 \mathcal{Z} に属する. 但し, 潜在変数 \mathbf{z} は, 一様分布や Gauss 分布などの既知の分布 $\gamma(\mathbf{z})$ に従うものとする. すなわち, implicit generative model は $\beta_\theta = G_\theta \# \gamma$ によって与えられる. 尚, $\#$ は push-forward operator [170] と呼ばれ, この場合は \mathcal{Z} における分布 γ が写像 G_θ により \mathcal{X} における確率分布へ移されることを表す. この implicit generative model は, モデル分布から生成されたサンプルセットが学習データセットへ近付くように学習される. この際, 下記のように, 何らかの近似的損失関数 \mathcal{L} が最小化するようパラメタ θ の学習を進める.

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\hat{\alpha}_M, \hat{\beta}_{\theta, M_g}). \quad (6.1)$$

但し, $\hat{\alpha}_M(\mathbf{x}), \hat{\beta}_{\theta, M_g} = G_\theta \# \hat{\gamma}_{M_g}(\mathbf{z})$ は, サンプルデータセット $\{\mathbf{x}_i\}_{i=1}^M$ と潜在変数の集合 $\{\mathbf{z}_i\}_{i=1}^{M_g}$ から定義される経験的分布を表す. 尚, それぞれ以下のような確率分布に従う.

$$\hat{\alpha}_M(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M \delta(\mathbf{x} - \mathbf{x}_i), \quad \hat{\gamma}_{M_g}(\mathbf{z}) = \frac{1}{M_g} \sum_{i=1}^{M_g} \delta(\mathbf{z} - \mathbf{z}_i). \quad (6.2)$$

6.1.2 最適輸送距離 (Optimal Transport Loss; OTL)

OTL は, 画像解析や自然言語処理, 金融 [171, 172, 170, 173] などの幅広い分野で用いられている metric である. また, 生成モデルの損失関数として広く用いられているが, これは特に連続値の経験分布の比較に用いることができる点が多い. OTL は, 確率分布 α を別の分布 β に移す際の最小限のコストで定義される:

定義 6.1 (最適輸送距離 [174])

$$\begin{aligned} \mathcal{L}_c(\alpha, \beta) &= \min_{\pi} \int c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}), \\ \text{subject to } &\int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \beta(\mathbf{y}), \\ &\int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \alpha(\mathbf{x}), \quad \pi(\mathbf{x}, \mathbf{y}) \geq 0, \end{aligned}$$

但し, $c(\mathbf{x}, \mathbf{y}) \geq 0$ は, $\mathcal{X} \times \mathcal{X}$ 上の \mathbf{x} から \mathbf{y} への輸送コストを表す非負の関数であり, ground cost と呼ばれる. また, $\mathcal{L}_c(\alpha, \beta)$ を最小化する集合 π を最適輸送計画 (optimal transport plan) と呼ぶ.

また, 一般に OTL は距離の公理を満たさないが, ground cost を以下のような metric function に取れば距離の公理を満たす:

定義 6.2 (p-Wasserstein distance [175]) ground cost $c(\mathbf{x}, \mathbf{y})$ が, metric function $d(\mathbf{x}, \mathbf{y})$ と正定数 p を用いて $c(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathbf{y})^p$ と表される時, OTL は距離の公理を満たし

$$\mathcal{W}_p(\alpha, \beta) = \mathcal{L}_{d^p}(\alpha, \beta)^{1/p}$$

を p-Wasserstein distance と呼ぶ.

p-Wasserstein distance は確率分布間の距離を表す条件を満たす. すなわち, 任意の分布 α, β, γ に対し, 以下の特性を満たす:

$$\begin{aligned}\mathcal{W}_p(\alpha, \beta) &\geq 0, \\ \mathcal{W}_p(\alpha, \beta) &= \mathcal{W}_p(\beta, \alpha), \\ \mathcal{W}_p(\alpha, \beta) &= 0 \Leftrightarrow \alpha = \beta, \\ \mathcal{W}_p(\alpha, \gamma) &\leq \mathcal{W}_p(\alpha, \beta) + \mathcal{W}_p(\beta, \gamma).\end{aligned}$$

また, 一般に OTL $\mathcal{L}_c(\alpha, \beta_\theta)$ を確率分布 α, β_θ に対して最小化することは難しい. そのため, 式 (6.1) に示したように, 経験分布 (6.2) で OTL を近似したものを最小化することを考える.

定義 6.3 (OTL の経験的推定器 [175])

$$\begin{aligned}\mathcal{L}_c(\hat{\alpha}_M, \hat{\beta}_{\theta, M_g}) &= \min_{\{\pi_{i,j}\}_{i,j=1}^{M, M_g}} \sum_{i=1}^M \sum_{j=1}^{M_g} c(\mathbf{x}_i, G_\theta(\mathbf{z}_j)) \pi_{i,j}, \\ \text{subject to } \sum_{i=1}^M \pi_{i,j} &= \frac{1}{M_g}, \quad \sum_{j=1}^{M_g} \pi_{i,j} = \frac{1}{M}, \quad \pi_{i,j} \geq 0.\end{aligned}\tag{6.3}$$

経験的推定器 $\mathcal{L}_c(\hat{\alpha}_M, \hat{\beta}_{\theta, M})$ は $M = M_g \rightarrow \infty$ の極限で $\mathcal{L}_c(\alpha, \beta_\theta)$ に収束する. ここで, 一般に収束速度がサンプル空間 \mathcal{X} の次元 N_x に対して $O(M^{-1/N_x})$ となることを思い出す [176]. 一方, [177] の定義 4 で定義される upper Wasserstein dimension $d_p^*(\alpha)$ を用いると, p-Wasserstein distance の収束速度が次のように表せる:

定理 6.1 (p-Wasserstein distance の収束速度) upper Wasserstein dimension $d_p^*(\alpha)$ に対し, s が $d_p^*(\alpha)$ より大きいという条件の下では, 以下の関係が成立する:

$$\mathbb{E}[\mathcal{W}_p(\alpha, \hat{\alpha}_M)] \lesssim O(M^{-1/s}),$$

但し, 期待値 \mathbb{E} は経験分布 $\hat{\alpha}_M$ に対して取るものとする.

これは直感的に考えると, upper Wasserstein dimension $d_p^*(\alpha)$ は, 確率分布 α のサポート次元と解釈することができ, それは implicit generative model の潜在空間の次元 N_z に対応すると解釈できる. さらに, p-Wasserstein distance の特性を活用すると, 定理 6.1 から以下が直ちに帰結される:

系 6.1 (p-Wasserstein distance の収束速度: 同一の分布からサンプルされた場合) 経験分布 $\hat{\alpha}_{1, M}, \hat{\alpha}_{2, M}$ を同一の分布 α からサンプルされた二つの経験分布とする. 両者のサンプル数を M とした時, $s > d_p^*(\alpha)$ に対して以下の関係が成立する:

$$\mathbb{E}[\mathcal{W}_p(\hat{\alpha}_{1, M}, \hat{\alpha}_{2, M})] \lesssim O(M^{-1/s}),\tag{6.4}$$

但し, 期待値 \mathbb{E} は経験分布 $\hat{\alpha}_{1, M}$ と $\hat{\alpha}_{2, M}$ に対して取るものとする.

系 6.2 (p-Wasserstein distance の収束速度: 異なる分布からサンプルされた場合) 確率分布 α, β_θ の upper Wasserstein dimension を最大 d_p^* とすると, $s > d_p^*$ に対して以下の関係が成立する:

$$\mathbb{E} \left[|\mathcal{W}_p(\alpha, \beta_\theta) - \mathcal{W}_p(\hat{\alpha}_M, \hat{\beta}_{\theta, M})| \right] \lesssim O(M^{-1/s}),\tag{6.5}$$

但し, 期待値 \mathbb{E} は経験分布 $\hat{\alpha}_M$ と $\hat{\beta}_{\theta, M}$ に対して取るものとする.

これらの系は、学習データの intrinsic dimension と潜在空間の次元 N_z が十分小さい場合には、経験的推定器 (6.3) が良い推定器となることを示している。なぜなら、Wasserstein dimension d_p^* が、学習データの intrinsic dimension と潜在空間の次元とほぼ一致するからである。6.3 節では、OTL が p-Wasserstein distance でない場合であっても、同様の収束則を示すことを数値実験により確認する。

6.2 提案手法

本節では、量子状態の分布を直接的に扱うための新たな OTL と、それをを用いた量子生成モデルを提案する。

6.2.1 Local ground cost を用いた Optimal transport loss

まず、量子状態の分布を直接扱うための OTL を検討する。発想の起点はシンプルであり、式 (6.3) を古典データベクトル $c(\mathbf{x}, \mathbf{y})$ ではなく、量子状態 $c(|\psi\rangle, |\phi\rangle)$ へ適用し、ground cost を計算することを考える。結果、量子アンサンブル $\{|\psi_i\rangle\}$ と $\{|\phi_i\rangle\}$ 間の OTL として、以下が得られる：

$$\begin{aligned} \mathcal{L}_c(\{|\psi_i\rangle\}, \{|\phi_i\rangle\}) &= \min_{\{\pi_{i,j}\}} \sum_{i,j} c(|\psi_i\rangle, |\phi_j\rangle) \pi_{i,j}, \\ \text{subject to } \sum_i \pi_{i,j} &= q_j, \quad \sum_j \pi_{i,j} = p_i, \quad \pi_{i,j} \geq 0, \end{aligned} \quad (6.6)$$

但し、 p_i, q_j は $|\psi_i\rangle, |\phi_j\rangle$ が得られる確率を表す。この OTL は、[178] で議論されているように、混合状態間の距離、例えば $\sum_i p_i |\psi_i\rangle \langle \psi_i|$ と $\sum_j q_j |\phi_j\rangle \langle \phi_j|$ の距離を扱える。しかしながら、構成する量子状態の対称性が高い場合（例えば、ブロッホ球上の赤道上に一様に純粋状態が分布する場合など）では、アンサンブルに相当する混合状態は、多くの情報を失ってしまい、元の量子状態の分布を再構成することは不可能である。従って、この OTL では、本研究の目的である量子アンサンブルの個々の分布を保ったまま学習を行うことはできない。

そこで考えるべきは、ground cost $c(|\psi\rangle, |\phi\rangle)$ をどのように構成するかである。直ちに考えられる候補は trace distance であろう：

定義 6.4 (純粋状態に対する trace distance [179])

$$c_{\text{tr}}(|\psi\rangle, |\phi\rangle) = \sqrt{1 - |\langle \psi | \phi \rangle|^2}. \quad (6.7)$$

trace distance は距離の公理を満たすため、量子アンサンブルに対する p-Wasserstein distance を

$$\mathcal{W}_p(\{|\psi_i\rangle\}, \{|\phi_i\rangle\}) = \mathcal{L}_{d^p}(\{|\psi_i\rangle\}, \{|\phi_i\rangle\})^{1/p},$$

と定義することができる。これにより系 6.1 の有用な特性を利用することができる。また、trace distance は、swap test [180] や inversion test [74] を用いることで、量子コンピュータにより比較的簡便に計算が行える。

ここで重要な点を指摘する。以下でも改めて記すが、本研究の目的は、与えられた量子アンサンブルを最も良く近似するような量子アンサンブルを生成できる量子回路を設計することである。このタスクは、parameterized quantum circuit (PQC) に対して最急降下法を適用することで実行できるが、ナイーブな実装では勾配消失が起こる。すなわち、qubit 数が大きくなるにつれ指数関数的に勾配ベクトルの大きさが消失していく [92]。これを回避するために様々な手法（例えば [94] など）が提案されているが、それらに共通する要件は“local cost”を用いることである。この意味するところを説明するために、以下の設定を考える。まず、量子状態 $|\phi\rangle$ が $|\phi\rangle = U|0\rangle^{\otimes n}$ により与えられるとす

る。但し, U はユニタリ演算子 (具体的には, 生成モデルを特徴づける PQC $U(\theta)$), n は qubit 数である。ここで trace distance を fidelity $|\langle\psi|\phi\rangle|^2 = |\langle\psi|U|0\rangle^{\otimes n}|^2$ に基づいて計算する。これは, 状態 $U^\dagger|\psi\rangle$ に対して, 計算基底の下で “Global” 測定を行った際に全ての qubit で zero が観測される確率で求められる。すなわち, trace distance は global cost なのである。従って, fidelity-based な学習手法には勾配消失問題が生じてしまう。そこで上記で定義した損失関数に, 以下の手法を導入することで, localized fidelity measurement に基づいた損失関数に変換する。

定義 6.5 (local 測定のみで計算される ground cost [181, 182])

$$\begin{aligned} c_{\text{local}}(|\psi\rangle, |\phi\rangle) &= c_{\text{local}}(|\psi\rangle, U|0\rangle^{\otimes n}) \\ &= \sqrt{\frac{1}{n} \sum_{k=1}^n (1 - p^{(k)})}, \\ p^{(k)} &= \text{Tr} [P_0^k U^\dagger |\psi\rangle \langle\psi| U], \\ P_0^k &= \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \overbrace{|0\rangle\langle 0|_k}^{\text{k-th bit}} \otimes \cdots \otimes \mathbb{I}_n, \end{aligned} \tag{6.8}$$

但し, n は qubit 数, \mathbb{I}_i は恒等演算子, $|0\rangle\langle 0|_i$ は第 i 番目の qubit へ作用する射影演算子, $p^{(k)}$ は, 第 k 番目の qubit で 0 が観測される確率を表す。

式 (6.8) は確かに local cost になっており, 6.4 節で触れられるいくつかの追加条件を満たせば, 勾配消失問題を回避し, 効率的に学習を行うことができる。しかし, ここで注意が必要な点がある。それは $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ は二つの量子状態の距離を表している訳では無いという点である。なぜならこの metric には対称性がなく, 三角不等式を満たさないからである。但し, $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ は常に非負であり, $|\psi\rangle = |\phi\rangle$ の時のみ zero になる。すなわち, $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ の関数は divergence の要件は満たす。一般に, ground cost が divergence の場合, それによって定義される OTL も divergence になることを示すことができる。証明は [167] の Appendix A を参照されたい。

命題 6.1 ground cost $c(\mathbf{x}, \mathbf{y})$ が

$$\begin{aligned} c(\mathbf{x}, \mathbf{y}) &\geq 0, \\ c(\mathbf{x}, \mathbf{y}) &= 0 \quad \text{iff} \quad \mathbf{x} = \mathbf{y}, \end{aligned}$$

を満たす divergence の場合, OTL $\mathcal{L}_c(\alpha, \beta)$ with $c(\mathbf{x}, \mathbf{y})$ も divergence である。すなわち, $\mathcal{L}_c(\alpha, \beta)$ は, 任意の確率分布 α と β に対し, 以下を満たす:

$$\begin{aligned} \mathcal{L}_c(\alpha, \beta) &\geq 0, \\ \mathcal{L}_c(\alpha, \beta) &= 0 \quad \text{iff} \quad \alpha = \beta. \end{aligned}$$

従って, 式 (6.8) で与えられる local ground cost $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ を用いて計算される式 (6.6) に基づく OTL $\mathcal{L}_c(\{|\psi_i\rangle\}, \{|\phi_i\rangle\})$ は divergence である。これは, 与えられた量子アンサンブルと生成モデルの出力による集合の違いを評価するのに, $\mathcal{L}_c(\{|\psi_i\rangle\}, \{|\phi_i\rangle\})$ を用いることができることを意味している。同時に, 勾配消失も回避する目的のために, fidelity-based measurement の適用と OTL の距離の特性を失った点に注意されたい。従って, これまで述べてきた以下のような好ましい特性, 具体的には定理 6.1, 系 6.1, 系 6.2 である。しかし, 6.3 節において, ground cost が divergence であっても類似の性質を示すことが分かる。

6.2.2 学習アルゴリズム

本タスクの目的は、与えられたアンサンブル $\{|\psi_i\rangle\}_{i=1}^M$ を近似する量子アンサンブルを出力するような implicit generative model を学習することである。提案する生成モデルは、6.1.1 項で古典のケースに関して記したように、調整可能なパラメタと潜在変数を含む。具体的には、以下で記される implicit generative model である：

定義 6.6 (量子回路を用いた implicit generative model) 初期状態 $|0\rangle^{\otimes n}$ とパラメタ付き量子回路 $U(\mathbf{z}, \theta)$ を用いて、implicit generative model を

$$|\phi_\theta(\mathbf{z})\rangle = U(\mathbf{z}, \theta) |0\rangle^{\otimes n}. \quad (6.9)$$

と定義する。但し、 θ は調整可能なパラメタベクトル、 \mathbf{z} は既知の確率分布に従う潜在変数ベクトルを表す。尚、 θ と \mathbf{z} は回転ゲート $U(\mathbf{z}, \theta)$ の回転角として埋め込まれる。

ちなみに、似た回路モデルが meta-VQE [183] でも提案されている。これはランダムな潜在変数 \mathbf{z} の代わりに、例えば原子核間距離などの物理パラメタを埋め込むモデルである。また、[184] で提案されているモデルでは、潜在変数 \mathbf{z} を $|\phi_\theta(\mathbf{z})\rangle = U(\theta) |\mathbf{z}\rangle$ の形で計算基底に対応付けている。しかし、この埋め込み方では異なる潜在変数に対応する状態が互いに直交しており、Hilbert 空間中の小さな変化を潜在変数の変化として表現することができない。一方、本研究の提案モデル (6.9) では、対象の状態が \mathbf{z} に対応するモデルの表現空間に含まれていれば、対象の状態の変化が \mathbf{z} の変化として対応付けられる。さらに、提案モデルは、調整パラメタの解析微分が parameter shift rule [105, 185] で計算できるだけでなく、潜在変数 \mathbf{z} の微分も同様に計算できるという利点がある。この特徴は 6.4 節で示される、異常検知タスクにおいて重要な役割を果たす。

次に学習に用いる損失関数について考える。OTL の経験的推定器として、以下の値を用いる。この値は学習データ $\{|\psi_i\rangle\}_{i=1}^M$ とサンプルされた潜在変数 $\{\mathbf{z}_j\}_{j=1}^{M_g}$ により計算される。：

$$\begin{aligned} \mathcal{L}_{c_{\text{local}}} \left(\{|\psi_i\rangle\}_{i=1}^M, \{|\phi_\theta(\mathbf{z}_j)\rangle\}_{j=1}^{M_g} \right) &= \min_{\{\pi_{i,j}\}_{i,j=1}^{M, M_g}} \sum_{i=1}^M \sum_{j=1}^{M_g} c_{\text{local}, i, j} \pi_{i, j}, \\ \text{subject to } \sum_{i=1}^M \pi_{i, j} &= \frac{1}{M_g}, \quad \sum_{j=1}^{M_g} \pi_{i, j} = \frac{1}{M}, \quad \pi_{i, j} \geq 0. \end{aligned} \quad (6.10)$$

但し、 $c_{\text{local}, i, j}$ は

$$\begin{aligned} c_{\text{local}, i, j} &= \sqrt{\frac{1}{n} \sum_{k=1}^n (1 - p_{i, j}^{(k)})}, \\ p_{i, j}^{(k)} &= \text{Tr} [P_0^k U^\dagger(\mathbf{z}_j, \theta) |\psi_i\rangle \langle \psi_i| U(\mathbf{z}_j, \theta)], \\ P_0^k &= \mathbb{I}_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \overbrace{|0\rangle \langle 0|}_k \otimes \cdots \otimes \mathbb{I}_n. \end{aligned} \quad (6.11)$$

で与えられる ground cost である。尚、 $c_{\text{local}, i, j}$ は実用上は有限回の測定によって推定されることに注意されたい。そこで理想的には $c_{\text{local}, i, j}$ であるものを、 N_s 回の測定によって推定したものを $\tilde{c}_{\text{local}, i, j}^{(N_s)}$ と記すことにする。また、この場合の OTL は $\mathcal{L}_{\tilde{c}_{\text{local}}^{(N_s)}}$ と記す。

提案 OTL (6.10) に基づいて、目的のタスクを行うアルゴリズムの pseudo-code を Algorithm 1 に記す。パラメタ更新に必要な量子状態 $|\psi_i\rangle$ の数は、step-3 の $O(MM_gN_s)$ と step-5 の

$O(\max(M, M_g)N_sN_p)$ で決まる. 尚, これらには parameter shift rule [105, 185] に必要な数も含まれる.

Algorithm 1 最適輸送距離 (6.10) を用いた学習アルゴリズム

Input: 量子回路モデル $U(\mathbf{z}, \theta)$, 初期パラメタ θ , 学習率 ε , アンサンブル $\{|\psi_i\rangle\}_i^M$

Output: 入力アンサンブルをよく近似するアンサンブルを出力する量子回路

- 1: **repeat**
 - 2: 潜在空間の分布からサンプルされた潜在変数 $\{\mathbf{z}_j\}_{j=1}^{M_g}$ を生成.
 - 3: N_s 回の測定から得られた $\{|\psi_i\rangle\}_{i=1}^M$ と $\{U(\mathbf{z}_j, \theta)\}_{j=1}^{M_g}$ から ground cost $\{\tilde{c}_{\text{local},i,j}^{(N_s)}\}_{i,j=1}^{M,M_g}$ を式 (6.8) に基づいて推定.
 - 4: 線形計画 (6.10) を解くことで最適輸送計画 $\pi_{i,j}$ を算出.
 - 5: $\pi_{i,j}$ と $\left\{\frac{\partial}{\partial\theta_k}\tilde{c}_{\text{local},i,j}^{(N_s)}\right\}_{i,j,k=1}^{M,M_g,N_p}$ から parameter shift rule を用いて勾配 $\left\{\frac{\partial}{\partial\theta_k}\mathcal{L}_{\text{cllocal}}\right\}_{k=1}^{N_p}$ を算出.
 - 6: 勾配 $\left\{\frac{\partial}{\partial\theta_k}\mathcal{L}_{\text{cllocal}}\right\}_{k=1}^{N_p}$ と学習率 ε を用いて $\{\theta_k\}_{k=1}^{N_p}$ を更新する.
 - 7: **until** 収束
-

6.3 数値実験：損失関数とその勾配

本節では, 提案 OTL(6.10) とその勾配ベクトルについて解析する. まず, 6.3.1 項にて, OTL の推定誤差に関しデータの intrinsic dimension と qubit 数の観点から数値実験を通して調べる. 特に, 提案 OTL が distance ではなく divergence であっても, 定理 6.1 や系 6.1, 系 6.2 と同様の性質を示すかどうかを確認する. 次に, 6.3.2 節にて, OTL の推定誤差に関し測定数 (shot noise) の観点から数値実験と理論解析を通して調べる. 最後に, 6.3.3 節にて, 提案 OTL が勾配消失を回避できることを数値実験を通して確認する. すなわち, 損失関数の局所性により, 勾配が qubit 数に対して指数関数的に消失しないことを確認する. これら全ての解析は, 学習過程におけるある瞬間における特性に着目して行う. 学習過程や学習結果の性能の確認は次節で行う.

ここでは, パラメタ付きユニタリ演算子 $U(\mathbf{z}, \theta)$ (図 6.1) を用いて implicit generative model (6.9) を構成する. [92] と似ているが, 本研究のモデルには潜在変数 \mathbf{z} が含まれる点が異なる. 具体的には, 以下のように, 前述のユニタリ演算子を N_L 回繰り返したモデルを用いる (ここでは第 ℓ 番目のユニタリ演算子を第 ℓ 層と呼ぶ):

$$U_{N_L,\xi,\eta}(\mathbf{z}, \theta) = \prod_{\ell=1}^{N_L} W V_{\xi_\ell,\eta_\ell}(\mathbf{z}, \theta_\ell),$$

但し, $\theta_\ell = \{\theta_{\ell,j}\}_{j=1}^n$, $\xi_\ell = \{\xi_{\ell,j}\}_{j=1}^n$, $\eta_\ell = \{\eta_{\ell,j}\}_{j=1}^n$ は第 ℓ 層目の n 次元のパラメタベクトルを表す. 簡単のため, これらのベクトルは, $\theta = \{\theta_\ell\}_{\ell=1}^{N_L}$, $\xi = \{\xi_\ell\}_{\ell=1}^{N_L}$, $\eta = \{\eta_\ell\}_{\ell=1}^{N_L}$ と表すことにする. 但し, θ は調整パラメタ, \mathbf{z} は潜在変数である. W は controlled- Z gate で構成される ladder 型の entangling unitary gate である. すなわち, W は隣接 qubit に作用する two-qubit controlled- Z gate である:

$$W = \prod_{i=1}^{n-1} CZ_{i,i+1},$$

但し, $CZ_{i,i+1}$ は第 i 番目と第 $(i+1)$ 番目の qubit に作用する controlled- Z gate であり, 演算子

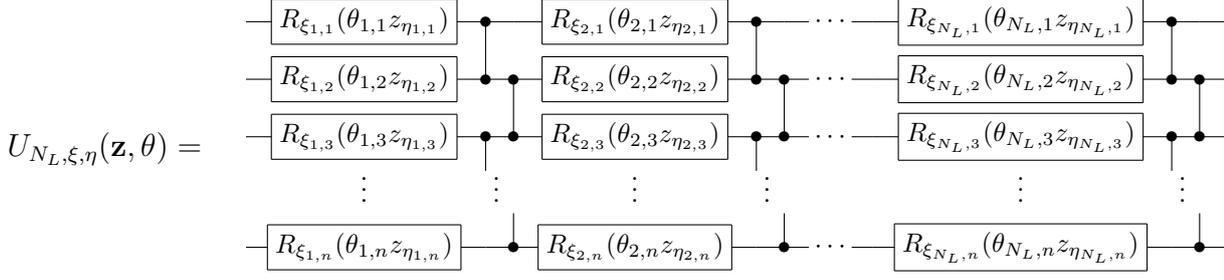


図 6.1: 6.3 節と 6.4 節で用いるパラメタ付き量子回路 (ansatz) . この ansatz は同じ構造をもつ層を繰り返して構成される. 第 ℓ 層目には, 回転角 $\{\theta_{\ell,j} \cdot z_{\eta_{\ell,j}}\}_{j=1}^n$ で回転方向 $\{\xi_{\ell,j}\}_{j=1}^n$ の single-qubit 回転演算子が各 qubit に作用された後, ladder 構造の controlled-Z gate が作用される. 回転方向 $\xi = \{\xi_{\ell,j}\}_{\ell,j=1}^{N_L,n}$ と潜在変数の index パラメタ $\eta = \{\eta_{\ell,j}\}_{\ell,j=1}^{N_L,n}$ は学習過程の前にランダムに選択され, 学習過程では固定される. Reprinted figure from [46]. Reproduced with permission from Springer Nature.

$V_{\xi_\ell, \eta_\ell}(\mathbf{z}, \theta_\ell)$ は single-qubit gate で構成される:

$$V_{\xi_\ell, \eta_\ell}(\mathbf{z}, \theta_\ell) = \prod_{i=1}^n R_{\xi_{\ell,i}}(\theta_{\ell,i} z_{\eta_{\ell,i}}),$$

但し, $R_{\xi_{\ell,i}}(\theta_{\ell,i} z_{\eta_{\ell,i}})$ は, 第 ℓ 層目の回転角 $\theta_{\ell,i} z_{\eta_{\ell,i}}$ の single-qubit 回転演算子であり, 回転方向は $\xi_{\ell,i} \in \{X, Y, Z\}$ である. 例えば, $R_X(\theta_{\ell,3} z_5) = \exp(-i\theta_{\ell,3} z_5 \sigma_x)$ のような具合である. index parameter $\eta_{\ell,i} \in \{0, 1, \dots, N_z\}$ と $\xi_{\ell,i} \in \{X, Y, Z\}$ は学習過程の初めにランダムに選択され, 学習過程の中は固定される. また, ansatz $|\phi_\theta(\mathbf{z})\rangle = U(\mathbf{z}, \theta)|0\rangle^{\otimes n}$ の表現力を高めるために, $z_0 = 1$ として回転角のバイアスを導入した. (例えば, もしバイアス項がない状態で $z_1, \dots, z_{N_z} \sim 0$ となると, $|\phi_\theta(\mathbf{z})\rangle$ は $|0\rangle^{\otimes n}$ の周辺に限定されてしまう.) 全ての量子状態シミュレーションには Qiskit [87] を用いた.

6.3.1 データ数による近似誤差

6.1.2 節では, p-Wasserstein distance において, データ数の観点からの近似誤差の収束則は $\mathcal{O}(M^{-1/N_x})$ ではなく, $\mathcal{O}(M^{-1/s})$ であることを見た. 但し, M はデータ数, N_x はサンプル空間 \mathcal{X} の次元, s はデータの intrinsic dimension または潜在空間の次元数である. 一般に $s \ll N_x$ であるため, これは好ましい特徴である. ところが, 提案 OTL (6.10) は distance ではなく divergence であるため, 同様のスケール則が当てはめられるかは理論的には非自明である. しかし, 本項では数値実験を通して, この conjecture をサポートするポジティブな結果を示していく.

本項では, 以下の 2 種類の数値シミュレーションを行う. 1 つ目の実験 (実験 A) の目的は, 提案 OTL が式 (6.4) と同様の性質を持つかどうかを調べることである. すなわち, 同一の分布からサンプルされた二つの異なる経験分布の差を調べることになる. 2 つ目の実験 (実験 B) の目的は, 式 (6.5) と同様の性質を持つかを調べることである. すなわち, 無限個のサンプル数を仮定した際に得られる理想的な OTL と経験分布から算出される OTL の差を調べることになる. 本実験では, 無限回の測定回数に相当する statevector simulator を用いて OTL の算出を行い, 次項で有限の測定回数による影響について調べる. 全ての数値実験ではパラメタ $\xi, \eta, \theta, \tilde{\xi}, \tilde{\eta}, \tilde{\theta}$ をランダムに選択し, 実験中は値を固定して行った.

表 6.1: 6.3.1 項の数値シミュレーションにおけるパラメタ設定

測定回数	N_s	∞ (statevector)
学習データ数	M	$\{2^i i \in \{0, 1, 2, \dots, 10\}\}$
試行回数	N_{Monte}	100
qubit 数	n	$\{1, 2, 4, 6, 8, 10\}$: 実験 A $\{1, 2, 4, 6, 8\}$: 実験 B
潜在空間の次元数	N_z	$\{1, 2, 4, 6, 8\}$: 実験 A $\{1, 2, 4, 6, 10, 14\}$: 実験 B
層数	N_L	$3 + \lfloor N_z/n \rfloor$

実験 A 式 (6.4) を参考にして, 式 (6.10) で定義される経験的 OTL の期待値に着目する:

$$\mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}} \left[J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) \right], \quad (6.12)$$

但し,

$$J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) = \mathcal{L}_{\text{Cllocal}} \left(\{U_{N_L, \tilde{\xi}, \tilde{\eta}}(\tilde{\mathbf{z}}_i, \tilde{\theta}) | 0\rangle^{\otimes n}\}_{i=1}^M, \{U_{N_L, \xi, \eta}(\mathbf{z}_j, \theta) | 0\rangle^{\otimes n}\}_{j=1}^M \right). \quad (6.13)$$

式 (6.12) では, 式 (6.13) の右辺 $\mathcal{L}_{\text{Cllocal}}$ の引数に現れる二つのユニタリ演算子として $\xi = \tilde{\xi}$ と $\eta = \tilde{\eta}$ を考える. これは式 (6.12) 中の $J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M)$ が, 学習データ数が無限大の極限 ($M \rightarrow \infty$) でゼロになることを意味する.

式 (6.12) 中の期待値は, 潜在変数 $\tilde{\mathbf{z}}_i$ に対して取る. \mathbf{z}_j は N_z 次元の一様分布 $U(0, 1)^{N_z}$ からサンプルされるが, 試行回数 N_{Monte} は有限であり, モンテカルロサンプリングにより近似した. 他の条件は表 6.1 の通りである.

図 6.2 に, 式 (6.12) において様々な N_z (潜在変数の次元), n (qubit 数) を用いた値をプロットした. 図中の破線は M^{-1/N_z} に対応する曲線である. 注目すべきは, qubit 数に関係なく, 学習データ数が大きな領域ではプロットが破線に一致する点である. これは提案 OTL が, 式 (6.12) で与えられる異なる二つのアンサンブルにおいて, qubit 数 n にはほぼ依存せず, 主に潜在次元 N_z に依存していることを示している. これは系 6.1 で distance-based な損失関数に対して示されたのと同様の性質である.

実験 B 次に提案 OTL の近似誤差が, 式 (6.5) と同様のスケールング則を満たすかを調べる. 具体的には, 学習データ数 M に対する以下の期待値の依存性を数値的に確認する:

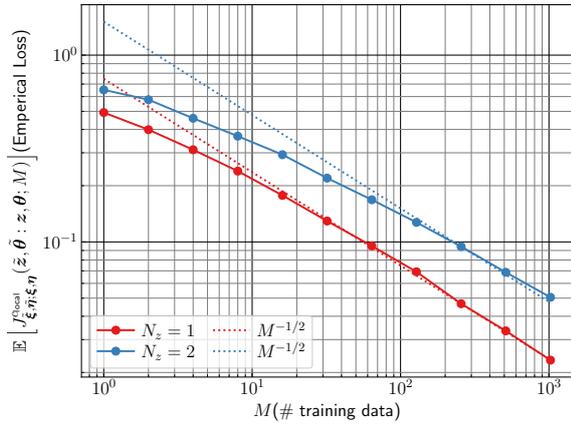
$$\mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}} \left[\lim_{K \rightarrow \infty} J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; K) - J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) \right], \quad (6.14)$$

但し, J^{Cllocal} は式 (6.13) で定義される量である. この場合, 式 (6.13) 中の二つのユニタリ演算子として, 異なるパラメタ $\xi \neq \tilde{\xi}$ と $\eta \neq \tilde{\eta}$ を考える. 数値実験で用いた各種パラメタは表 6.1 に記した.

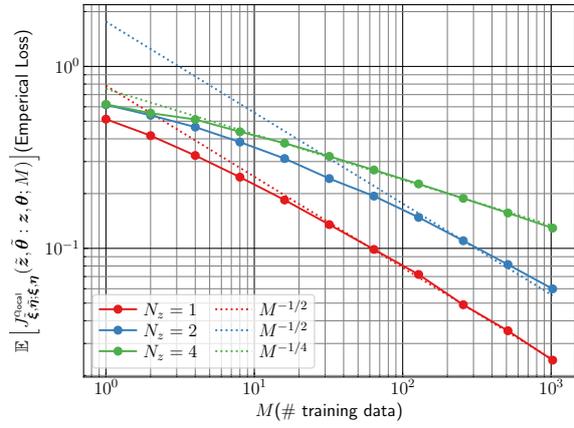
式 (6.14) の第一項目は, 無限個のデータ数を用いて推定した場合の理想的な量である. 第二項目は, 式 (6.5) で示される期待値であり, 以下のように表される:

$$\mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}} \left[J_{\xi, \eta; \xi, \eta}^{\text{Cllocal}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) \right] = aM^{-1/b} + c. \quad (6.15)$$

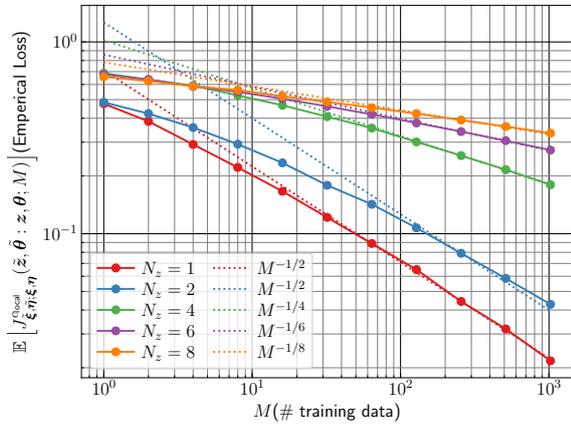
パラメタ b を同定するために, 式 (6.15) の左辺をデータ数 M の関数としてモンテカルロ法により計算し, 関数 $aM^{-1/b} + c$ によりフィッティングした. 同様のことを, qubit 数 n や潜在次元 N_z 毎に繰り返した. 詳細な実験結果は付録 B を参照されたい. パラメタ同定の結果は図 6.3 に示した. すなわ



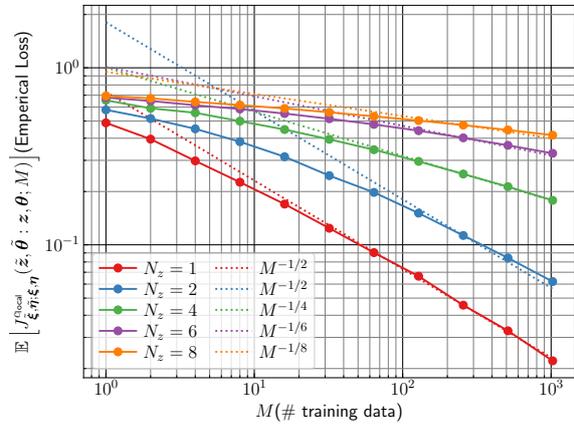
(a) $n = 1$



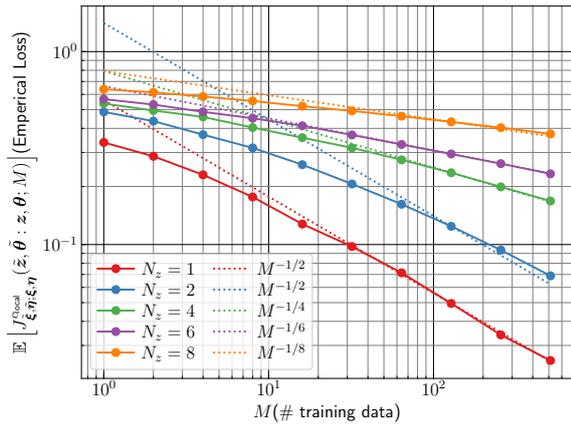
(b) $n = 2$



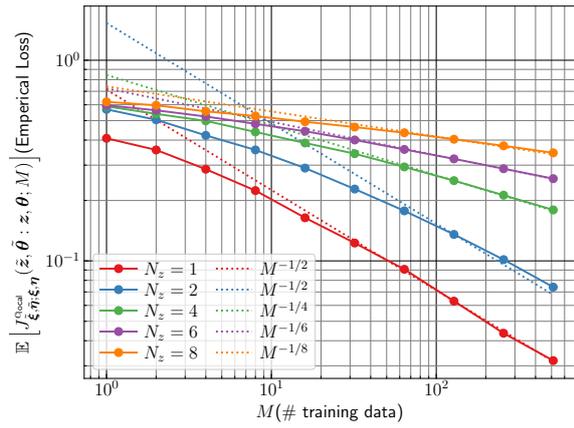
(c) $n = 4$



(d) $n = 6$



(e) $n = 8$



(f) $n = 10$

図 6.2: 学習データ数と経験的 OTL の関係 (数値シミュレーション結果) . 破線は M^{-1/N_z} に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.

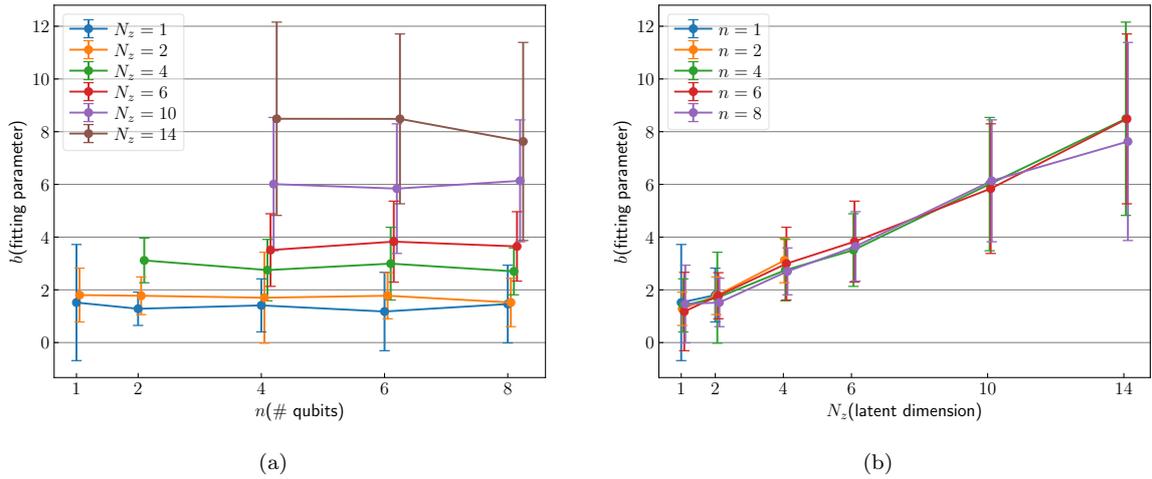


図 6.3: フィッティングパラメタ b のシミュレーション結果. (a) qubit 数依存性, (b) 潜在次元依存性. Reprinted figure from [46]. Reproduced with permission from Springer Nature.

表 6.2: 6.3.2 項の数値シミュレーションにおけるパラメタ設定

測定回数	N_s	$\{2^{i+7} \mid i \in \{0, 1, 2, \dots, 7\}\}, \infty(\text{statevector})$
学習データ数	M	$\{2^i \mid i \in \{0, 1, 2, \dots, 10\}\}$
試行回数	N_{Monte}	256
qubit 数	n	8
潜在空間の次元	N_z	$\{1, 2, 4\}$
層数	N_L	$3 + \lfloor N_z/n \rfloor$

ち、フィッティングパラメタ b は qubit 数 n にはほぼ依存せず、潜在次元 N_z に対して線形にスケールする。この結果は確かに式 (6.5) に一致する。

実験 A, B の結果が示すことは、以下の conjecture である：

Conjecture 6.1 学習データ数 M に対する OTL(6.10) の近似誤差は、潜在次元 N_z について以下のようにスケールする：

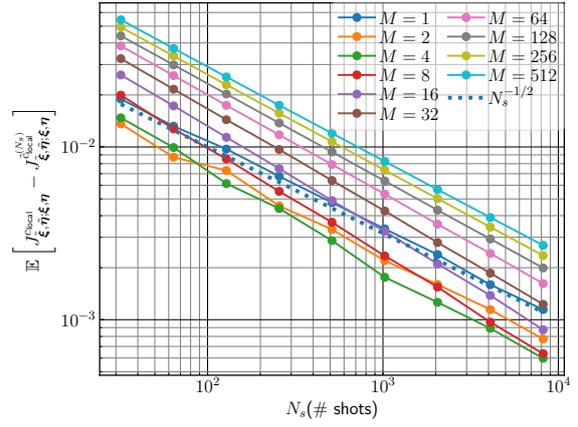
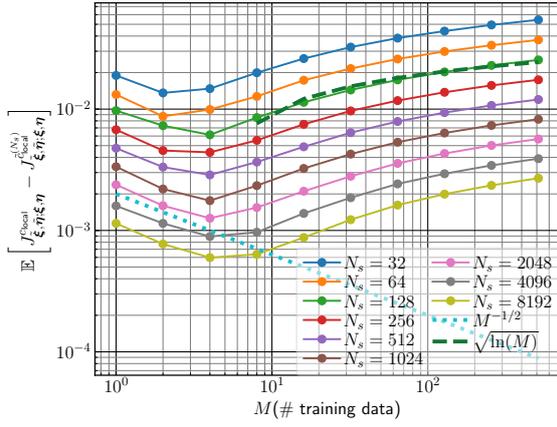
$$\mathbb{E}_{\bar{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}} \left[\mathcal{L}_{\text{clonal}} \left(\{U(\bar{\mathbf{z}}_i, \theta) |0\rangle^{\otimes n}\}_{i=1}^{\infty}, \{U(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}\}_{j=1}^M \right) \right] \lesssim O(M^{-1/N_z}),$$

すなわち、qubit 数には依存しない。

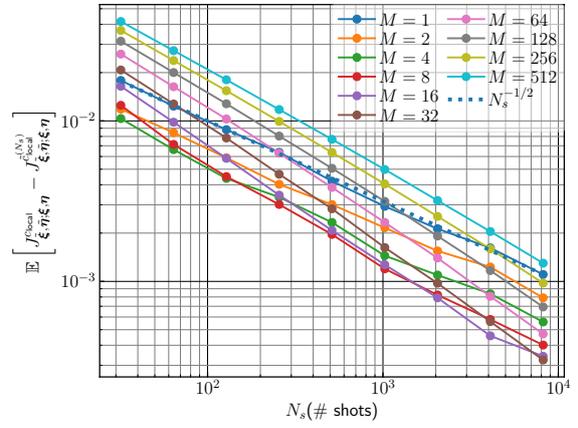
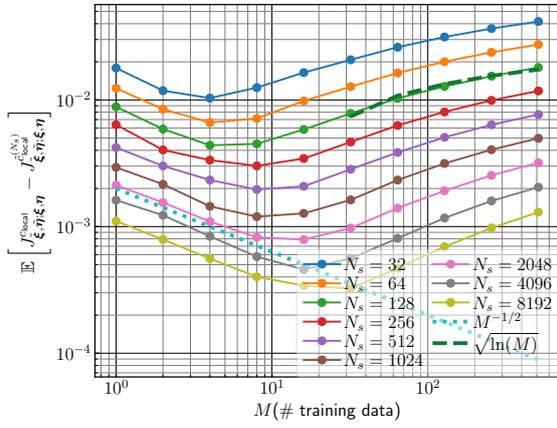
以上の結果は、データの intrinsic dimension や潜在次元が十分に小さい場合は、サンプリングによって効率的に提案 OTL を計算できることを意味している。これを理論的に証明することは簡単ではないため、ここでは今後の課題としておく。

6.3.2 Shot 数による近似誤差

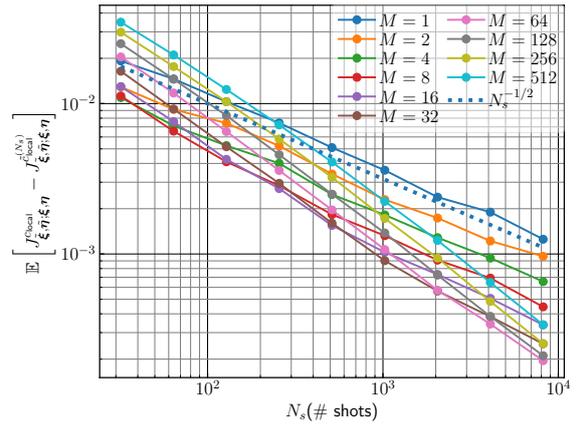
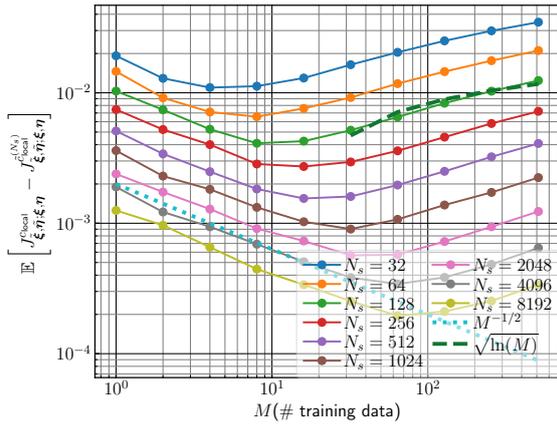
6.3.1 節における誤差解析では、測定回数 (shot 数) を無限大と仮定しており、量子状態間の ground cost が完全に同定できるものとした。ここでは shot 数が有限であることが近似誤差に与える影響について解析する。解析の起点となるのは、以下の命題である。



(a) $N_z = 1$



(b) $N_z = 2$



(c) $N_z = 4$

図 6.4: 測定回数と経験的 OTL の関係 (数値シミュレーション結果). 左列図中の点線は $M^{-1/2}$ に従う関係を表し, 破線は $N_s = 128$ の結果を $\sqrt{c_1 \ln(M) + c_2}$ でフィッティングした結果を表す. 右列図中の点線は $N_s^{-1/2}$ に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.

表 6.3: 6.3.3 項の数値シミュレーションにおけるパラメタ設定

測定回数	N_s	∞ (statevector)
学習データ数	M	$\{2^i \mid i \in \{1, 2, 3, 4\}\}$
試行回数	N_{Monte}	300
qubit 数	n	$\{2, 4, 6, 8, 10, 12, 14\}$
潜在空間の次元数	N_z	$\{1\}$
層数	N_L	$\{10, 25, 50, 75, 100, 200\}$

命題 6.2 今, $\tilde{c}_{\text{local}}^{(N_s)}$ を式 (6.8) における ground cost c_{local} を, N_s サンプルで推定した値とする. ここで異なる二つの確率分布のサポートが厳密に分離されているとする場合, 任意の $i, j \in \{1, 2, \dots, M\}$ に対する ground cost について $g > 0$ なる下界が存在する. すなわち, $(c_{\text{local}}(|\psi_i\rangle, U(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}) > g, \forall i, j)$ とする. この時, 任意の正の値 δ について, 以下の不等式が成立する:

$$P\left(\left|\mathcal{L}_{c_{\text{local}}}\left(\{|\psi\rangle_i\}_{i=1}^M, \{U(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}\}_{j=1}^M\right) - \mathcal{L}_{\tilde{c}_{\text{local}}^{(N_s)}}\left(\{|\psi\rangle_i\}_{i=1}^M, \{U(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}\}_{j=1}^M\right)\right| \geq \sqrt{\frac{2M}{\delta}} \sqrt{\frac{1-g}{N_s} + \frac{(1-g)^2}{4N_s^2 g} + \frac{1-g}{2N_s \sqrt{g}}}\right) \leq \delta.$$

この証明は [167] の Appendix C を参照されたい. 命題 6.2 は, 提案 OTL の近似誤差は, $M \gg 1$ かつ $N_s \gg 1$ の条件の下で, $O(\sqrt{M/N_s})$ の定数が上界であることを示している. 従って, Conjecture 6.1 が正しい場合, shot 数 N_s とデータ数 M が有限であることによる近似誤差の上界は $O(M^{-1/N_s}) + O(\sqrt{M/N_s})$ である. 但し, N_z は潜在次元である.

次に, データ数 M に加えて shot 数 N_s の関数として, 平均的な近似誤差について数値シミュレーションの結果を示す. 設定としては, 6.3.1 節の図 6.1 に記した hardware efficient ansatz を量子回路として用いた. この数値シミュレーションの目的は, 以下の期待値のデータ数 M と shot 数 N_s に対する依存性を確認することである.

$$\mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}} \left[\left| J_{\tilde{\xi}, \tilde{\eta}; \xi, \eta}^{\tilde{c}_{\text{local}}^{(N_s)}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) - J_{\tilde{\xi}, \tilde{\eta}; \xi, \eta}^{c_{\text{local}}}(\tilde{\mathbf{z}}, \tilde{\theta} : \mathbf{z}, \theta; M) \right| \right],$$

但し, $J_{\tilde{c}_{\text{local}}^{(N_s)}}$ は式 (6.13) により計算される値である. 6.3.1 節の数値シミュレーションで示したように, 期待値 $\mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{z} \sim U(0,1)^{N_z}}[\cdot]$ は一様分布 $U(0,1)^{N_z}$ から潜在変数 $\tilde{\mathbf{z}}, \mathbf{z}$ をサンプル数してこすることで, モンテカルロサンプリングにより近似できる. また, パラメタ $\xi, \eta, \theta, \tilde{\xi}, \tilde{\eta}, \tilde{\theta}$ はシミュレーション前にランダムに設定し, シミュレーション中は固定される. その他のパラメタは表 6.2 の通りである. 図 6.4 にシミュレーション結果を示した. 着目すべき点は以下の通りである.

- データ数 M が小さい領域では, 近似誤差がおおよそ $M^{-1/2}$ に比例する.
- データ数 M が大きい領域では, 近似誤差が $\sqrt{c_1 \ln M + c_2}$ で表される. 但し, c_1, c_2 は定数.
- 学習誤差は, shot 数 N_s に対しておおよそ $N_s^{-1/2}$ に比例する.

これらの結果の直感的な説明を付録 C に記す. 特に重要なことは, 近似誤差を抑えるには適切なサンプル数を選択する必要があるという点である.

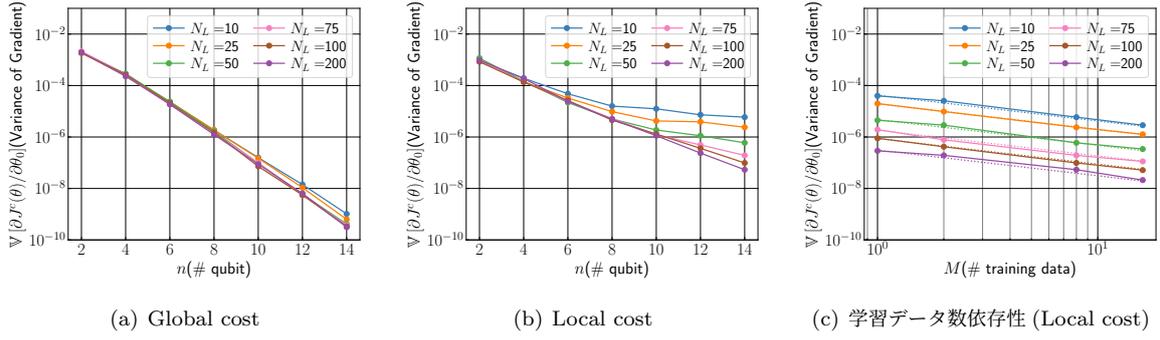


図 6.5: 損失関数の勾配と qubit 数の関係 (数値シミュレーション) . 図中の点線は M^{-1} に従う関係を表す. Reprinted figure from [46]. Reproduced with permission from Springer Nature.

6.3.3 勾配消失の回避

6.2.1 項では, 勾配消失のための必要条件となる local measurement による損失関数を選択した. 但し, local cost を選択するだけでは勾配消失を完全には回避できないことには注意が必要である. 例えば [94] では, local cost に加えて *alternating layered ansatz (ALA)* なるタイプのパラメタ付き量子回路を用いる方法が提案されており, この場合は勾配消失が回避できることが示されている. しかしながら, そのような特殊な条件を導入しなくとも, 提案手法により global cost に比べて確かに勾配の減少が緩和されることを, 数値実験を用いて示す.

より具体的には, 提案 OTL(6.10) の偏微分の分散の期待値を, 学習データのアンサンブル $\{|\psi_i\rangle\}_{i=1}^M$ と生成モデルからサンプルされたデータのアンサンブル $\{U(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}\}_{j=1}^M$ に基づいて計算する.

$$\mathbb{V}_{\xi, \eta, \theta, \mathbf{z}} \left[\frac{\partial}{\partial \theta} \mathcal{L}_{\text{local}} \left(\{|\psi_i\rangle\}, \{U_{N_L, \xi, \eta}(\mathbf{z}_j, \theta) |0\rangle^{\otimes n}\} \right) \right]. \quad (6.16)$$

偏微分は parameter shift rule [105] により計算する. 分散の期待値は, $\mathbf{z}, \xi, \eta, \theta$ に対するモンテカルロ計算により近似的に算出する. 但し, \mathbf{z} は一様分布 $U(0, 1)$ からサンプルされ, ξ, η, θ は $\xi \in \{X, Y, Z\}^{nN_L}$, $\eta \in \{0, 1, \dots, N_z\}^{nN_L}$ からサンプルされ, $\theta \in [0, 2\pi]^{nN_L}$ とする. 生成モデル $U(\mathbf{z}, \theta) |0\rangle^{\otimes n}$ の構造は, 図 6.1 と同じものを用い, 微分は $\theta_{1,1}$ について取るものとする. 学習データのアンサンブル $\{|\psi_i\rangle\}_{i=1}^M$ は以下のように準備する:

$$|\psi\rangle_i = W'V_2'(\zeta_2^i)W'V_1'(\zeta_1^i)|0\rangle^{\otimes n},$$

但し, $\zeta_1^i = \{\zeta_{1,j}^i\}_{j=1}^n$ と $\zeta_2^i = \{\zeta_{2,j}^i\}_{j=1}^n$ は一様分布 $[0, 2\pi]$ からランダムに選択され, モンテカルロ計算の間は固定される. 演算子 W', V_1', V_2' は以下のように定義される:

$$\begin{aligned} W' &= \prod_{i=1}^{n-1} CX_{j,j+1}, \\ V_1'(\zeta_1^i) &= \prod_{j=1}^n R_{j,Y}(\zeta_{1,j}^i), \\ V_2'(\zeta_2^i) &= \prod_{j=1}^n R_{j,Z}(\zeta_{2,j}^i), \end{aligned}$$

但し, $CX_{j,k}$ は controlled- X gate であり, 第 j 番目の qubit を制御 qubit とし第 k 番目の qubit に対して X gate を作用させる. $R_{j,Y}, R_{j,Z}$ は, それぞれ x 軸, y 軸周りの single qubit Pauli 回転を表す. その他のパラメータは, 表 6.3 に示す.

図 6.5 に勾配の微分の期待値の分散 (6.16) の計算結果を示した. それぞれ, (a) global cost (6.7) を用いた場合, (b) local cost (6.8) を用いた場合である. 学習データ数は $M = 8$ とした. global cost を用いた場合は, 層数 N_L に依らず, 勾配の分散が指数関数的に明確に消失していることが見て取れる. それとは対照的に local cost を用いた場合は, 比較的浅い回路 ($N_L = 10, 25$) では $n \geq 10$ の領域でほぼ一定になっている. $N_L \geq 50$ のような深い回路でも, global cost の場合と比べると勾配消失の度合いは緩やかであり, $n \geq 8$ の領域でも大きな値を取っている. この結果は, local な ground cost を用いた提案 OTL が, この目的のために特別な回路を導入しなくとも勾配消失の問題を回避できることを示している. 尚, これは本研究において単一の ground cost により損失関数を構成した場合について調べた [121] の結果とも合致している.

図 6.5(c) には, 分散 (6.16) を学習データ数 M の関数としてプロットした ($n = 14$). この図ではモンテカルロ計算の結果を点で, スケーリング曲線として M^{-1} を点線で記した. このフィッティング結果は, 勾配が M に対して緩やかなスケーリングに従っており, 学習データ数が大きな領域でも効率的に学習が可能であることを示している.

6.4 デモンストレーション：量子異常検知

6.4.1 量子異常検知タスク

異常検知は, 与えられたテストデータ $\mathbf{x}^{(t)}$ が通常とは異なるデータ (anomalous, rare) かどうかを, 過去の学習データ $\mathbf{x}_i, (i = 1, 2, \dots, M)$ に基づいて判別するタスクである. 通常のカテゴリ分類タスクと異なる点は, この問題では通常データ (normal data) の数と異常データ (anomalous data) の数が大きく不均衡な点である (前者は後者に比べて非常に多い). 従って, 典型的な分類手法を用いてもこのタスクを上手く解くことができず, いくつかの特化した手法が開発されている [186].

異常検知問題は, 量子テクノロジーの文脈では特に重要である. すなわち, 正確な状態準備や制御を実現するためには, 余計な量子状態を検知し, できるだけ早く取り除く必要がある. 既存の異常検知手法は, 測定に基づいたデータ処理 [41, 42] を行っており, 特に量子トモグラフィを用いているために多くの測定回数を必要とする. 一方, 本研究で提案する異常検知の手法は, 構築した生成モデルに対して量子状態を直接入力するため, 非常に少ない測定回数で異常の診断が行える.

まず以下に, 生成モデルを用いた (古典データに対する) 異常検知の既存手法 [187] を記す. 本研究では, これを量子データに応用する.

1. (分布の推定): 正常データセットの確率分布モデルを構築する.
2. (異常値の設計): Anomaly Score (AS; 異常値) を正常分布モデルに基づいて定義する.
3. (閾値の設定): 異常と判定するための AS の閾値を設定する.

ここでステップ 1 における確率分布モデルの構築に, 6.2.2 項に記した学習アルゴリズムを適用する. ステップ 2 における AS の設定には, 古典機械学習の AnoGAN [188] を参考にする. すなわち, 損失関数として

$$\mathcal{L} \left(U(\mathbf{z}, \theta) |0\rangle^{\otimes n}, |\psi^{(t)}\rangle \right)$$

をテストデータ $|\psi^{(t)}\rangle$ に対して定義し, 生成モデル $U(\mathbf{z}, \bar{\theta})$ は学習パラメータ $\bar{\theta}$ を備え, 学習データを

用いて構築される。そこで、AS が最小の値を取るような潜在変数 \mathbf{z} を求める：

$$(\text{Anomaly Score}) = \min_{\mathbf{z}} \mathcal{L} \left(U(\mathbf{z}, \bar{\theta}) |0\rangle^{\otimes n}, |\psi^{(t)}\rangle \right).$$

損失関数 \mathcal{L} と同様に、ここでも式 (6.8) にある local ground cost $c_{\text{local}}(|\psi^{(t)}\rangle, U(\mathbf{z}, \bar{\theta}) |0\rangle^{\otimes n})$ を用いる。上記の最小化は、潜在変数 \mathbf{z} に対して最急降下法を用いて実行することができるが、 θ に対する微分の計算と同様にここでも parameter shift rule を適用することができる。Algorithm 2 の要約を記す：

Algorithm 2 Anomaly Score を計算するアルゴリズム

Input: 学習する量子回路 $U(\mathbf{z}, \bar{\theta})$, テストデータ $\{|\psi^{(t)}\rangle\}$

Output: Anomaly Score

- 1: \mathbf{z} を初期化
 - 2: **repeat**
 - 3: 式 (6.8) に従って、 $\{U(\mathbf{z}, \bar{\theta})\}$ と $|\psi^{(t)}\rangle$ から ground cost \mathcal{L} を計算する。
 - 4: parameter shift rule により、勾配 $\left\{ \frac{\partial \mathcal{L}}{\partial z_k} \right\}_{k=1}^{N_z}$ を計算する。
 - 5: $\left\{ \frac{\partial \mathcal{L}}{\partial z_k} \right\}_{k=1}^{N_z}$ に基づく最急降下法により \mathbf{z} を更新する。
 - 6: **until** 収束
-

6.4.2 分散したデータの場合

1 つ目のデモンストレーションでは、ブロッホ球上の赤道に分布する量子アンサンブルを学習する生成モデルを構成する。すなわち、学習アンサンブル（正常データ） $\{|\psi_j\rangle\}_{j=1}^M$ として以下のようなものを考える：

$$|\psi_j\rangle = \cos(\pi/4) |0\rangle + e^{2\pi i \phi^j} \sin(\pi/4) |2^n - 1\rangle,$$

但し、 ϕ^j は $[0, 1]$ の一様分布から生成され、 $|x\rangle$ は 2^n 次元の Hilbert 空間における第 x 番目の基底を表す。注意すべきは、このアンサンブルは mixed-state-based の既存の量子異常検知手法 [41, 42] では使えない点である。すなわち、この状態に対応する混合状態は、ほぼ完全混合状態になるために、既存手法では元のアンサンブルを復元することができないのである。

Ansatz としては 6.3.3 節と同じものを用い、パラメタは表 6.4 の設定とした。生成モデルは Algorithm 1 に従って構築した。一旦、正常アンサンブルに対応するモデルが構築できれば、その後は異常検知に用いることができる。テストデータ $\{|\psi^{(t)}\rangle\}$ は

$$|\psi^{(t)}\rangle = \cos\left(\frac{\pi}{2}\theta^{(t)}\right) |0\rangle + e^{2\pi i \phi^{(t)}} \sin\left(\frac{\pi}{2}\theta^{(t)}\right) |2^n - 1\rangle, \quad (6.17)$$

で与えた。但し、 $\theta^{(t)}, \phi^{(t)} \in \{0, 0.1, 0.2, \dots, 2\}$ である。AS は Algorithm 2 に従って算出した。その他のパラメタ設定は表 6.4 の通りである。

図 6.6, 6.7 に、それぞれ $n = 2, n = 10$ の結果を示した。学習アンサンブルの様子を (a) に、generalized Bloch vector (2^n 次元の Hilbert 空間を 2 次元空間へ射影した様子) として青い点でプロットした。合わせて、生成モデルから出力された量子状態 ($z \in [0, 1]$ に対応するもの) を (b) にプロットした。(c) の青と赤のプロットはテストデータの量子状態 (6.17) に対応している。(d) は算出された AS を示しており、青や赤の色が (c) の色と対応している。(d) 中の破線は、学習データを完全に学習できた場合の理論曲線である。

表 6.4: 6.4.2 項の数値シミュレーションにおけるパラメタ設定

測定回数 (学習時)	N_s	100 ($n = 2$), ∞ ($n = 10$)
測定回数 (テスト時)	N_{ad}	50 ($n = 2$), 100 ($n = 10$)
学習データ数	M	30
qubit 数	n	2, 10
潜在次元数	N_z	1
層数	N_L	10

まず, 図 6.6 において, 算出された AS と理論曲線が非常によく一致していることが分かる. これは提案手法により, 適切に AS が計算できていることを示している. 実用的な場面では, ユーザーが所望のタスクに応じて閾値を設定し, その値と算出された AS を比較することで与えられた量子状態が異常かどうかを判別する. ここでは例として, 閾値を $AS = 0.3$ した. この値を用いると, 図 6.6 においては, $0.3 \leq \theta^{(t)}/\pi \leq 0.7$ の領域が正常, それ以外が異常と判別される. 次に, 図 6.7 で示される $n = 10$ の結果を見てみる. (b) の様子からは完全な学習には失敗しているように見えるが, (a) に示されている学習データと見比べると, ある程度の相関は見られる. すなわち, モデルの出力状態は $|0\rangle^{\otimes 10}$ と $|1\rangle^{\otimes 10}$ で張られる generalized Bloch sphere における xy 平面に限定されており, Hilbert 空間の次元が 2^{10} 次元あることを考えると, それなりに適切に学習が進んでいる. また, (d) にある AS の様子を見ると, (b), (c) の状態の差に応じて AS 自体は適切に算出ができていていることが分かる. ここで着目すべきは, $n = 10$ であっても, わずか $N_s = 100$ で異常検知が実行できている点である. これは提案手法が qubit 数に応じてスケールアップできることを示しており, 実用面では重要な利点である.

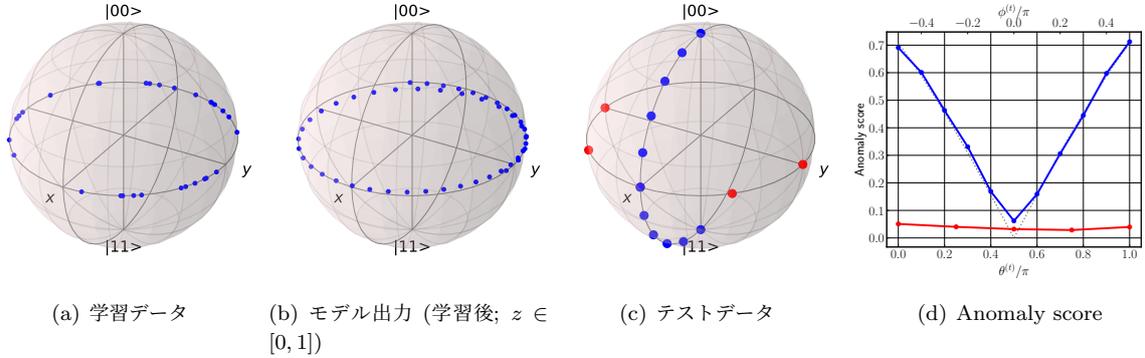


図 6.6: 分散データ, $n = 2$ における異常検知の結果 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.

6.4.3 局在したデータの場合

次に局在した量子アンサンブルの場合について見てみる. ここでは, 正常状態に対応する量子アンサンブル $\{|\psi_j\rangle\}_{j=1}^M$ として

$$|\psi_j\rangle = \cos\left(\frac{\pi}{2}\Delta\theta_j\right)|0\rangle + e^{2\pi i\Delta\phi_j}\sin\left(\frac{\pi}{2}\Delta\theta_j\right)|2^n - 1\rangle,$$

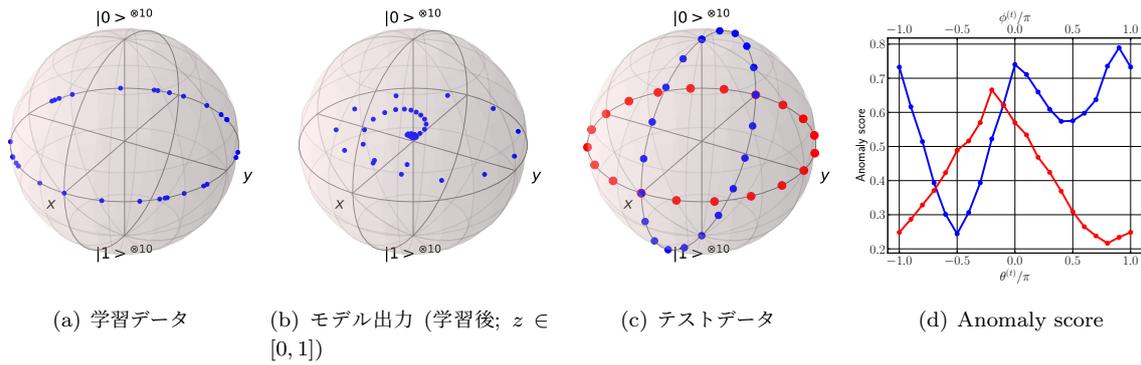
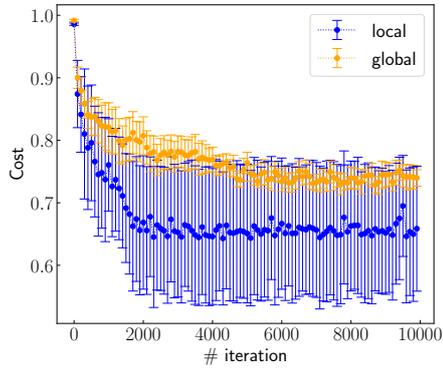


図 6.7: 分散データ, $n = 10$ における異常検知の結果 (数値シミュレーション). Reprinted figure from [46]. Reproduced with permission from Springer Nature.

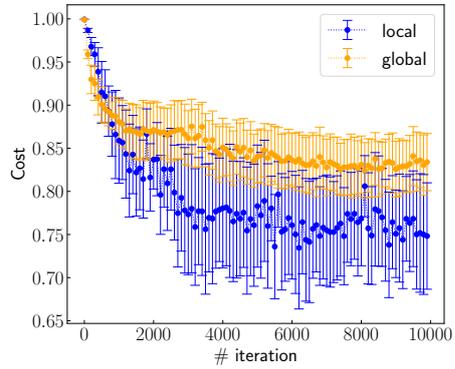
を考える. 但し, n は qubit 数であり, $\Delta\theta_j, \Delta\phi_j$ は, それぞれ正規分布 $N(\mu, \sigma)$ (μ は平均, σ は分散) と一様分布 $U(a, b)$ に対応するものとする. 本項では次の二つのケース; $n = 6$ における $(\mu, \sigma, a, b) = (0, 0.02, 0, 0.1)$, $n = 10$ における $(\mu, \sigma, a, b) = (0, 0.02, 0, 0.2)$ について見る. これら 2 つの設定は, 図 6.9 (a, d) にあるように, generalized Bloch sphere でほぼ 2 次元に分布していることに注意されたい. 他の設定は表 6.5 に記した.

この 2 次元分布を学習する生成モデルを構築するために, 潜在変数の次元は $N_z = 2$ とした. また, 本項では勾配消失の観点を検証するために, 以下の二つのケースを比べた. 1 つ目が, local cost に加えていわゆる alternating layered ansatz (ALA) を用いたものであり, 勾配消失を回避できる [94, 189] ことが知られている. 2 つが, global cost かつ hardware efficient ansatz (HEA) と呼ばれているものである. これらを学習速度の観点から比較した. 図 6.8 に典型的な学習曲線を示した. “local” とラベルが付いた青いプロットは, ALA かつ local cost (6.8) による結果であり, 以下では L-ALA と呼ぶ. 一方の “global” とラベルが付いたオレンジのプロットは, HEA かつ global cost (6.7) の結果であり, 以下では G-HEA と呼ぶ. 比較の際に注意が必要なのが, 両コストの定義の違いである. 定義が異なるコストを比較することはできないため, 図中のコストは, 各 iteration における量子状態に対応するパラメータを用いて算出した global cost で統一してある. すなわち, 図中にプロットした値は直接比較して差し支えない. すると, L-ALA が G-HEA に比べて明らかに収束速度が速い様子が見て取れる. この結果は 6.3.3 節の議論の内容と合致しており, local cost の優位性が確認できる. さらに, 収束先の値も G-HEA に比べて L-ALA の方が小さい. 尚, 学習性能は初期ランダムシードに大きく依存するため, G-HEA については学習が完全に成功する設定を見付けることができなかった. すなわち, 全ての学習過程が局所最適解にトラップされているように見えた. このような減少も G-HEA の勾配の分散が L-ALA に比べて小さいことで説明がつく.

次に, 得られた生成モデルを用いて異常検知タスクを行った. 図 6.9 の (b), (e) に, それぞれ $n = 6$, $n = 10$ の場合のテスト量子データを示した. 各状態に対応した異常値の結果を図 6.9 (c, f) に示した. 両者とも適切に学習が行えたと言える. 特に, $\{|\psi_j\rangle\}_{j=1}^M$ の分布の分散, すなわち (a, d) における青プロットの分布が (c, f) の赤線で示される dip の幅に対応しており, よく特徴が捉えられている. 最後に, 本項の設定について記す. 数値シミュレーションには QASM シミュレータを用い, 測定回数 (shot 数) の影響も含めた結果を示している. shot 数は, 学習過程では 1000 とし, 異常検知の際は 50 ($n = 10$ 含む) とした. この値は量子状態トモグラフィーを用いる場合と比較して, 明らかに大幅に少ない. また, 提案手法は学習過程で必要な shot 数としても少ない数ですむ.

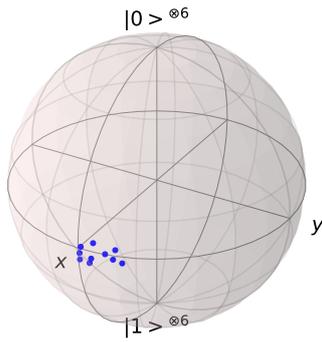


(a) $n = 6$ の場合

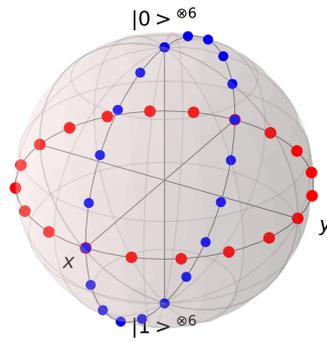


(b) $n = 10$ の場合

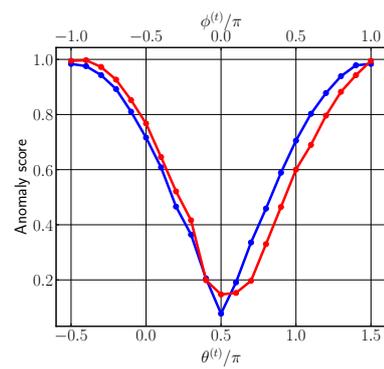
図 6.8: 学習曲線 (数値シミュレーション) . Reprinted figure from [46]. Reproduced with permission from Springer Nature.



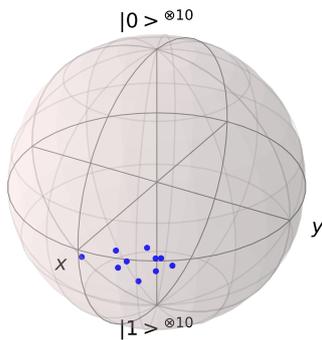
(a) 学習データ ($n = 6$)



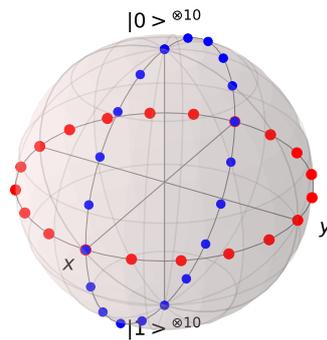
(b) テストデータ ($n = 6$)



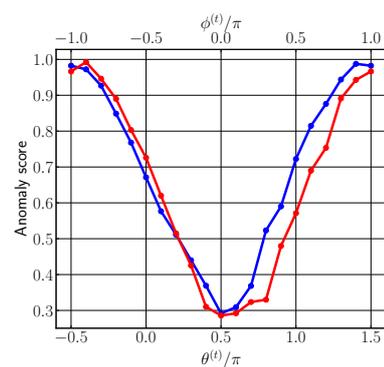
(c) Anomaly score ($n = 6$)



(d) 学習データ ($n = 10$)



(e) テストデータ ($n = 10$)



(f) Anomaly score ($n = 10$)

図 6.9: 局在データ, $n = 6, 10$ における異常検知の結果 (数値シミュレーション) . Reprinted figure from [46]. Reproduced with permission from Springer Nature.

表 6.5: 6.4.3 項の数値シミュレーションにおけるパラメタ設定

測定回数 (学習過程)	N_s	1000
測定回数 (異常検知の際)	N_{ad}	50
学習データ数	M	10
qubit 数	n	6, 10
潜在空間の次元数	N_z	2
層数	N_L	10

6.5 結論

古典機械学習では、多くの生成モデルが精力的に研究されている一方で、量子データを対象とした生成モデルに関する研究はいまだ少数である。本研究では、教師なし学習の枠組みで量子アンサンブルを学習する生成モデルの構築方法を新たに提案した。この目的のために、既存手法のように混合状態間ではなく、モデルから出力される純粋状態の集合と学習データとなる純粋状態の集合の間の距離を適切に測定するための最適輸送距離 (optimal transport loss; OTL) を提案した。さらに、学習過程における勾配消失問題を回避し、学習容易性を向上させるために、提案 OTL を local cost の和の形に修正した。但し、提案 OTL はこの修正のために距離の公理を満たさなくなる。そこで、この localized OTL が確率分布間の divergence の性質を満たし、生成モデルの学習コストとして好適であることを示した。すなわち、理論解析と数値シミュレーションの両面から提案 OTL の性質について解析を行った。結果として、学習データとなる量子アンサンブルが比較的低次元の多様体上にある場合には、提案 OTL が良い損失関数を構成できることが分かった。さらに数値シミュレーションの結果、cost が local であることから、提案 OTL を用いることで勾配消失問題を回避できることが示された。最後に、既存手法では実現不可能な問題設定における、量子状態に対する異常検知タスクが行えることを示した。

第7章

結論と展望

7.1 結論

本論文では、量子計算機を使用するにあたり基本的かつ重要な課題について考え、その対応について検討した。

第2章では、量子計算を考えるにあたり必要な基礎事項について説明した。

第3章では、最大の課題の一つとして知られている“効率的なデータロード・初期状態準備”を実現するための方法の一つとして、[49]で著者らによって提案されている手法を説明した。既存手法では使用する量子回路のサイズ n のスケーリングに対し、必要な計算ステップが指数関数的 $O(2^n)$ に増加してしまう問題を抱えていた。これに対し、提案手法である「Approximate amplitude encoding (AAE)」では、必要な計算ステップを $O(\text{poly } n)$ 程度に抑えることができる。本手法では、エンコーダを PQC で構成し、この回路のパラメタ学習に機械学習的手法を導入した。これにより、学習結果には有限の誤差が含まれることになるが、その誤差が許容できればデータロードコストが大幅に削減できる。これは、実装性を考えた場合には利点が多い。また、必ずしも高い精度のデータ埋込が必要な問題ばかりではないことに加え、ユーザー自身が必要な精度を選択できる点で実効性がある。

第4章では、Oracleの実装性の問題にアプローチする方法の一つとして、[44]で著者らによって提案されている手法を説明した。モチーフとして、有名かつ重要な量子アルゴリズムである Grover's algorithm を選択し、具体的な問題設定として量子画像におけるデータベース検索を検討した。ここでは一般的な Grover's algorithm をベースに、第3章で説明した AAE を応用することで、汎用性を担保したままブラックボックス化が可能なことを示した。加えて、以下の観点で議論を行った。

- 回路実装の観点から、CNOT gate の必要数を従来手法と比較した。結果、AAE の採用により大幅に CNOT gate 数が削減できることを示し、特に NISQ/early FTQC 時代における実装上の有効性を示した。
- データベース探索の問題設定におけるエンコード時のノイズ感度の観点から、amplitude encoding (AM) と basis encoding (BE) を比較した。結果、AE ではノイズ耐性は高いが識別性が悪く、BE では識別性が高いがノイズ耐性が若干悪いことが分かった。結論として、目的に応じた選択が必要であることが分かった。

第5章では、量子機械学習を効率的に実装するアプローチする方法の一つとして、[45]によって著者らによって提案している手法を説明した。ここでは古典 NTK 理論が直接適用できる量子-古典のハイブリッドフレームワークを提案し、以下を示した。

- 量子、古典の両部分をランダムに初期化し、古典部分のノード数（幅）を無限大の極限に取ることで、系全体に対して量子 NTK が定義でき、それが時不変になるため、学習過程のダイナ

ミクスを明示的に解析できる。

- 量子パートの最終的な出力が Gaussian process で表され, projected quantum kernel (PQK) の非線形関数になる。尚, PQK は, 汎化性を備える量子カーネルとされており, 提案モデルは高い機械学習性能を備えることが期待される。

そして, 実際に数値実験から, 提案ハイブリッドモデルが, 純量子・純古典のモデルに比べて明確に高い性能を示す事例を確認した。但し, 既存研究からも指摘されているように, 機械学習タスクの性能はデータエンコーダやデータセットの構造に強く依存する点には注意が必要である。提案モデルでは, 勾配消失の問題を回避するために量子パートは学習を行わず, 古典パートのみ学習を行ったが, データ構造に応じたエンコーダを得るために量子パートの学習も行うということもありえる。いずれにせよ, 古典機械学習の領域で開発されている成熟した機械学習のスキームを使いながらモデルの学習を行いつつ, PQK を活用できるフレームワークは有用性が高い。

第6章では, 量子ハードウェアの簡便なヘルスチェックの実現に向けて, 量子状態の異常検知を効率的に行う手法を検討した。具体的には, [167] によって著者らによって提案された手法を説明した。問題設定として, 量子状態のアンサンブルとしてサンプルデータセットを用意し, それらを純粋状態の集合として学習することをモチーフとし, 潜在変数を含む PQC で表現される量子生成モデルを提案した。本モデルのポイントは, 以下の通りである。

- 損失関数として, 量子状態のアンサンブルを適切に学習できるような最適輸送距離 (optimal transport loss; OTL) を新たに提案した。
- inversion-test を用いて入力データとモデル出力の間の ground cost を求め, その値を用いて OTL を算出する。
- 提案 OTL の導入により, ある精度を達成するために必要なサンプル数がシステムサイズには依存せず, 潜在変数次元のみに依存することを示した。すなわち, 問題設定に依っては高次元系であっても少量のサンプルから適切に学習が行えることを示した。
- cost の local 化を行うことで, 高次元系における勾配消失 (barren plateau) が回避できることを示した。

そして, このモデルの有効性を確認するために, 数値実験によって量子状態の異常検知タスクを示した。結果として, 既存手法に比べて大幅に少量のサンプル数で異常検知が行えることを示した。

7.2 展望

最後に, 今後の展望を記す。本論文では量子コンピュータを活用した効率的なワークフローを実現するために必要な要素技術を検討した。しかし, 未だ多くの課題が存在する。

まず, データ入力について記す。第3章に記した手法 [49] は, 任意の実ベクトル (の近似値) を従来比で指数関数的に少ないゲートで生成する手法を提供するが, 回路の生成に学習過程が必要なため, リアルタイムな処理には向かない。この点を踏まえ, 今後の研究には以下の方向性がある。

- 1つ目は, データベース検索のような同じ回路を何度も使い回すタスクを意識した方向である。生成するのに小さくない学習コストが必要となるデータベース状態の活用を考えると, オンライン学習 [190] や転移学習 [191] など, 既存の回路を再利用しつつ活用の幅を広げるための技術開発が重要だろう。加えて, 効率的に学習を進めるためのサブルーチンの開発も期待される。
- 2つ目は, 汎化能をもつ回路圧縮を利用する方向がある。筆者らが提案した手法は, 広義の回路圧縮技術とみることができる。これを入出力関係について汎化できればリアルタイムに効率的なデータ入力が可能になる。あらゆるクラスに対応するようなかたちでの実現には大きなハ-

ドルがあるが、事前に定義したある範囲の入力であれば可能性があるだろう。

補足として、[49]の適用クラスを任意の量子状態へと拡張した[192]について付記する。本手法では、学習に用いるコスト関数に fidelity を採用し、学習に classical shadow tomography [38, 39] を用いた。すなわち、原理的には埋め込みたい量子状態が分かっているならば、任意のエンコーダを置換できるようになった。一方、依然として大規模なデータを高精度でエンコードする回路を簡便に生成するには新しいアイデアが必要である。合わせて、これらを用いた多様なアプリケーションの提案も期待される。

次に、Oracle の実装と Grover's algorithm (GA) について記す。第 4 章では Oracle を汎用的な問題設定で利用可能な明示的な実装例を示した。また、モチーフとして GA を用いたが、GA を NISQ 上で実現するには様々な工夫が必要である。本研究の展開として、以下の方向性がある。

- 本論文で記した Oracle の明示的な実装方法を、他アルゴリズムの Oracle へ展開するための検討。本研究ではデータ構造を上手く利用して Oracle を実装した。すなわち、Oracle の引数にあたる領域とデータ領域を空間方向で分離し、両者がエンタングルするようにエンコードすることで意図した機能を実現した。同様の考え方で問題が解消するケースが他にもあるだろう。
- 量子計算を用いたデータベース検索が実益を生むアプリケーションの探索。本研究ではデモンストレーションとして画像検索を行ったが、世の中には扱うデータ空間が広く、本技術が好適なアプリケーションが存在すると考える。具体的には、DNA シーケンサや化学空間（ケミカルスペース）における検索などがあり得る。
- ハードウェアによる大規模実装へ向けた要素技術開発。GA は産業上重要な用途が多く、ハードウェアによる早期実現が期待されているが、特に Grover 演算子の実装がボトルネックとなる。これらの解消へ向けた新たな実装方法の検討は有益であろう。参考までにこれまでの関連研究を記す [193, 194, 195]。

続いて、量子-古典ハイブリッド計算について記す。現在、多くのハードウェアベンダーが NISQ の多ビット化と低ノイズ化、そしてその先にある誤り訂正の実装と FTQC の実現へ向けたロードマップを描いているが、大規模 FTQC の実現には未だ 10 年から 20 年以上がかかると言われている。このような状況と足元の古典機械学習の目覚ましい発展を鑑みれば、しばらくの間は古典機械学習と量子計算を合わせて活用していくことは必然だろう。第 5 章では、その一つの例として NTK 理論を応用した量子-古典ハイブリッド計算のフレームワークを提案した。本研究に関連する今後の展開として、以下の方向性がある。

- 量子優位性を示す実用的なデータセット・データ構造と、それに対応するエンコーダの明確化。本論文では提案手法のバリデーションのために経験論的にエンコーダを選択したが、処理対象のデータ構造に応じた適切なエンコーダの設計指針の明確化は、社会実装上とても価値がある。これに関し、本論文で示したようにモデル性能を解析的に評価できるという点は、より性能の高いエンコーダ・モデルを効率的に設計する上で有益である。
- 他の機械学習モデルとの接続。本研究では理論的に扱いやすい系として NTK 理論を用いたが、古典機械学習の領域では目下様々なフレームワークが提案されている。それらと量子計算の融合について調べることは大変興味深い。特に最近では拡散モデル [196] が画像処理や自然言語処理の領域を席卷しているが、量子計算へ応用することでどのような利点が生まれるかは興味深い。

補足として量子カーネル法に乱択測定を応用することでスケーラビリティを拡張した [197] について付記する。第 5 章 5.5.4 項でも触れた通り、量子カーネル法の欠点としてデータ数 N_D に対して

$O(N_D^2) \sim O(N_D^3)$ の計算量がかかる点がある。この問題に対し “random Fourier feature” [198] から着想を得た “deterministic quantum feature (DQF)” や “random quantum feature (RQF)” を用いることで計算量を $O(N_D D^2)$ に抑えることができる手法を提案した。ここで $D (\ll N_D)$ は特徴量次元である。このような、量子機械学習の適用範囲を拡張する方向の研究も重要だろう。

最後に、異常検知と量子生成モデルについて記す。量子状態の異常検知を考えた場合、量子コンピュータのヘルスチェックに加えて、例として以下の研究の方向性が挙げられる。

- 本研究では個々の量子状態を個別に測定するという前提を置いているが、量子優位性を検討する上で近年着目されているコヒーレントアクセスを許す問題設定も興味深い [199, 200]。尚、このシナリオでは、量子センサーのような外部プロセッサから届く量子状態を処理することを想定しており、将来的には社会的にも有用となる可能性がある。
- コストを entropy regularized Wasserstein distance [201, 202, 203, 204, 205] へ拡張することも興味深い。このコストの導入により、古典生成モデルではより高次元のデータを効率的に扱えるようになる。しかしながら、この場合も本論文で記したように localized cost が距離の公理を満たさなくなるため、詳細な解析が必要である。

実用的な電子計算機（コンピュータ）の登場から 75 年余りが経ち、我々は半導体技術に基づく非常に高度な計算技術を手に入れている。しかし、計算量等の問題から未だに解くことのできない問題が多くの領域に存在する。それらを解決し、学術・産業のさらなる発展に貢献しうる一つの可能性を秘めるのが量子コンピュータである。この技術が実際に社会実装され、広く活用されるには多くの課題・残件があるが、様々な分野からの取り組みにより、近い将来解決されることを期待する。合わせて、本研究がその一助になることを期待する。

謝辞

本研究は、多くの方のお力添えなくしては進めることができませんでした。この場をお借りし心より感謝申し上げます。

指導教官の山本直樹教授には、ハードウェアエンジニアだった私を社会人博士として受け入れて頂き、量子アルゴリズムの研究者として教育すべく、タイムリーかつ重要な課題を継続的に与えて頂きました。山本先生の幅広い領域に及ぶ知見や好奇心に導かれ、本研究を行うことができたように思います。また、夜間・休日にもディスカッションや論文の確認・査読対応に時間を割いて頂き、誠にありがとうございました。

武岡正裕教授、田中宗准教授、渡辺宙志准教授に副査をお受け頂き、大変感謝します。審査中には、研究コンセプトの中核を射るようなご質問、まとめ方やタイトルへのご助言を頂き、本論文にも数多くのフィードバックを頂きました。ありがとうございました。

本研究の活動拠点である慶應義塾大学量子コンピューティングセンター（KQCC）のメンバーからは、日々多くの刺激を頂き、研究を進めるエネルギーになりました。機械学習チームを牽引下さった上西慧理子氏、川口英明氏、金融チームを牽引下さった鈴木洋一氏、いつでも快くディスカッションに応じて下さった菅原道彦氏、渡邊宙志氏に感謝します。また、私と同じく企業から派遣された優秀な研究員の方々と苦楽を共にし、フランクに切磋琢磨し合えたことは、私にとって大変幸運でした。田中智樹氏、高王己氏、大西裕也氏、光田尚生氏、由本美香氏、長井遼氏、大塩耕平氏、加藤準平氏、佐藤勇氣氏、近藤瑠歩氏、市村達大氏、佐久間怜氏、前田洋太氏に感謝します。共にKQCCで活動し、博士同期生となる矢野碩志氏、鈴木雄大氏にも有意義な議論の場を頂きました。感謝します。日本IBMの小野寺民也副所長、渡辺日出雄氏（当時）、Raymond Harry Putra Rudy氏（当時）にも、研究の場に閉じず、様々な話題から刺激を頂く機会が多々ありました。大変感謝します。また、KQCC運営を支えて頂いている新美新一さん、秘書の相沢志乃さん、大上直子さんにも大変お世話になりました。

そして、共同研究者である宇野隼平氏、中路紘平氏、佐藤貴彦准教授には感謝してもしきれません。研究に対する姿勢やアプローチは研究者としての私のロールモデルになりました。共に活動する中で多くを学ばせて頂いただけでなく、刺激的でとても楽しい時間を過ごさせて頂きました。ありがとうございました。

さらに、KQCCへの参画を決めて頂いた鉄川弘樹統括部長、平山照峰主席技監（当時）、吉田憲一郎CEOは、量子コンピューティングという不確実性が高いテーマに対して、アルゴリズムの門外漢であった私を訪問研究員という形で送り出してくださいました。大きなチャレンジの機会を与えて頂いたことを大変感謝しています。そして、2018年にKQCCを設立され、その後弊社に参画のご提案をして下さったのが伊藤公平塾長（当時理工学部長、学部・修士時代の恩師）でした。修士を修了し、ちょうど10年目の節目に母校に戻り量子の研究に携われたのは強い因縁を感じざるを得ません。これまでの経験を糧とし、今後とも量子コンピューティング・量子テクノロジーの産業応用を目指し、微力ながら社会問題の解決に貢献していく所存です。

最後に、家族のサポートに心から感謝します。特に9歳の長男 丈翔、5歳の長女 千晴の元気一杯で好奇心に満ち溢れた姿は、いつも私を鼓舞してくれました。ありがとうございました。

付録

A 5.6.3 項の数値実験の詳細

第5章の5.6.3項「純古典モデルと純量子モデルに対するqcNNの優位性」を示す実験の詳細を記す。まず、モデルに含まれるパラメタ数 N_p と、学習に必要な量子回路の数 N_{qc} を表A.1に示す。第1行目と第2行目はモデルの構造に基づいて計算されるものであり、第3行目は本実験設定における具体的なパラメタ数 N_p である。ここで N_p がモデルによって異なる点に注意されたい。qNNの N_p は他の二つに比べて非常に小さいように見えるが、これは量子パートのパラメタ学習の計算コストを実装性を考慮して選択した結果である（量子デバイスのオペレーション速度は最先端の古典コンピュータに比べると劇的に遅いことに注意）。表A.2は、学習時のiteration数 N_{ite} とその値を用いて表A.1の N_{qc} から計算された、具体的な N_{qc} の値を示している。最後に、5.5.2項の数値実験で用いた擬似的なランダムユニタリ回路を図Aに示す。

表 A.1: 第5章 5.6.3 項で用いたモデルパラメタ数

		qcNN	qNN	cNN
パラメタ数 [*]	N_p	n_0	$3nL_q$	$(n+2)n_0$
学習に必要な回路数	N_{qc}	n_0	$2N_pN_{ite}$	-
パラメタ数 ^{*,**}	N_p	1k	60-150	4k-7k

^{*} 出力ノードのスケーリング、バイアスパラメタは含まず。

^{**} $n_0 = 1000$, $n = \{2, 3, 4, 5\}$, $L_q = 10$.

表 A.2: 第5章 5.6.3 項の数値実験で用いた iteration 数と学習に必要な回路数

Task \ Model		qcNN	qNN	cNN	
回帰 ^{**}	iteration 数	N_{ite}	2k-20k	1k	2k-20k
	学習に必要な回路数	N_{qc}	1k	120k-300k	-
分類 ^{**}	iteration 数	N_{ite}	1k	500	1k
	学習に必要な回路数	N_{qc}	1k	60k-150k	-

^{**} $n_0 = 1000$, $n = \{2, 3, 4, 5\}$, $L_q = 10$.

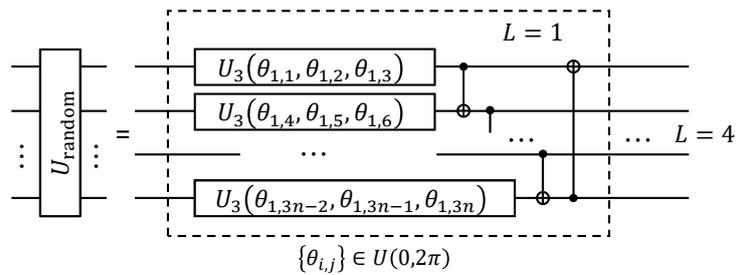


図 A: ランダムユニタリ回路. $U_3(\alpha, \beta, \gamma)$ は 3 つのオイラー角による single-qubit rotation gate であり, 角度 α, β, γ は一様分布 $U(0, 2\pi)$ からランダムに選択される. Reprinted figure from [45]. ©IOP Publishing. Reproduced with permission. All rights reserved.

B 式 (6.14) のシミュレーション結果

第 6 章の 6.3.1 節「データ数による近似誤差」の実験 B に関するシミュレーション結果を記す。図 B に式 (6.14) で表される経験的損失関数の近似誤差と式 (6.15) で表されるフィッティングの結果を示す。シミュレーションに用いたパラメタは表 6.1 の通りである。尚、図 6.3 は図 B を元に作成した。

C 図 6.4 の直感的な理解

第 6 章の 6.3.2 節「データ数による近似誤差」における図 6.4 で示されている、学習データ数と損失関数の近似誤差の関係について直感的な解釈を与える。尚、以下では測定回数 N_s は漸近理論が成り立つ程度に十分大きいことを仮定する。

まず、学習データ数 M が少ない場合の振る舞いが $M^{-1/2}$ に比例することを説明する。この場合、学習データ $\{\mathbf{x}_i\}_{i=1}^M$ は各々が十分離れており、最適輸送計画 $\{\pi_{i,j}\}_{i,j=1}^M$ は測定回数 N_s に影響を受けないことが予想される。多くの場合、最適輸送計画 $A = \{(i,j) | \pi_{i,j} > 0\}$ の non-zero な要素の数は M であり、各要素の値は $1/M$ である。

式 (6.11) において測定回数を N_s とした際の ground cost $c_{\text{local},i,j}$ の推定値は $\hat{c}_{\text{local},i,j}^{(N_s)} = \sqrt{\frac{1}{n} \sum_{k=1}^n \frac{1}{N_s} \sum_{s=1}^{N_s} X_{i,j,k}^{(s)}}$ で与えられる。但し、 $X_{i,j,k}^{(s)}$ は Bernoulli 分布 $1 - p_{i,j}^{(k)}$ に従うランダムな変数である。ここで平方根の中身に着目し $Y_{i,j}^{(s)} = \frac{1}{n} \sum_{k=1}^n X_{i,j,k}^{(s)}$ とおく。すると、中心極限定理により $\sqrt{N_s}(\sum_{s=1}^{N_s} Y_{i,j}^{(s)}/N_s - \mu_{i,j}) \sim \mathcal{N}(0, \sigma_{i,j}^2)$ が漸近的に成立する。但し、平均 $\mu_{i,j} = \sum_{k=1}^n (1 - p_{i,j}^{(k)})$ 、分散 $\sigma_{i,j}^2 = \sum_{k=1}^n (1 - p_{i,j}^{(k)})p_{i,j}^{(k)}$ である。すなわち、Delta method [206] から、ground cost の各成分の近似誤差は $\sqrt{N_s}(\hat{c}_{\text{local},i,j}^{(N_s)} - c_{\text{local},i,j}) \sim \mathcal{N}(0, \frac{\sigma_{i,j}^2}{4\mu_{i,j}})$ に従うことが分かる。

ここで全ての要素の平均や分散の大きさが同程度 $\mu_{i,j} \approx \mu$, $\sigma_{i,j} \approx \sigma \forall i, j$ だとすると、測定回数に起因する最適輸送距離の近似誤差は以下のように表せる。

$$\begin{aligned} \mathcal{L}_{\hat{c}_{\text{local}}^{(N_s)}} - \mathcal{L}_{c_{\text{local}}} &= \frac{1}{M} \sum_{(i,j) \in A} \left(\hat{c}_{\text{local},i,j}^{(N_s)} - c_{\text{local},i,j} \right) \\ &\sim \mathcal{N} \left(0, \sum_{(i,j) \in A} \frac{\sigma_{i,j}^2}{4N_s \mu_{i,j} M^2} \right) \\ &\approx \mathcal{N} \left(0, \frac{\sigma^2}{4N_s \mu M} \right). \end{aligned}$$

従って、以上の仮定の下では、学習データが少ない場合の近似誤差が $(N_s M)^{-1/2}$ と振る舞うことが予想される。

他方、学習データ数 M が多い領域における近似誤差の振る舞いは極値分布の理論 [207] から説明できる。学習データが多い場合は、同程度の大きさの ground cost $c_{\text{local},i,j}$ が多く存在すると考えられる。極端な場合として、全ての ground cost $\hat{c}_{\text{local},i,j}^{(N_s)}$ が同一の定数になることを考えてみる。ここで再び測定回数 N_s が十分に大きいと仮定すると、ground cost $\hat{c}_{\text{local},i,j}^{(N_s)}$ は正規分布に従う。その値を $\mathcal{N}(c, \frac{\sigma^2}{N_s})$ と記すことにする。すると、正規分布 $\mathcal{N}(0, \frac{\sigma^2}{N_s})$ に従う独立同分布な乱数 $\{X_{i,j}\}_{i,j=1}^M$ を

用いて、近似誤差は

$$\begin{aligned} \mathcal{L}_{\tilde{c}_{\text{local}}^{(N_s)}} - \mathcal{L}_{c_{\text{local}}} &\approx \min_{\{\pi_{i,j}\}_{i,j=1}^M} \sum_{i,j=1}^M X_{i,j} \pi_{i,j}, \\ \text{subject to } \sum_{i=1}^M \pi_{i,j} &= \frac{1}{M}, \sum_{j=1}^{M_g} \pi_{i,j} = \frac{1}{M}, \pi_{i,j} \geq 0. \end{aligned}$$

と書ける。今、貪欲法によりこの値を最小化することを考える。すなわち、まず初めに M^2 の要素の中の最小値 X_{i_1, j_1} を得て、次に i 行及び j 列を除いた残りの $(M-1)^2$ の要素の中から 2 番目に小さい値 X_{i_2, j_2} を得る、という具合である。ランダム変数 $X_{i,j}$ の累積分布関数を $F(x)$ と記すとすると、 k 個のデータの最小値の累積分布関数 $G(x, k)$ は

$$\begin{aligned} G(x, k) &= 1 - (1 - F(x))^k, \\ p(x, k) &= \frac{dG(x, k)}{dx} = k \frac{dF(x)}{dx} (1 - F(x))^{k-1}. \end{aligned}$$

と書ける。この時、貪欲法により小さい方から順に $x_1, x_2, x_3, \dots, x_M$ が得られる確率分布 $p(x_1, x_2, \dots, x_M)$ は

$$\begin{aligned} p(x_1, x_2, \dots, x_M) &= p(x_1, M^2) \frac{p(x_2, (M-1)^2) \theta(x_2 - x_1)}{1 - G(x_1, (M-1)^2)} \frac{p(x_3, (M-2)^2) \theta(x_3 - x_2)}{1 - G(x_2, (M-2)^2)} \times \dots \times \frac{p(x_M, 1^2) \theta(x_M - x_{M-1})}{1 - G(x_{M-1}, 1^2)} \\ &= \prod_{k=1}^M \frac{k^2}{2k-1} p(x_k, 2k-1) \theta(x_k - x_{k-1}), \end{aligned}$$

で表される。但し、 $\theta(x)$ は階段関数であり、最終行では $x_0 = -\infty$ とした。最後に、この確率分布の期待値を最頻値 (mode) で近似することを考える。 $p(x, k)$ の最頻値 (mode) は $x_{\text{mode}} \approx -\sigma \sqrt{2 \ln M / N_s}$ と書け、最終的に

$$\begin{aligned} \mathcal{L}_{\tilde{c}_{\text{local}}^{(N_s)}} - \mathcal{L}_{c_{\text{local}}} &\approx \frac{\sigma}{\sqrt{N_s}} \left(\frac{1}{M} \sum_{k=1}^M \sqrt{2 \ln(2k-1)} \right) \\ &\approx \frac{\sigma}{\sqrt{N_s}} \sqrt{2 \ln(2M-1)}. \end{aligned}$$

が得られる。以上から、学習データが多い領域における近似誤差は $N_s^{-1/2} \sqrt{\ln(M)}$ と振る舞うことがおおまかに理解できる。

D 6.3.3 項のシミュレーション結果

第 6 章の 6.3.3 項「提案 OTL の勾配の分散」に関するシミュレーション結果を記す。6.3.3 項で議論した通り、提案手法では式 (6.8) で表されるような local cost function から計算される ground cost を用いた損失関数を導入することで、勾配消失を回避する。勾配の分散のデータ数依存性を図 D に示す。用いたパラメタ設定は表 6.3 の通りである。尚、図 6.5(c) は図 D のデータを用いたものである。

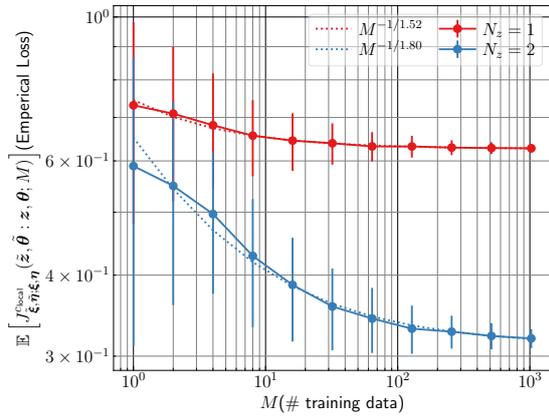
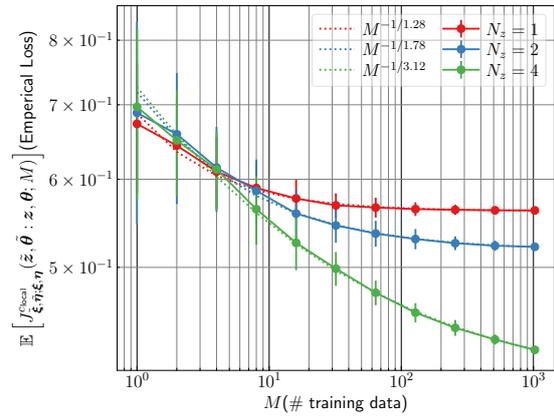
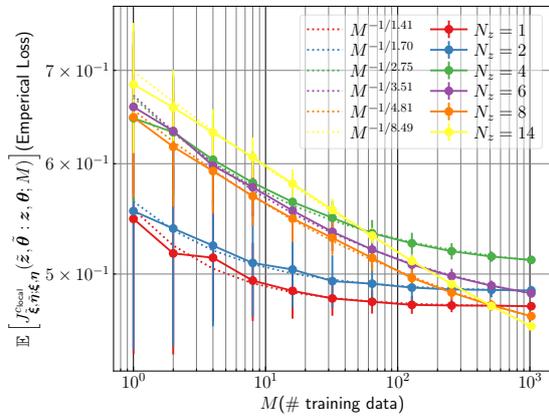
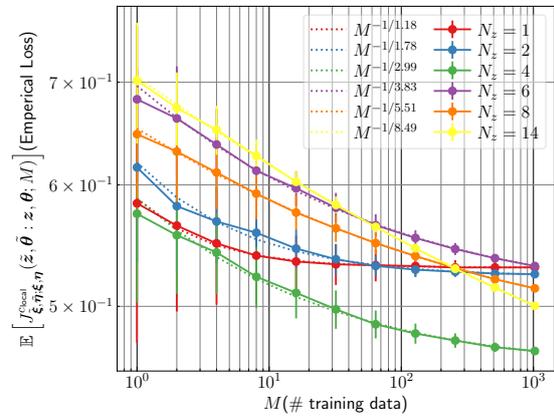
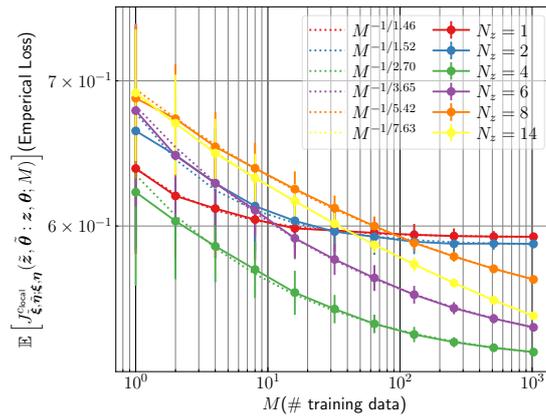
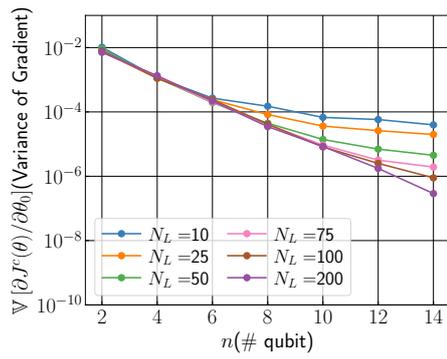
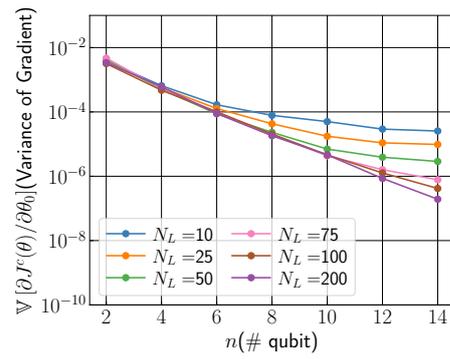
(a) $n = 1$ (b) $n = 2$ (c) $n = 4$ (d) $n = 6$ (e) $n = 8$

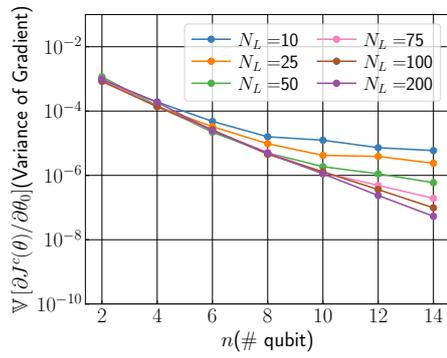
図 B: 提案 OTL の近似誤差 (6.14) と学習データ数 M の関係 (数値シミュレーション) . プロットは経験的 OTL の数値実験結果, 点線はフィッティング結果. フィッティングによって得られた式 (6.15) のパラメタ b は点線の凡例に記した. Reprinted figure from [46]. Reproduced with permission from Springer Nature.



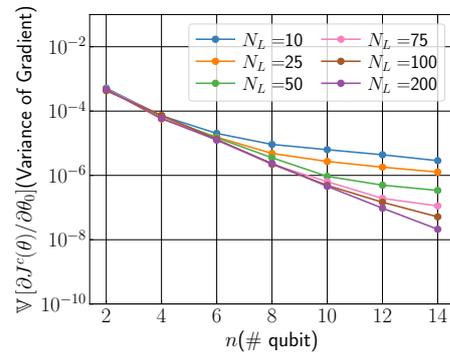
(a) $M = 1$



(b) $M = 2$



(c) $M = 8$



(d) $M = 16$

図 D: qubit 数と local cost に基づく提案 OTL の分散の関係 (数値シミュレーション) . Reprinted figure from [46]. Reproduced with permission from Springer Nature.

参考文献

- [1] Richard P. Feynman. Simulating physics with computers. In *Feynman and computation*, pages 133–153. CRC Press, 2018.
- [2] David Deutsch. Quantum theory, the church–turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117, 1985.
- [3] Peter W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. IEEE, 1994.
- [4] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219, 1996.
- [5] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15):150502, 2009.
- [6] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated quantum computation of molecular energies. *Science*, 309(5741):1704–1707, 2005.
- [7] Sam McArdle, Suguru Endo, Alán Aspuru-Guzik, Simon C. Benjamin, and Xiao Yuan. Quantum computational chemistry. *Rev. Mod. Phys.*, 92:015003, 2020.
- [8] Emmanuel E Haven. A discussion on embedding the Black–Scholes option pricing model in a quantum physics setting. *Physica A: Statistical Mechanics and its Applications*, 304(3-4):507–524, 2002.
- [9] Belal E Baaquie. *Quantum finance: Path integrals and Hamiltonians for options and interest rates*. Cambridge University Press, 2007.
- [10] Román Orús, Samuel Mugel, and Enrique Lizaso. Quantum computing for finance: Overview and prospects. *Reviews in Physics*, 4:100028, 2019.
- [11] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, 2014.
- [12] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Physical Review Letters*, 113(13):130503, 2014.
- [13] Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. Quantum-enhanced machine learning. *Physical Review Letters*, 117(13):130501, 2016.
- [14] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [15] John Preskill. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, 2018.
- [16] Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al.

- Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.
- [17] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L. O’ Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.
- [18] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O’Brien. Quantum error mitigation. *arXiv:2210.00921*, 2022.
- [19] Tianyi Peng, Aram W. Harrow, Maris Ozols, and Xiaodi Wu. Simulating large quantum circuits on a small quantum computer. *Physical Review Letters*, 125(15):150504, 2020.
- [20] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, 2022.
- [21] Martin Plesch and Časlav Brukner. Quantum-state preparation with universal gate decompositions. *Physical Review A*, 83(3):032302, 2011.
- [22] Vivek V. Shende, Stephen S. Bullock, and Igor L. Markov. Synthesis of quantum-logic circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25(6):1000–1010, 2006.
- [23] Lov Grover and Terry Rudolph. Creating superpositions that correspond to efficiently integrable probability distributions. *arXiv preprint quant-ph/0208112*, 2002.
- [24] M. Möttönen, J. Vartiainen, V. Bergholm, and M. Salomaa. Transformation of quantum states using uniformly controlled rotations. *Quantum Information and Computation.*, 5:467–473, 2005.
- [25] V. Shende and I. Markov. Quantum circuits for incompletely specified two-qubit operators. *Quantum Information and Computation.*, 5:49–57, 2005.
- [26] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52:3457–3467, 1995.
- [27] Juha J. Vartiainen, Mikko Möttönen, and Martti M. Salomaa. Efficient decomposition of quantum gates. *Physical Review Letters*, 92:177902, 2004.
- [28] Vivek V. Shende, Igor L. Markov, and Stephen S. Bullock. Minimal universal two-qubit controlled-not-based circuits. *Physical Review A*, 69:062321, 2004.
- [29] Gui-Lu Long and Yang Sun. Efficient scheme for initializing a quantum register with an arbitrary superposed state. *Physical Review A*, 64(1):014303, 2001.
- [30] Ville Bergholm, Juha J. Vartiainen, Mikko Mottonen, and Martti M. Salomaa. Quantum circuits with uniformly controlled one-qubit gates. *Physical Review A*, 71(5):052330, 2005.
- [31] Raban Iten, Roger Colbeck, Ivan Kukuljan, Jonathan Home, and Matthias Christandl. Quantum circuits for isometries. *Physical Review A*, 93:032318, 2016.
- [32] Xiaoming Sun, Guojing Tian, Shuai Yang, Pei Yuan, and Shengyu Zhang. Asymptotically optimal circuit depth for quantum state preparation and general unitary synthesis. *arXiv:2108.06150*, 2021.
- [33] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Architectures for a quantum random access memory. *Physical Review A*, 78(5):052310, 2008.
- [34] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory.

- Physical Review Letters*, 100(16):160501, 2008.
- [35] A Yu Kitaev. Quantum measurements and the abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.
 - [36] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2001.
 - [37] Andreas Elben, Steven T Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, 2023.
 - [38] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.
 - [39] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Physical Review Letters*, 127:030503, 2021.
 - [40] Andrew Wack, Hanhee Paik, Ali Javadi-Abhari, Petar Jurcevic, Ismael Faro, Jay M Gambetta, and Blake R Johnson. Quality, speed, and scale: three key attributes to measure the performance of near-term quantum computers. *arXiv:2110.14108*, 2021.
 - [41] Satoshi Hara, Takafumi Ono, Ryo Okamoto, Takashi Washio, and Shigeki Takeuchi. Anomaly detection in reconstructed quantum states using a machine-learning technique. *Physical Review A*, 89(2):022104, 2014.
 - [42] Satoshi Hara, Takafumi Ono, Ryo Okamoto, Takashi Washio, and Shigeki Takeuchi. Quantum-state anomaly detection for arbitrary errors using a machine-learning technique. *Physical Review A*, 94(4):042341, 2016.
 - [43] Kouhei Nakaji, Shumpei Uno, Yohichi Suzuki, Rudy Raymond, Tamiya Onodera, Tomoki Tanaka, Hiroyuki Tezuka, Naoki Mitsuda, and Naoki Yamamoto. Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicator. *arXiv:2103.13211*, 2021.
 - [44] Hiroyuki Tezuka, Kouhei Nakaji, Takahiko Satoh, and Naoki Yamamoto. Grover search revisited: Application to image pattern matching. *Physical Review A*, 105:032440, 2022.
 - [45] Kouhei Nakaji, Hiroyuki Tezuka, and Naoki Yamamoto. Quantum-classical hybrid neural networks in the neural tangent kernel regime. *Quantum Sci. Technol.*, 9(1), 2024.
 - [46] Hiroyuki Tezuka, Shumpei Uno, and Naoki Yamamoto. Generative model for learning quantum ensemble with optimal transport loss. *Quantum Machine Intelligence*, in press, 2024.
 - [47] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.
 - [48] Maria Schuld and Francesco Petruccione. *Supervised learning with quantum computers*. Springer, 2018.
 - [49] Kouhei Nakaji, Shumpei Uno, Yohichi Suzuki, Rudy Raymond, Tamiya Onodera, Tomoki Tanaka, Hiroyuki Tezuka, Naoki Mitsuda, and Naoki Yamamoto. Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators. *Physical Review Research*, 4:023136, 2022.
 - [50] Jian Zhao, Yu-Chun Wu, Guang-Can Guo, and Guo-Ping Guo. State preparation based on quantum phase estimation. *arXiv:1912.05335*, 2019.
 - [51] Gregory Rosenthal. Query and depth upper bounds for quantum unitaries via Grover search. *arXiv:2111.07992*, 2021.

- [52] Xiao-Ming Zhang, Man-Hong Yung, and Xiao Yuan. Low-depth quantum state preparation. *Physical Review Research*, 3:043200, 2021.
- [53] Zhicheng Zhang, Qisheng Wang, and Mingsheng Ying. Parallel quantum algorithm for Hamiltonian simulation. *arXiv:2105.11889*, 2021.
- [54] Lov K Grover. Synthesis of quantum superpositions by quantum computation. *Physical Review Letters*, 85(6):1334, 2000.
- [55] Yuval R Sanders, Guang Hao Low, Artur Scherer, and Dominic W Berry. Black-box quantum state preparation without arithmetic. *Physical Review Letters*, 122(2):020502, 2019.
- [56] Shengbin Wang, Zhimin Wang, Guolong Cui, Shangshang Shi, Ruimin Shang, Lixin Fan, Wendong Li, Zhiqiang Wei, and Yongjian Gu. Fast black-box quantum state preparation based on linear combination of unitaries. *Quantum Information Processing*, 20(8):1–14, 2021.
- [57] Johannes Bausch. Fast black-box quantum state preparation. *arXiv:2009.10709*, 2020.
- [58] Andrei N. Soklakov and Rüdiger Schack. Efficient state preparation for a register of quantum bits. *Physical Review A*, 73:012307, 2006.
- [59] Christa Zoufal, Aurélien Lucchi, and Stefan Woerner. Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 5(1):1–9, 2019.
- [60] Nasir Ahmed, Kamisetty Ramamohan Rao, Nasir Ahmed, and Kamisetty Ramamohan Rao. Walsh-hadamard transform. *Orthogonal transforms for digital signal processing*, pages 99–152, 1975.
- [61] Robin Scheibler, Saeid Haghhighatshoar, and Martin Vetterli. A fast hadamard transform for signals with sublinear sparsity in the transform domain. *IEEE Transactions on Information Theory*, 61(4):2115–2132, 2015.
- [62] Bharath K. Sriperumbudur, A. Gretton, K. Fukumizu, G. Lanckriet, and B. Schölkopf. Injective Hilbert space embeddings of probability measures. In *COLT*, 2008.
- [63] K. Fukumizu, A. Gretton, Xiaohai Sun, and B. Schölkopf. Kernel measures of conditional dependence. In *NIPS*, 2007.
- [64] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi. The Born supremacy: quantum advantage and training of an Ising Born machine. *npj Quantum Information*, 6(1):1–11, 2020.
- [65] Gavin E Crooks. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv:1905.13311*, 2019.
- [66] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.
- [67] Nathan Wiebe, Ashish Kapoor, and Krysta Svore. Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning. *Quantum Inf. Comput.*, 15:316–356, 2015.
- [68] Austin Gilliam, Stefan Woerner, and Constantin Gionciulea. Grover adaptive search for constrained polynomial binary optimization. *Quantum*, 5(428):428, 2021.
- [69] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA, July 1996. Association for Computing Machinery.

- [70] L K Grover. Synthesis of quantum superpositions by quantum computation. *Physical Review Letters*, 85(6):1334–1337, 2000.
- [71] Yuval R Sanders, Guang Hao Low, Artur Scherer, and Dominic W Berry. Black-Box quantum state preparation without arithmetic. *Physical Review Letters*, 122(2):020502, 2019.
- [72] Martin Plesch and Āaslav Brukner. Quantum-state preparation with universal gate decompositions. *Physical Review A*, 83(3):032302, 2011.
- [73] V V Shende, S S Bullock, and I L Markov. Synthesis of quantum-logic circuits. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, 25(6):1000–1010, June 2006.
- [74] Vojtĉch Havlĉek, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, 2019.
- [75] A A Ezhov, A V Nifanova, and Dan Ventura. Quantum associative memory with distributed queries. *Inf. Sci.*, 128(3):271–293, 2000.
- [76] Kouhei Nakaji. Faster amplitude estimation. *Quantum Inf. Comput.*, 20:1109–1122, 2020.
- [77] Yohichi Suzuki, Shumpei Uno, Rudy Raymond, Tomoki Tanaka, Tamiya Onodera, and Naoki Yamamoto. Amplitude estimation without phase estimation. *Quantum Inf. Process.*, 19(2):1–17, 2020.
- [78] Adriano Barenco, Charles H Bennett, Richard Cleve, David P DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A Smolin, and Harald Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52(5):3457, 1995.
- [79] Ville Bergholm, Juha J Vartiainen, Mikko MÖttönen, and Martti M Salomaa. Quantum circuits with uniformly controlled one-qubit gates. *Physical Review A*, 71(5):052330, 2005.
- [80] Zhaobin Wang, Minzhe Xu, and Yaonan Zhang. Review of quantum image processing. *Arch. Comput. Methods Eng.*, pages 1–25, 2021.
- [81] Phuc Q Le, Fangyan Dong, and Kaoru Hirota. A flexible representation of quantum images for polynomial preparation, image compression, and processing operations. *Quantum Inf. Process.*, 10(1):63–84, 2011.
- [82] Bo Sun, A Ilyasu, Fei Yan, Fangyan Dong, and Kaoru Hirota. An rgb multi-channel representation for images on quantum computers. *J. Adv. Comput. Intell. Intell. Inform.*, 17(3), 2013.
- [83] Fei Yan, Sihao Jiao, Abdullah M Ilyasu, and Zhengang Jiang. Chromatic framework for quantum movies and applications in creating montages. *Frontiers of Computer Science*, 12:736–748, 2018.
- [84] Xi-Wei Yao, Hengyan Wang, Zeyang Liao, Ming-Cheng Chen, Jian Pan, Jun Li, Kechao Zhang, Xingcheng Lin, Zhehui Wang, Zhihuang Luo, Wenqiang Zheng, Jianzhong Li, Meisheng Zhao, Xinhua Peng, and Dieter Suter. Quantum image processing and its application to edge detection: Theory and experiment. *Physical Review X*, 7(3):031041, 2017.
- [85] Yue Ruan, Xiling Xue, and Yuanxia Shen. Quantum image processing: Opportunities and challenges. *Math. Probl. Eng.*, 2021, 2021.
- [86] Yi Zhang, Kai Lu, Yinghui Gao, and Mo Wang. NEQR: a novel enhanced quantum representation of digital images. *Quantum Inf. Process.*, 12(8):2833–2860, 2013.
- [87] MD SAJID ANIS and Héctor Abraham, et al. Qiskit: An Open-source Framework for

- Quantum Computing, 2021.
- [88] Diederik P. Kingma and Jimmy Ba. Adam: A method for Stochastic Optimization. *arXiv:1412.6980*, 2014.
 - [89] Dmitri Maslov. Advantages of using relative-phase toffoli gates with an application to multiple control toffoli optimization. *Physical Review A*, 93(2):022311, 2016.
 - [90] Ho Lun Tang, V O Shkolnikov, George S Barron, Harper R Grimsley, Nicholas J Mayhall, Edwin Barnes, and Sophia E Economou. qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansätze on a quantum processor. *PRX Quantum*, 2(2):020310, 2021.
 - [91] Nikolay V Tkachenko, James Sud, Yu Zhang, Sergei Tretiak, Petr M Anisimov, Andrew T Arrasmith, Patrick J Coles, Lukasz Cincio, and Pavel A Dub. Correlation-informed permutation of qubits for reducing ansatz depth in the variational quantum eigensolver. *PRX Quantum*, 2(2):020337, 2021.
 - [92] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):1–6, 2018.
 - [93] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019.
 - [94] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12(1):1–12, 2021.
 - [95] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Sci. Technol.*, 6(2):025008, 2021.
 - [96] Lov K Grover. Trade-offs in the quantum search algorithm. *Physical Review A*, 66(5):052314, 2002.
 - [97] Ji Liu and Huiyang Zhou. Hardware efficient quantum search algorithm. *arXiv:2103.14196*, 2021.
 - [98] Marcin Briański, Jan Gwinner, Vladyslav Hlembotskyi, Witold Jarnicki, Szymon Pliś, and Adam Szady. Introducing structure to expedite quantum searching. *Physical Review A*, 103(6):062425, 2021.
 - [99] Vladyslav Hlembotskyi, Rafał Burczyński, Witold Jarnicki, Adam Szady, and Jan Tułowiecki. Efficient unstructured search implementation on current ion-trap quantum processors. *arXiv:2010.03841*, 2020.
 - [100] Jan Gwinner, Marcin Briański, Wojciech Burkot, Łukasz Czerwiński, and Vladyslav Hlembotskyi. "Benchmarking 16-element quantum search algorithms on superconducting quantum processors". *arXiv:2007.06539*, 2020.
 - [101] Kun Zhang, Pooja Rao, Kwangmin Yu, Hyunkyung Lim, and Vladimir Korepin. Implementation of efficient quantum search algorithms on NISQ computers. *Quantum Inf. Process.*, 20(7):233, 2021.
 - [102] UCI machine learning repository: Optical recognition of handwritten digits data set. <https://archive.ics.uci.edu/ml/datasets/optical+recognition+of+handwritten+digits>.
 - [103] MV Altaisky. Quantum neural network. *arXiv preprint quant-ph/0107012*, 2001.

- [104] Edward Farhi and Hartmut Neven. Classification with quantum neural networks on near term processors. *arXiv:1802.06002*, 2018.
- [105] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.
- [106] Maria Schuld, Alex Bocharov, Krysta M Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308, 2020.
- [107] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017.
- [108] Kwok Ho Wan, Oscar Dahlsten, Hlér Kristjánsson, Robert Gardner, and MS Kim. Quantum generalisation of feedforward neural networks. *npj Quantum information*, 3(1):1–8, 2017.
- [109] Guillaume Verdon, Jason Pye, and Michael Broughton. A universal training algorithm for quantum deep learning. *arXiv:1806.09729*, 2018.
- [110] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao. Expressive power of parametrized quantum circuits. *Physical Review Research*, 2(3):033125, 2020.
- [111] Jonathan Romero and Alán Aspuru-Guzik. Variational quantum generators: Generative adversarial quantum machine learning for continuous distributions. *Advanced Quantum Technologies*, 4(1):2000003, 2021.
- [112] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J. Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. *Nature communications*, 11(1):1–6, 2020.
- [113] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021.
- [114] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, 2019.
- [115] Rongxin Xia and Sabre Kais. Hybrid quantum-classical neural network for calculating ground state energies of molecules. *Entropy*, 22(8):828, 2020.
- [116] Andrea Mari, Thomas R Bromley, Josh Izaac, Maria Schuld, and Nathan Killoran. Transfer learning in hybrid classical-quantum neural networks. *Quantum*, 4:340, 2020.
- [117] J. McClean, S. Boixo, V. Smelyanskiy, R. Babbush, and H. Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9, 2018.
- [118] Stefan H. Sack, Raimel A. Medina, Alexios A. Michailidis, Richard Kueng, and Maksym Serbyn. Avoiding barren plateaus using classical shadows. *PRX Quantum*, 3(2):020365, 2022.
- [119] Andrea Skolik, Jarrod R. McClean, Masoud Mohseni, Patrick van der Smagt, and Martin Leib. Layerwise learning for quantum neural networks. *Quantum Machine Intelligence*, 3(1):1–11, 2021.
- [120] James Dborin, Fergus Barratt, Vinul Wimalaweera, Lewis Wright, and Andrew G Green. Matrix product state pre-training for quantum machine learning. *Quantum Science and Technology*, 7(3):035014, 2022.
- [121] Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3(1):010313, 2022.

- 2022.
- [122] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement-induced barren plateaus. *PRX Quantum*, 2(4):040316, 2021.
 - [123] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1):1–11, 2021.
 - [124] Taylor L. Patti, Khadijeh Najafi, Xun Gao, and Susanne F. Yelin. Entanglement devised barren plateau mitigation. *Physical Review Research*, 3(3):033090, 2021.
 - [125] Zoë Holmes, Andrew Arrasmith, Bin Yan, Patrick J. Coles, Andreas Albrecht, and Andrew T. Sornborger. Barren plateaus preclude learning scramblers. *Physical Review Letters*, 126(19):190501, 2021.
 - [126] Hsin-Yuan Huang, Michael Broughton, Masoud Mohseni, Ryan Babbush, Sergio Boixo, Hartmut Neven, and Jarrod R. McClean. Power of data in quantum machine learning. *Nature communications*, 12(1):1–9, 2021.
 - [127] Jonas Kübler, Simon Buchholz, and Bernhard Schölkopf. The inductive bias of quantum kernels. In *Advances in Neural Information Processing Systems 34*, pages 12661–12673, 2021.
 - [128] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, 2021.
 - [129] Evan Peters, João Caldeira, Alan Ho, Stefan Leichenauer, Masoud Mohseni, Hartmut Neven, Panagiotis Spentzouris, Doug Strain, and Gabriel N Perdue. Machine learning of high dimensional data on a noisy quantum processor. *npj Quantum Information*, 7(1):1–5, 2021.
 - [130] Teresa Sancho-Lorente, Juan Román-Roche, and David Zueco. Quantum kernels to learn the phases of quantum matter. *Physical Review A*, 105(4):042432, 2022.
 - [131] Maria Schuld and Nathan Killoran. Is quantum advantage the right goal for quantum machine learning? *PRX Quantum*, 3:030101, 2022.
 - [132] Maria Schuld. Supervised quantum machine learning models are kernel methods. *arXiv:2101.11020*, 2021.
 - [133] Gian Gentinetta, Arne Thomsen, David Sutter, and Stefan Woerner. The complexity of quantum support vector machines. *arXiv:2203.00031*, 2022.
 - [134] Supanut Thanasilp, Samson Wang, M Cerezo, and Zoë Holmes. Exponential concentration and untrainability in quantum kernel methods. *arXiv:2208.11060*, 2022.
 - [135] Arthur Jacot, Franck Gabriel, and Clement Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems 31*, pages 8571–8580, 2018.
 - [136] Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. Deep neural networks as gaussian processes. *arXiv:1711.00165*, 2017.
 - [137] Sanjeev Arora, Simon Du, Wei Hu, Zhiyuan Li, and Ruosong Wang. Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *International Conference on Machine Learning*, pages 322–332. PMLR, 2019.
 - [138] Zeyuan Allen-Zhu, Yuanzhi Li, and Yingyu Liang. Learning and generalization in overparameterized neural networks, going beyond two layers. In *Advances in Neural Information*

- Processing Systems 32*, pages 6155–6166, 2019.
- [139] Yuan Cao and Quanquan Gu. Generalization error bounds of gradient descent for learning over-parameterized deep relu networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3349–3356, 2020.
 - [140] Atsushi Nitanda, Geoffrey Chinot, and Taiji Suzuki. Gradient descent can learn less over-parameterized two-layer neural networks on classification problems. *arXiv:1905.09870*, 2019.
 - [141] Jaehoon Lee, Lechao Xiao, Samuel S. Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems 32*, pages 8570–8581, 2019.
 - [142] Junyu Liu, Francesco Tacchino, Jennifer R Glick, Liang Jiang, and Antonio Mezzacapo. Representation learning via quantum neural tangent kernels. *PRX Quantum*, 3(3):030323, 2022.
 - [143] Junyu Liu, Khadijeh Najafi, Kunal Sharma, Francesco Tacchino, Liang Jiang, and Antonio Mezzacapo. Analytic theory for the dynamics of wide quantum neural networks. *Physical Review Letters*, 130:150601, 2023.
 - [144] Norihito Shirai, Kenji Kubo, Kosuke Mitarai, and Keisuke Fujii. Quantum tangent kernel. *arXiv:2111.02951*, 2021.
 - [145] Junyu Liu, Zexi Lin, and Liang Jiang. Laziness, barren plateau, and noise in machine learning. *arXiv:2206.09313*, 2022.
 - [146] Xinbiao Wang, Junyu Liu, Tongliang Liu, Yong Luo, Yuxuan Du, and Dacheng Tao. Symmetric pruning in quantum neural networks. *arXiv:2208.14057*, 2022.
 - [147] John Charles Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.
 - [148] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022.
 - [149] Michael Ragone, Paolo Braccia, Quynh T Nguyen, Louis Schatzki, Patrick J Coles, Frederic Sauvage, Martin Larocca, and M Cerezo. Representation theory for geometric quantum machine learning. *arXiv:2210.07980*, 2022.
 - [150] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009.
 - [151] Yudai Suzuki, Hideaki Kawaguchi, and Naoki Yamamoto. Quantum fisher kernel for mitigating the vanishing similarity issue. *arXiv:2210.16581*, 2022.
 - [152] Bernhard Schölkopf, Alexander J Smola, Francis Bach, et al. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
 - [153] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.
 - [154] Yasunari Suzuki, Yoshiaki Kawase, Yuya Masumura, Yuria Hiraga, Masahiro Nakadai, Jiabao Chen, Ken M Nakanishi, Kosuke Mitarai, Ryosuke Imai, Shiro Tamiya, et al.

- Qulacs: a fast and versatile quantum circuit simulator for research purpose. *Quantum*, 5:559, 2021.
- [155] qulacs.gate.randomunitary.
- [156] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103:032430, 2021.
- [157] Francisco Javier Gil Vidal and Dirk Oliver Theis. Input redundancy for parameterized quantum circuits. *Frontiers in Physics*, 8:297, 2020.
- [158] qiskit.ml.datasets.ad_hoc_data.
- [159] Quantum-enhanced support vector machine (qsvm). https://qiskit.org/documentation/stable/0.24/tutorials/machine_learning/01_qsvm_classification.html.
- [160] UCI machine learning repository: Heart disease data set. <https://archive.ics.uci.edu/ml/datasets/heart+disease>.
- [161] Subir Sachdev. Quantum phase transitions. *Physics world*, 12(4):33, 1999.
- [162] F Duncan M Haldane. Nonlinear field theory of large-spin heisenberg antiferromagnets: semiclassically quantized solitons of the one-dimensional easy-axis néel state. *Physical Review Letters*, 50(15):1153, 1983.
- [163] Frank Pollmann and Ari M Turner. Detection of symmetry-protected topological phases in one dimension. *Physical review b*, 86(12):125441, 2012.
- [164] Yusen Wu, Bujiao Wu, Jingbo Wang, and Xiao Yuan. Quantum Phase Recognition via Quantum Kernel Methods. *Quantum*, 7:981, 2023.
- [165] Ken N Okada, Keita Osaki, Kosuke Mitarai, and Keisuke Fujii. Identification of topological phases using classically-optimized variational quantum eigensolver. *arXiv:2202.02909*, 2022.
- [166] Johannes Herrmann, Sergi Masot Llima, Ants Remm, Petr Zapletal, Nathan A. McMahon, Colin Scarato, Francois Swiadek, Christian Kraglund Andersen, Christoph Hellings, Sebastian Krinner, Nathan Lacroix, Stefania Lazar, Michael Kerschbaum, Dante Colao Zanuz, Graham J. Norris, Michael J. Hartmann, Andreas Wallraff, and Christophe Eichler. Realizing quantum convolutional neural networks on a superconducting quantum processor to recognize quantum phases. *Nature Communications*, 13(4144), 2022.
- [167] Hiroyuki Tezuka, Shumpei Uno, and Naoki Yamamoto. Generative model for learning quantum ensemble via optimal transport loss. *arXiv:2210.10743*, 2022.
- [168] Leon Bottou, Martin Arjovsky, David Lopez-Paz, and Maxime Oquab. Geometrical insights for implicit generative modeling. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, pages 229–268. Springer, 2018.
- [169] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27:2672–2680, 2014.
- [170] Gabriel Peyre and Marco Cuturi. *Computational Optimal Transport: With Applications to Data Science (Foundations and Trends in Machine Learning)*. Now Publishers, paperback edition, 2 2019.
- [171] Yann Ollivier, Pajot Herve, and Cedric Villani, editors. *Optimal Transport: Theory and Applications*. London Mathematical Society Lecture Note Series. Cambridge University

- Press, 2014.
- [172] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87. Birkhäuser Cham, 2015.
 - [173] Gabriel Peyre and Marco Cuturi. Editorial IMA IAI - Information and Inference special issue on optimal transport in data sciences. *Information and Inference: A Journal of the IMA*, 8(4):655–656, 12 2019.
 - [174] Leonid V Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.
 - [175] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
 - [176] Richard Mansfield Dudley. The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1):40–50, 1969.
 - [177] Jonathan Weed and Francis Bach. Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648, 2019.
 - [178] Shouvanik Chakrabarti, Huang Yiming, Tongyang Li, Soheil Feizi, and Xiaodi Wu. Quantum Wasserstein Generative Adversarial Networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
 - [179] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, paperback edition, 9 2000.
 - [180] Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001.
 - [181] Sumeet Khatri, Ryan LaRose, Alexander Poremba, Lukasz Cincio, Andrew T. Sornborger, and Patrick J. Coles. Quantum-assisted quantum compiling. *Quantum*, 3:140, 2019.
 - [182] Kunal Sharma, Sumeet Khatri, Marco Cerezo, and Patrick J Coles. Noise resilience of variational quantum compiling. *New Journal of Physics*, 22(4):043006, 2020.
 - [183] Alba Cervera-Lierta, Jakob S Kottmann, and Alán Aspuru-Guzik. Meta-variational quantum eigensolver: Learning energy profiles of parameterized hamiltonians for quantum simulation. *PRX Quantum*, 2(2):020329, 2021.
 - [184] Pierre-Luc Dallaire-Demers and Nathan Killoran. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.
 - [185] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019.
 - [186] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.
 - [187] T Ide. Introduction to Anomaly Detection Using Machine Learning—a Practical Guide With R (in Japanese). *Corona Publishing*, pages 132–139, 2015.
 - [188] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer, 2017.
 - [189] Kouhei Nakaji and Naoki Yamamoto. Expressibility of the alternating layered ansatz for quantum computation. *Quantum*, 5:434, 2021.
 - [190] Léon Bottou and Yann Cun. Large scale online learning. *Advances in neural information*

- processing systems*, 16, 2003.
- [191] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.
 - [192] Naoki Mitsuda, Kohei Nakaji, Yohichi Suzuki, Tomoki Tanaka, Rudy Raymond, Hiroyuki Tezuka, Tamiya Onodera, and Naoki Yamamoto. Approximate complex amplitude encoding algorithm and its application to classification problem in financial operations. *arXiv:2211.13039*, 2022.
 - [193] Ken M Nakanishi, Takahiko Satoh, and Synge Todo. Quantum-gate decomposer. *arXiv:2109.13223*, 2021.
 - [194] Takahiko Satoh, Yasuhiro Ohkura, and Rodney Van Meter. Subdivided phase oracle for nisq search algorithms. *IEEE Transactions on Quantum Engineering*, 1:1–15, 2020.
 - [195] Ji Liu and Huiyang Zhou. Hardware efficient quantum search algorithm. *arXiv:2103.14196*, 2021.
 - [196] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.
 - [197] Kouhei Nakaji, Hiroyuki Tezuka, and Naoki Yamamoto. Deterministic and random features for large-scale quantum kernel machine. *arXiv:2209.01958*, 2022.
 - [198] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. *Advances in neural information processing systems*, 20, 2007.
 - [199] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature Communications*, 13(1):1–9, 2022.
 - [200] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. Quantum advantage in learning from experiments. *Science*, 376(6598):1182–1186, 2022.
 - [201] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In C.J. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, pages 2292–2300. Curran Associates, Inc., 2013.
 - [202] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trounev, and Gabriel Peyré. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences. *arXiv:1810.08278*, 2018.
 - [203] Aude Genevay, Lénaïc Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of Sinkhorn divergences. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR, 16–18 Apr 2019.
 - [204] Shun-ichi Amari, Naotsugu Tsuchiya, and Masafumi Oizumi. Geometry of information integration. In *Information Geometry and its Applications IV*, pages 3–17. Springer, 2016.
 - [205] Shun-ichi Amari, Ryo Karakida, and Masafumi Oizumi. Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem. *Information Geometry*, 1(1):13–37, 2018.
 - [206] Robert J Serfling. *Approximation theorems of mathematical statistics*, volume 162. John

Wiley & Sons, 2009.

- [207] Laurens De Haan, Ana Ferreira, and Ana Ferreira. *Extreme value theory: an introduction*, volume 21. Springer, 2006.