

最適輸送を用いた量子状態学習

宇野 隼平ⁱ, 手塚 宙之^{ii,iii}, 山本 直樹^{iv}

Learning Quantum States using Optimal Transport

Shumpei UNO, Hiroyuki TEZUKA, Naoki YAMAMOTO

近年, 既存のデータにもとづいて新たなデータを生成する機械学習モデル (生成モデル) が注目を集めており, 画像生成, 異常検知, 分子構造の生成等, 幅広い分野への活用が期待されている. 本稿では, 通常のデータの代わりに, 量子状態を量子コンピュータにより生成する機械学習モデルについて概説する. 特に損失関数として最適輸送距離を用いて, 比較的低次元に量子状態が局在する場合の学習手法を示す.

(キーワード): 量子コンピュータ, 量子アルゴリズム, 学習理論, 生成モデル, 異常検知

1 はじめに

近年の量子コンピュータのハードウェアの進展に伴い, 将来の活用に向けた量子アルゴリズム開発が急速に進展している. 特に, 量子コンピュータを用いた機械学習 (量子機械学習) は近い将来の量子コンピュータである NISQ (Noisy Intermediate-Scale Quantum Computer) において活用が有力視されている適用領域の一つである¹⁻⁴⁾.

量子機械学習は, 教師データとして量子データを扱うものと, 古典データを扱うものに大別される. 古典データは文書や画像データのような, 日常的に既存のコンピュータで取り扱っているデータである. 一方で, 量子データは, ミクロな物理現象等の量子状態をデータとして取り扱うものである. 特に量子機械学習では, 量子データを用いた場合, 既存のコンピュータを用いた機械学習 (古典機械学習) と比較して大きな優位性を有することが期待されている. 量子データを用いた機械学習では, 例えば, 分光実験などで得られる光を直接量子コンピュータで処理することで, 既存の機械学習より遥かに少ない実験回数で, 高精度な原子構造や物性値の計測を行うこと

などが期待される. 少量の実験回数による高精度な予測は, 網羅的に材料探索を行うハイスループット実験で必要不可欠な要素であり, 今後, 量子実験技術やロボティクスとの組合せなどにより, さらなる発展が望まれる.

機械学習の手法は, その目的に応じて, 回帰, 分類, 生成等の種類に分別される. 近年, 特に, 古典データの生成に関して, 例えば, 画像生成⁵⁻⁷⁾, 異常検知⁸⁾, 分子構造⁹⁾の生成等において, 高い性能を示す学習モデルが提案されており, 大きな注目を集めている. 量子データにおいても, 生成モデルによる量子状態の異常検知や量子状態の生成は, 将来的に量子コンピュータ, 量子通信, 量子計測などにおいて, 量子状態を精密に制御するための重要な要素技術の一つに成ると考えられる. 例えば, 量子通信において, 送られる量子状態は周辺の環境の影響により大きくノイズを受ける可能性があり, ノイズを受けた量子状態を検知し, 除去することは, 実用化に向けた大きな鍵となると考えられる.

本稿では, 量子状態を生成する生成モデルの新たなフレームワーク¹⁰⁾を紹介する. これまでの量子状態の生成モデルでは, 主に単一の状態を学習するも

ⁱ サイエンスソリューション部 デジタルエンジニアリングチーム 上席主任コンサルタント 博士 (理学)

ⁱⁱ ソニーグループ株式会社 R&D センター 先端研究部

ⁱⁱⁱ 慶應義塾大学 量子コンピューティングセンター 共同研究員

^{iv} 慶應義塾大学 量物理情報工学科 教授

のであったため、適用先には大きな制限があった。本稿の手法では、高次元の全ヒルベルト空間中の比較的低次元の部分空間上のみ量子状態が分布すると仮定することで、量子状態の分布を学習出来るようなフレームワークを提供する。

本稿の構成は以下のとおりである。まず 2 章において、生成モデルや最適輸送損失の基礎的な事項について簡単に述べる。次に、3 章において、最適輸送損失を用いた量子状態の分布の学習アルゴリズムについて述べる。4 章では、生成モデルの応用例として、量子状態の異常検知アルゴリズムについて述べる。最後に、5 章において、まとめと今後の展望を述べる。

2 Preliminaries

本章では、まず、古典機械学習における生成モデルの問題設定をレビューし、暗黙的な生成モデルの量子モデルについて述べる。その後、生成モデルの学習におけるコスト関数の有力な候補の一つである、最適輸送損失について述べる。最後に、ビット数が多い場合の学習において問題となる勾配消失問題について簡単に述べ、解決策の一つであるローカルコスト関数を紹介する。

2.1 暗黙的な生成モデル

生成モデルは、教師なし学習の主要なタスクの一つであり、与えられたデータセット $\{\mathbf{x}_i\}_{i=1}^{N_r}$ をもとに、そのデータの背後にある分布を近似することを目的とする。より具体的には、与えられたデータセットの背後にある確率分布 $\alpha(\mathbf{x})$ を仮定し、確率モデルを θ でパラメトライズされた確率分布 $\beta_\theta(\mathbf{x})$ としたときに、適切な損失関数 $\mathcal{L}(\alpha, \beta, \theta)$ をなるべく小さくするようにパラメータ θ を学習する。古典機械学習において、生成モデルは異常検知、データの生成、データ構造の学習、次元削減などの目的で広く研究されているが、量子コンピュータへの適用事例はほとんどない。

生成モデルは大きく、明示的モデル (prescribed model) と暗黙的モデル (implicit model) に分類される。明示的モデルでは、パラメータ θ を用いて、データセット $\{\mathbf{x}_j\}_{j=1}^{N_r}$ の背後にある確率密度 $\beta_\theta(\mathbf{x})$ を関数として明示的に記述する。このため、明示的モデルでは、対数尤度を計算することが可能であり、最尤推定などにより、パラメータ θ を求めることが可能となる。一方で、暗黙的モデルでは、確率密度を明示的

に記述せずに、生成モデルに従うサンプルを生成する。暗黙的モデルの利点としては、広大な全標本空間のうち、比較的低次元の超平面上にデータセットが分布するような現象を容易に表すことができること、データを生成する物理的なプロセスとして解釈できること等がある。暗黙的モデルの代表的な例としては、や Generative Adversarial Networks¹¹⁾などが挙げられる。本稿では、暗黙的モデルに注目して議論をすすめる。

暗黙的モデルでは、多くの場合、全標本空間のうち、データは比較的低次元の多様体上に存在すると仮定する。つまり、標本空間 \mathcal{X} の次元 N_x に比べて十分次元が低い次元 $N_z (\ll N_x)$ の潜在空間 \mathcal{Z} 上のランダム潜在変数 $\mathbf{z} = \{z_i\}_{i=1}^{N_z}$ を仮定し、潜在変数 \mathbf{z} を標本空間に移す複雑な写像 (Generator $G_\theta(\mathbf{z}) \in \mathcal{X}$) として確率モデルを表す (図 1 を参照)。ここで、ランダム変数 \mathbf{z} はガウス分布や一様分布等の既知の分布 $\gamma(\mathbf{z})$ に従うと仮定される。暗黙的モデルでは、適当なコスト関数を用いて、Generator のパラメータ θ をデータに適合させることで学習を行う。

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\hat{\alpha}_{N_r}, \hat{\beta}_{\theta, N_g}) \quad (1)$$

ここで、 $\hat{\beta}_{\theta, N_g}$ は push-forward operator #¹²⁾を通じて標本空間上に定義された確率分布 $G_\theta \# \hat{\gamma}_{N_g}$ を表しており、大まかに言えば、潜在変数 $\mathbf{z} \in \mathcal{Z}$ に対して定義されている確率分布 $\hat{\gamma}$ を Generator G_θ を通じて標本空間 \mathcal{X} 上の確率分布に移したものである。 $\hat{\alpha}_{N_r}(\mathbf{x}), \hat{\gamma}_{N_g}(\mathbf{z})$ はそれぞれ確率分布 $\alpha(\mathbf{x})$ 及び $\gamma(\mathbf{z})$ からデータ $\{\mathbf{x}_j\}_{j=1}^{N_r}$ 及び $\{\mathbf{z}_j\}_{j=1}^{N_g}$ を用いて定義される経験分布である：

$$\hat{\alpha}_{N_r}(\mathbf{x}) = \frac{1}{N_r} \sum_{i=1}^{N_r} \delta(\mathbf{x} - \mathbf{x}_i) \quad (2)$$

$$\hat{\gamma}_{N_g}(\mathbf{z}) = \frac{1}{N_g} \sum_{i=1}^{N_g} \delta(\mathbf{z} - \mathbf{z}_i) \quad (3)$$

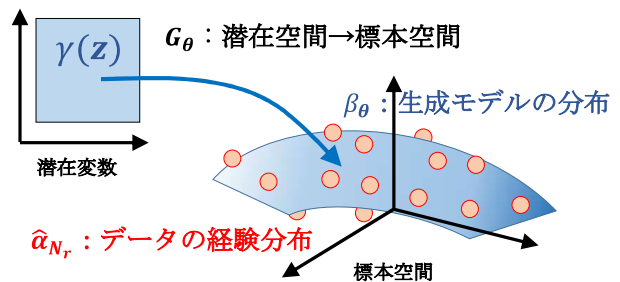


図 1 暗黙的な生成モデルのイメージ. 低次元の潜在空間上の確率変数を既知の分布 γ に従って発生, Generator $G_\theta(\mathbf{z})$ により標本空間に写像する. 写像した分布がデータセットに近づくようにパラメータ θ の学習を行う.

2.2 暗黙的な生成モデル

量子コンピュータを用いて, 量子状態の暗黙的な生成モデルを構築するにあたり, 量子データの Generator として, パラメータ付きの量子回路を用いることが考えられる. 例えば, 量子コンピュータを用いた量子 GAN¹³⁾では, パラメータ付きの量子回路 $U(\theta)$ を用いて, 以下のように暗黙的な生成モデルを定義することが提案されている.

$$|\psi_\theta(\mathbf{z})\rangle = U(\theta)|\mathbf{z}\rangle \quad (4)$$

ここで, $|\psi_\theta(\mathbf{z})\rangle$ は量子データの分布を表している. この状態を量子データの分布として用いるとすると, 潜在変数 \mathbf{z} の値がわずかに異なる量子データ同士が直交することとなるが, これは, 連続的な量子空間上の分布を表すことが困難であり, 現在の目的と合わない.

そこで, 本稿では, 以下のように, 潜在変数 \mathbf{z} をパラメータ付き量子回路に含めた形で, 量子状態の暗黙的な生成モデルを定めることとする.

$$|\psi_\theta(\mathbf{z})\rangle = U(\mathbf{z}, \theta)|\mathbf{0}\rangle \quad (5)$$

この形には, 量子空間上の連続的な状態を表すことが出来るということ以外にも, パラメータシフトルール^{14),15)}を使うことで, 潜在変数 \mathbf{z} に関する微分を計算できるという利点がある. 潜在変数 \mathbf{z} に関する微分を活用する例は, 生成モデルを用いた異常検知の例(4章)で示す.

パラメータ付き量子回路の具体例としては, 例えば, 図 2 の形が考えられる. 図 2 では, 「 N_L 」は,

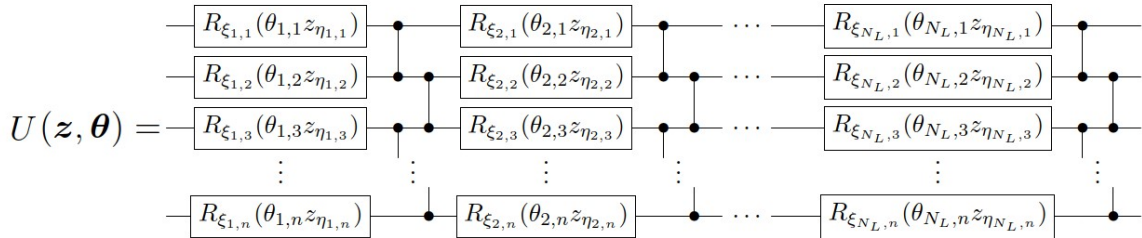


図 2 Generator $U(\mathbf{z}, \theta)$ に用いるパラメータ付き量子回路の例. この回路では, 同じ構造を持つ層を繰り返し N_L 回作用する. 各層では, 全ビットにランダムな方向の単一ビットパウリ回転をし, その後, 全ビットに対する ladder controlled Z を作用する.

Generator $U(\mathbf{z}, \theta)$ を構成する階層の数である. この Generator $U(\mathbf{z}, \theta)$ では, 同じ構造を持つ層を N_L 回繰り返し作用する.

$$U(\mathbf{z}, \theta) = \prod_{\ell=1}^{N_L} W V_{\xi_\ell, \eta_\ell}(\mathbf{z}, \theta_\ell) \quad (6)$$

ここで, $\theta_l = \{\theta_{l,j}\}_{j=1}^n$, $\xi_l = \{\xi_{l,j}\}_{j=1}^n$ 及び $\eta_l = \{\eta_{l,j}\}_{j=1}^n$ は, すべて n 次元のパラメータであり, $\theta = \{\theta_\ell\}_{\ell=1}^{N_L}$, $\xi = \{\xi_\ell\}_{\ell=1}^{N_L}$ 及び $\eta = \{\eta_\ell\}_{\ell=1}^{N_L}$ である. 式(6)のパラメータのうち, θ は学習する対象のパラメータであり, \mathbf{z} は潜在変数である. ξ 及び η は学習が始まる前にランダムに初期化され, 学習中には変化しないパラメータである.

ℓ 番目の層内ではまず, 演算子 $V_{\xi_\ell, \eta_\ell}(\mathbf{z}, \theta_\ell)$ により, j 番目のビットに角度 $\theta_{\ell,j} \times z_{\eta_{\ell,j}}$, 方向 $\xi_{\ell,j}$ の単一量子ビットのパウリ回転を作用し, その後, 演算子 W により全ビットに対する ladder controlled Z ゲートを作用する. すなわち,

$$W = \prod_{i=1}^{n-1} CZ_{i,i+1} \quad (7)$$

$$V_{\xi_\ell, \eta_\ell}(\mathbf{z}, \theta_\ell) = \prod_{i=1}^{n-1} R_{\xi_{\ell,i}}(\theta_{\ell,i} z_{\eta_{\ell,i}}) \quad (8)$$

ここで, $CZ_{i,i+1}$ は, i 番目及び $i+1$ 番目のビットに作用する controlled Z ゲートである. また, $R_{\xi_{\ell,i}}(\theta_{\ell,i} z_{\eta_{\ell,i}})$ は, 角度 $\theta_{\ell,i} \times z_{\eta_{\ell,i}}$, 方向 $\xi_{\ell,i}$ のパウリ回転である. 方向 $\xi_{\ell,i} \in \{X, Y, Z\}$ 及び潜在変数の成分番号 $\eta_{\ell,i} \in \{0, 1, 2, \dots, N_z\}$ は学習が始まる前にランダムに初期化される. ここで, 便利のため, バイアス項 z_0 を導入した.

2.3 最適輸送損失

既存の機械学習における生成モデルでは、損失関数として最適輸送損失/ワッサーシュタイン距離 (Optimal Transportation loss/Wasserstein distance) が注目されており、画像解析, 自然言語処理, 金融等様々な分野で活用されている¹⁶⁾⁻²⁰⁾. 最適輸送損失は、特に、尤度が計算できない場合や、2つの確率分布のサポート同士が一致しない場合にも適用可能なこと、標本空間上の距離を自然に取り込めること等の利点²⁰⁾⁻²⁶⁾がある. 最適輸送損失は、確率分布 $\alpha(\mathbf{x})$ を別の確率分布 $\beta(\mathbf{x})$ に移す最小のコストとして、以下で定義される.

定義1 (最適輸送損失²⁷⁾)

$$\begin{aligned} \mathcal{L}_c(\alpha, \beta) &= \min_{\pi} \int c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) \\ \text{subject to } &\int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{x} = \beta(\mathbf{y}), \\ &\int \pi(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \alpha(\mathbf{x}), \\ &\pi(\mathbf{x}, \mathbf{y}) \geq 0 \end{aligned} \quad (9)$$

ここで、 $c(\mathbf{x}, \mathbf{y}) \geq 0$ は、データ \mathbf{x} から \mathbf{y} への“輸送コスト”を表す $\mathcal{X} \times \mathcal{X}$ 上の関数であり、ground cost と呼ばれる. また、式(9)を最小化する coupling π の集合を最適輸送計画と呼ぶ. 一般には最適輸送損失は、確率分布の間の“距離”の性質を満たさないが、以下のように、ground cost を距離から構成することで、確率分布の間の距離を構成できる.

定義2 (p-Wasserstein 距離²⁸⁾) ground cost $c(\mathbf{x}, \mathbf{y})$ が距離の性質を満たす関数 $d(\mathbf{x}, \mathbf{y})$ を用いて、 $c(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathbf{y})^p$ で表される時、p-Wasserstein 距離 $W_p(\alpha, \beta)$ は以下で定義される.

$$W_p(\alpha, \beta) = \mathcal{L}_{d_p}(\alpha, \beta)^{1/p} \quad (10)$$

このように定義される p-Wasserstein 距離 $W_p(\alpha, \beta)$ は、距離の性質を満たす、すなわち、任意の確率分布 α, β, γ に対して、正值性 ($W_p(\alpha, \beta) \geq 0$)、対称性 ($W_p(\alpha, \beta) = W_p(\beta, \alpha)$)、非退化性 ($W_p(\alpha, \beta) = 0 \iff \alpha = \beta$)、三角不等式 ($W_p(\alpha, \gamma) \leq W_p(\alpha, \beta) + W_p(\beta, \gamma)$)を満たす.

実際に、暗黙的な生成モデルの損失関数として、最適輸送損失を使う状況においては、データの背後にある分布 α や Generator の生成する分布 β_θ を直接扱えるわけではない. その代わりに、通常は以下のように、式(2)および式(3)の経験分布を用いて、最適輸送損失の近似値を求める.

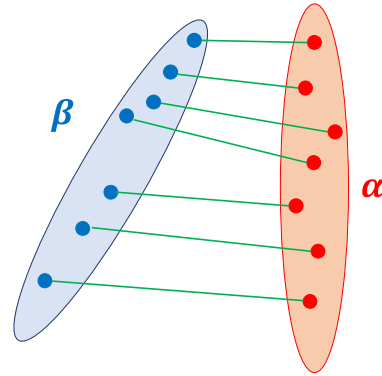


図3 最適輸送距離の empirical estimator のイメージ. α からサンプルした点を β からサンプルした点に移すとしたときの“距離”の総和として定義される.

定義3 (最適輸送損失の empirical estimator²⁸⁾)

$$\begin{aligned} \mathcal{L}_c(\alpha, \beta_\theta) &\simeq \mathcal{L}_c(\hat{\alpha}_{N_r}, \hat{\beta}_{\theta, N_g}) \\ &= \min_{\{\pi_{i,j}\}_{i,j=1}^{N_r, N_g}} \sum_{i,j=1}^{N_r, N_g} c(\mathbf{x}_i, G_\theta(\mathbf{z}_j)) \pi_{i,j}, \\ \text{subject to } &\sum_{i=1}^{N_r} \pi_{i,j} = \frac{1}{N_g}, \sum_{j=1}^{N_g} \pi_{i,j} = \frac{1}{N_r}, \\ &\pi_{i,j} \geq 0 \end{aligned} \quad (11)$$

この empirical estimator (図3を参照)は、サンプル数が無限大極限 ($N_r, N_g \rightarrow \infty$) で、 $\mathcal{L}_c(\hat{\alpha}_{N_r}, \hat{\beta}_{\theta, N_g}) \rightarrow \mathcal{L}_c(\alpha, \beta_\theta)$ と収束することが知られている. ただしこの収束速度は一般には非常に遅い²⁹⁾. ただし、定義2の p-Wasserstein 距離の empirical estimator に関しては、以下の収束法則が成り立つことが知られている³⁰⁾.

定理4 確率分布 α の upper Wasserstein 次元 $d^*(\alpha)$ (文献³⁰⁾の Definition 4)に対して、もし $s > d^*(\alpha)$ であれば、以下の関係式が成り立つ.

$$\mathbb{E}[W_p(\alpha, \hat{\alpha})] \leq O\left(N_r^{-\frac{1}{s}}\right) \quad (12)$$

ここで、upper Wasserstein 次元 $d^*(\alpha)$ は、大まかにいえば、確率分布 α のサポートの次元(データの存在する次元)である. p-Wasserstein 距離が距離の性質を満たすことを使うと、以下の補題が直ちに導かれる:

補題5 確率分布 α, β_θ の upper Wasserstein 次元がともに $d^*(\alpha)$ であるとすると、もし $s > d^*(\alpha)$ であれば、以下の関係式が成り立つ.

$$\mathbb{E}[|W_p(\alpha, \beta_\theta) - |W_p(\hat{\alpha}, \hat{\beta}_\theta)|]| \leq O\left(N_r^{-\frac{1}{s}}\right) \quad (13)$$

この補題により、データの内部次元や潜在次元 N_z が十分小さければ、empirical estimator (式(11))は最適輸送損失の精度のよい推定値となることがわかる。

2.4 変分量子回路の勾配消失問題

最適輸送損失を量子状態の生成モデルの損失関数として採用するにあたり、ある量子状態 $|\psi\rangle$ をある量子状態 $|\phi\rangle$ に輸送する際の ground cost $c(|\psi\rangle, |\phi\rangle)$ を何にするかを検討する必要がある。量子状態の輸送の最も単純なものとしては、量子状態間の距離であるトレース距離が考えられる。

定義 6 (純粋状態間のトレース距離³¹⁾) 量子状態 $|\psi\rangle$ と $|\phi\rangle$ の間のトレース距離は以下のように定義される。

$$c_{tr}(|\psi\rangle, |\phi\rangle) = \sqrt{1 - |\langle\psi|\phi\rangle|^2} \quad (14)$$

トレース距離 $c_{tr}(|\psi\rangle, |\phi\rangle)$ は距離の性質を満たすため、これを用いることにより、前章で示した収束速度(定理4)のような利便な性質を持つ p-Wasserstein 距離を定義することが出来る。また、トレース距離は、swap テスト³²⁾や inversion テスト³³⁾等により、量子コンピュータを用いて容易に算出することが可能である。

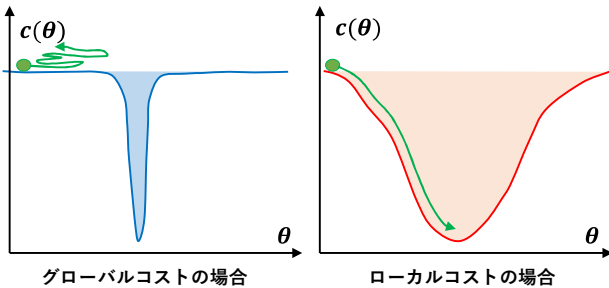


図 4 グローバルコストの場合のコスト関数は、ビット数が多くなると、ほとんどの θ でコスト関数 $c(\theta)$ が平坦であり、微分をもとにした最適化が困難である。ローカルコストでは、この問題が解決する可能性があることが知られている。

一方で、トレース距離は、グローバル測定(全部のビットを測定しないと得られない物理量)により得られる量であるため、勾配消失による次元の呪いを受けることが知られている^{34),35)}。すなわち、ランダムなパラメータ θ を持つ量子回路 $U(\theta)$ を用いて作成した量子状態 $|\phi_\theta\rangle = U(\theta)|0\rangle$ を、ターゲットの量子状態 $|\psi\rangle$ に近づけるように学習を行うときに、もしトレース距離をコスト関数として学習すると、微分の期待

値がビット数 n に対して指数関数的に0に近づくことと成る(図4)。これは、ビット数 n が大きい場合には、ほとんどの θ の値において、コスト関数の勾配の値がほぼ0となり、学習が困難になるということを示している。

勾配消失問題を回避する方法として、コスト関数を局所的な測定のみから構成する方法が知られている^{34),35)}。状態 $|\phi_\theta\rangle = U(\theta)|0\rangle$ の学習における局所的な測定のみから構成されるコスト関数としては、以下の式が提案されている。

定義 7 (局所的な測定のみから構成されるローカルコスト関数^{36),37)})

$$c_{\text{local}}(|\psi\rangle, |\phi\rangle) = \sqrt{\frac{1}{n} \sum_{k=1}^n (1 - p^{(k)})} \quad (15)$$

$$p^{(k)} = \text{Tr}[P_0^k U^\dagger(\theta)|\psi\rangle\langle\psi|U(\theta)]$$

$$P_0^k = I_1 \otimes I_2 \otimes \dots \otimes \overbrace{|0\rangle\langle 0|}_k \otimes \dots \otimes I_n$$

ここで、 n はビット数であり、 I_i は i 番目のビットに対する恒等演算子を表している、すなわち $p^{(k)}$ は k 番目のビットを観測した際に、0が出力される確率を表している。式(14)のトレース距離 $c_{tr}(|\psi\rangle, |\phi\rangle)$ は、距離の性質を満たす一方で式(10)のローカルコスト関数 $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ は、距離の性質を満たさない(対称性及び三角不等式を満たさない)。このため、ローカルコスト関数 $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ を ground cost として用いた場合の最適輸送損失 $\mathcal{L}_{c_{\text{local}}}$ は距離の性質を満たさない。ただし、以下のように、 $\mathcal{L}_{c_{\text{local}}}$ は、ダイバージェンスとしての性質を有することを示すことが出来る。

命題 8 ground cost $c(x, y)$ が以下の性質を満たすとする。

$$c(x, y) \geq 0, \quad (16)$$

$$c(x, y) = 0 \text{ iff } x = y$$

この性質を持つ $c(x, y)$ を ground cost とした最適輸送損失 $\mathcal{L}_c(\alpha, \beta)$ は、以下の性質を満たす。

$$\mathcal{L}_c(\alpha, \beta) \geq 0, \quad (17)$$

$$\mathcal{L}_c(\alpha, \beta) = 0 \text{ iff } \alpha = \beta$$

この命題から、ローカルコスト $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ を ground cost として用いた最適輸送損失 $\mathcal{L}_{c_{\text{local}}}$ は、ダイバージェンスの性質を満たすため、確率分布の比

較に用いることが出来ることがわかる。

以下、この最適輸送損失 $\mathcal{L}_{\text{local}}$ を使った量子状態の確率分布の学習アルゴリズム及びそれを利用した異常検知のアルゴリズムを紹介する。

3 量子状態の分布の学習アルゴリズム

ここで、対象とする量子状態の分布の学習は、 N_r 個からなる量子データセット $\{|\psi\rangle\}_{i=1}^{N_r}$ から、その量子データの背後にある分布やデータの隠れた構造を表す生成モデルを学習することである。ここで各量子データは有限個のコピー（例えば k 個のコピー $\{|\psi\rangle^{\otimes k}\}_{i=1}^{N_r}$ ）が用意できることを仮定する。もし、 $k \rightarrow \infty$ であれば、トモグラフィ等により、完全にすべての成分を確定することが出来るため、既存の機械学習モデルに帰着することに留意する³⁸⁾。

量子コンピュータを用いた暗黙的な生成モデルとしては、式(5)のモデルを採用することとする。式(5)の生成モデルを、最適輸送損失を用いて、なるべくデータセット $\{|\psi\rangle\}_{i=1}^{N_r}$ に近くなるようにパラメータ θ の学習を行う。学習アルゴリズムのフローチャートを図 5 に示す。

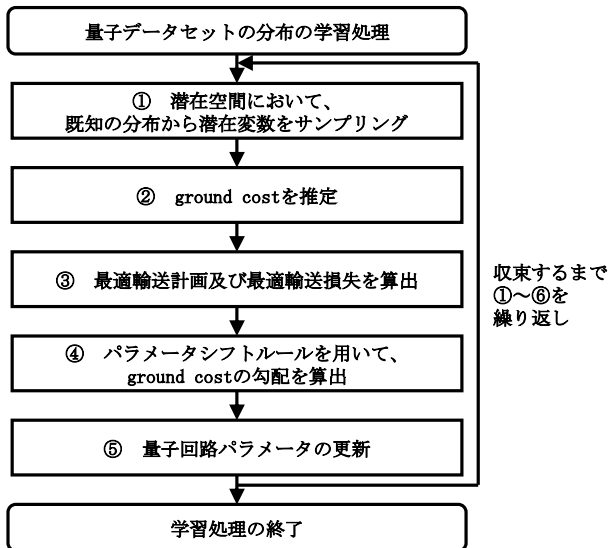


図 5 最適輸送損失を利用した量子データセットの学習アルゴリズムのフローチャート

このアルゴリズムでは、まず、既存のコンピュータを用いて、既知の分布から潜在変数を生成する（ステップ①）。具体的には、サンプル数 N_g 個の潜在変数のサンプル $\{\mathbf{z}_j\}_{j=1}^{N_g}$ を生成するものとする。

次に、ステップ②においては、ground cost の推定値の算出を行う。ここでは、量子コンピュータを用い

て、量子回路のパラメータ θ と潜在変数のサンプル $\{\mathbf{z}_j\}_{j=1}^{N_g}$ とから、量子状態のサンプル $\{U(\theta, \mathbf{z}_j)|0\rangle\}_{j=1}^{N_g}$ を生成する。次に、量子コンピュータにより、サンプル $\{U(\theta, \mathbf{z}_i)|0\rangle\}_{i=1}^{N_g}$ と、量子データセット $\{|\psi\rangle\}_{i=1}^{N_r}$ とから、ground cost c_{local} を算出する。これには、式(15)を使用する、すなわち、以下の量の計算を実施する：

$$c_{\text{local}}(|\psi_i\rangle, U(\theta, \mathbf{z}_j)|0\rangle) = \sqrt{\frac{1}{n} \sum_{k=1}^n (1 - p_{i,j}^{(k)})} \quad (18)$$

$$p_{i,j}^{(k)} = \text{Tr}[P_0^k U^\dagger(\theta, \mathbf{z}_j) |\psi_i\rangle \langle \psi_i| U(\theta, \mathbf{z}_j)]$$

$$P_0^k = I_1 \otimes I_2 \otimes \dots \otimes \overbrace{|0\rangle\langle 0|}_k \otimes \dots \otimes I_n$$

この ground cost c_{local} は量子状態のコピーが無制限個準備出来るときに算出可能な量であり、実際には、 N_s ショット（同一の状態を用意する個数）によりこの ground cost の近似値 $\{\tilde{c}_{\text{local},i,j}^{(N_s)}\}_{i,j=1}^{N_r, N_g}$ を算出する。ここで、

「 N_r 」は量子データセットのデータ数、「 N_g 」は潜在変数のサンプル数、「 N_s 」はショット数である。この ground cost の近似値は、確率 $(1 - p_{i,j}^{(k)})$ のベルヌーイ分布 $\text{Be}(1 - p_{i,j}^{(k)})$ に従う確率変数 $X_{i,j,k}^{(s)}$ を用いて以下のように算出される。

$$X_{i,j,k}^{(s)} (s = 1, 2, \dots, N_s) \sim \text{Be}(1 - p_{i,j}^{(k)})$$

$$\tilde{c}_{\text{local},i,j}^{(N_s)} = \sqrt{\frac{1}{n} \sum_{k=1}^n \frac{1}{N_s} \sum_{s=1}^{N_s} X_{i,j,k}^{(s)}} \quad (19)$$

次に、ステップ③において、得られた ground cost の近似値 $\{\tilde{c}_{\text{local},i,j}^{(N_s)}\}_{i,j=1}^{N_r, N_g}$ を用いて、既存のコンピュータにより、以下の線形計画問題を解くことで、量子回路 $U(\theta, \mathbf{z}_j)$ のパラメータ θ を固定したときの、最適輸送計画及び最適輸送損失の算出を行なう：

$$\mathcal{L}_{\tilde{c}_{\text{local}}^{(N_s)}} \left(\{|\psi\rangle_i\}_{i=1}^{N_r}, \{U(\theta, \mathbf{z}_j)|0\rangle\}_{j=1}^{N_g} \right)$$

$$= \min_{\{\pi_{i,j}\}_{i,j=1}^{N_r, N_g}} \sum_{i,j=1}^{N_r, N_g} \tilde{c}_{\text{local},i,j}^{(N_s)} \pi_{i,j} \quad (20)$$

subject to

$$\sum_{i=1}^{N_r} \sum_{j=1}^{N_g} \pi_{i,j} = \frac{1}{N_g}$$

$$\sum_{j=1}^{N_g} \pi_{i,j} = \frac{1}{N_r}, \quad \pi_{i,j} \geq 0$$

ここで $\mathcal{L}_{\text{local}}^{(N_s)}$ は、最適輸送損失、 $\pi_{i,j}$ は、データセットの i 番目のデータ $|\psi\rangle_i$ とサンプル $U(\boldsymbol{\theta}, \mathbf{z}_j)|0\rangle$ の間の輸送計画である。

次に、ステップ④では、量子コンピュータを用いて、ground cost $\tilde{c}_{\text{local},i,j}^{(N_s)}$ の勾配の算出を行なう。ここでは、例えば、パラメータシフトルール^{14),15)}を用いる。 N_s ショットにより得られる、ground cost の勾配の近似値を $\left\{ \frac{\partial}{\partial \theta_k} \tilde{c}_{\text{local},i,j}^{(N_s)} \right\}_{i,j,k=1}^{N_r, N_g, N_p}$ と表記する。ここで、「 N_p 」は、パラメータ「 $\boldsymbol{\theta}$ 」のパラメータ数である。

次に、ステップ⑤において、既存のコンピュータにより、ground cost の勾配 $\left\{ \frac{\partial}{\partial \theta_k} \tilde{c}_{\text{local},i,j}^{(N_s)} \right\}_{i,j,k=1}^{N_r, N_g, N_p}$ 及び最適輸送計画 $\{\pi_{i,j}\}_{i,j=1}^{N_r, N_g}$ を使って、最適輸送損失の勾配 $\left\{ \frac{\partial}{\partial \theta_k} \mathcal{L}_{\text{local}} \right\}_{k=1}^{N_p}$ を計算する。

最後に、ステップ⑥において、計算した勾配を用いて、パラメータ $\boldsymbol{\theta}$ をアップデートする。

収束するまで、以上のステップ①からステップ⑥の処理を繰り返し、得られたパラメータを $\bar{\boldsymbol{\theta}}$ とすると、最終的に最適輸送損失を（局所的に）最小化するような量子状態の分布 $U(\mathbf{z}, \bar{\boldsymbol{\theta}})|0\rangle$ を得ることが出来る。

本学習アルゴリズムの性能の理論的・数値的な評価については、筆者らの最近の論文¹⁰⁾に示されている。

4 量子状態の異常検知アルゴリズム

次に、量子状態の生成モデルの応用例として、量子状態の異常検知の例を示す。異常検知は、過去のデータ \mathbf{x}_i ($i = 1, 2, \dots, M$) から得られた情報を基に、新たなデータ $\mathbf{x}^{(R)}$ が異常（レアデータ）か正常データかを判定する問題ある。過去のデータのほとんどは正常データであり、異常データと正常データ間に大きな偏りがあるという点が、通常の二値分類問題と大きく異なる点であり、分類問題とは異なる手法が発展してきた³⁹⁾。

量子コンピュータを始め、量子通信、量子計測などの量子技術では、量子状態を精密に制御することが必要となる。一方で、量子状態は非常に繊細であり、周辺環境の僅かな変化により、大きくノイズを受ける。この異常を検知し、ノイズを受けた状態を除去することは、量子技術の実用化に向けた重要な鍵である。これまでの量子異常検知手法は主に、量子状態を観測して得られる古典的なデータをもとにして、異常検知を実行する^{40),41)}。一方で、本手法では、量子状態に対し

て、直接量子情報処理を実施することで、異常検知を行う。

一般に、異常検知は以下の流れで実施される⁴²⁾。

- ① (Distribution estimation): (主に) 正常なデータをもとに、正常データの確率分布モデルを構成する。
- ② (Anomaly score design): 得られた確率分布モデルをもとに、新たなデータの異常度 (Anomaly Score) を定義する。
- ③ (Threshold determination): 異常度がある値より大きい場合に異常、と判定できるような閾値を設定する。

今回は、上記の3ステップのうち、ステップ①を前述のアルゴリズムで学習する。ステップ③の閾値は問題に応じて、運用時に適切な値を決める必要がある。ステップ②に関して、ここでは、古典機械学習のAnoGAN⁴³⁾と同様の方法により異常検知を行う。すなわち、新たなデータ $|\psi^{(R)}\rangle$ と、学習した生成モデルから得られるサンプル $U(\mathbf{z}, \bar{\boldsymbol{\theta}})|0\rangle$ をもとに損失関数 $\mathcal{L}(U(\mathbf{z}, \bar{\boldsymbol{\theta}})|0\rangle, |\psi^{(R)}\rangle)$ を定義し、その損失関数の潜在変数 \mathbf{z} に関する最小値を異常度として用いる:

$$\begin{aligned} (\text{anomaly score}) \\ = \min_{\mathbf{z}} \mathcal{L}(U(\mathbf{z}, \bar{\boldsymbol{\theta}})|0\rangle, |\psi^{(R)}\rangle) \end{aligned} \quad (21)$$

損失関数としては、例えば、式(14)のトレース距離や式(15)のローカルコストを使用することが考えられる。式(21)において、 \mathbf{z} に関する最小化は、 \mathbf{z} に関する微分を使って勾配法等により実施可能である。 \mathbf{z} に関する微分は $\boldsymbol{\theta}$ に関する微分と同様に、パラメータシフトルールを用いて実施可能である。

具体的な異常検知アルゴリズムを図6に示す。

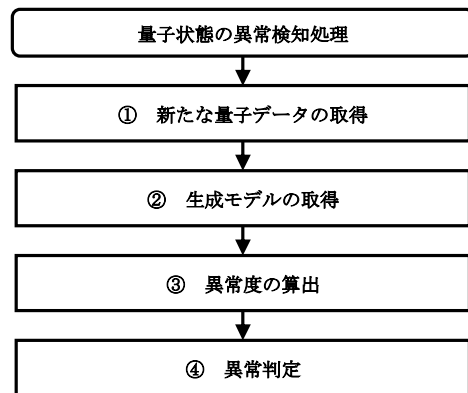


図6 量子状態の異常検知処理のフローチャート

このアルゴリズムでは、まずステップ①において、新たな量子データの取得を行なう。具体的には、学習に用いたデータセットとは別の新たな量子データ $|\psi^{(R)}\rangle$ を取得する。この新たな量子データ $|\psi^{(R)}\rangle$ としては、観測により外部から入力されるデータや、量子計算により出力される量子状態等が考えられる。

次に、ステップ②では、通常時（正常時）の量子データセット（教師データセット）を用いて、3章の方法により予め学習した生成モデル $U(\mathbf{z}, \bar{\theta})|0\rangle$ を取得する。

次に、ステップ③において、異常度の算出を行なう。例えば、 $U(\mathbf{z}, \bar{\theta})|0\rangle$ を用いて出力される量子状態の内、新たな量子データに最も近いデータ（式(21)を最小化する \mathbf{z} の値）を探索する。新たな量子データと、最も近いデータとの距離として異常度を定める。

最後にステップ④において、算出した異常度と、予め定めた異常判定値とを比較し、比較結果に基づいて、異常の有無を判定する。

この異常検知アルゴリズムを用いた実際の数値実験の結果やその他の応用例については筆者らの最近の論文¹⁰⁾に示す。

5 おわりに

本稿では、将来の量子時代に向けて、量子データを用いた生成モデルのアルゴリズムの提案を行った。最適輸送損失を用いた学習モデルを定義するとともに、応用先として、量子状態の異常検知の応用例を示した。

本稿の手法は量子コンピュータを用いた生成モデルの研究の第一ステップであり、今後さらなる発展が期待される。量子データを用いた生成モデルでは、実験等により得られる量子状態を教師データにすることにより、既存のコンピュータを用いた生成モデルに比べて、遥かに少ない実験回数で学習出来ると期待されている。

今回の提案では量子状態を個別に処理するようなアルゴリズムを考えたが、今後の方向性として、量子状態へのコヒーレントアクセスを許す場合²⁾や量子メモリを使う場合⁴⁾への拡張することにより、更に学習に必要な量子データの個数を減らすことが期待される。その他、online learning や混合状態学習⁴⁴⁾への拡張も必要不可欠である。

また、既存の機械学習において、より高次元のデータに対して有効なことが知られている、Sinkhorn

Distance 等の Entropy Regularized Wasserstein Distance⁴⁵⁾⁻⁴⁹⁾への拡張も重要な課題である。ただし、量子機械学習において、勾配消失を避けるためにローカルコストを ground cost として使う場合には、ground cost が距離の定義を満たさないため、これまで提案されている Sinkhorn divergence 等をそのまま使うだけでは、divergence の性質を満たさず、確率分布を比べる指標として不適切となる。例えば指数カネールなどにより半正定値性を回復した上で、Sinkhorn divergence を適用する等の処方箋が考えられる。

そのほか、本手法の応用例のさらなる探索も必要不可欠である。応用例としては例えば、以下のような方向性が期待される。

- ✓ 外部から入力された量子状態の異常検知(量子コンピュータの異常検知, 重力波検知, 量子センシングなど)
- ✓ 量子コンピュータの計算結果の学習(過去のVQE結果の波動関数の学習・初期値の設定, Born machineの波動関数の転移学習, 量子シミュレーションのサロゲートモデルなど)
- ✓ 古典データを埋め込んだ量子データ(例えば近似的な振幅埋め込み等による埋め込んだ状態を使った異常検知など)

引用文献

- 1) Maria S. & Francesco P.: Machine learning with quantum computers (Springer 2021).
- 2) Dorit Aharonov, Jordan Cotler, Xiao-Liang Qi. Quantum algorithmic measurement, *Nature Communications*, 13(1) (2022) 1–9,.
- 3) Yusen W. & Bujiao W. & Jingbo W. & Xiao Y.: Provable advantage in quantum phase learning via quantum kernel alphasatron, *arXiv:2111.07553*, (2021).
- 4) Hsin-Yuan H. et al.: Quantum advantage in learning from experiments, *arXiv:2112.00778*, (2021)
- 5) Jianmin B. & Dong C. & Fang W. & Houqiang L. & Gang H.: Cvae-gan: fine-grained image generation through asymmetric training, *In Proceedings of the IEEE international conference on computer vision*, (2017), 2745–2754.
- 6) Andrew B. & Jeff D. & Karen S.: Large scale gan training for high fidelity natural image synthesis, *arXiv:1809.11096* (2018).
- 7) Tejas D K. & William F W. & Pushmeet K. & Josh T.:

- Deep convolutional inverse graphics network, *Advances in neural information processing systems*, 28, (2015).
- 8) Chong Z. & Randy C P.: Anomaly detection with robust deep autoencoders, *In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, (2017) 665–674.
- 9) Rafael G-B & Jennifer N W. & David D. & Jose M. H.-L. & Benjamin S.-L. & Dennis S. & Jorge A.-I. & Timothy D H. & Ryan P A. & Alan A.-G.: Automatic chemical design using a data-driven continuous representation of molecules, *ACS central science*, 4(2), (2018), 268–276.
- 10) Hiroyuki T. & Shumpei U. & Naoki Y.: Quantum Generative model with Optimal Transport, XXXXX, (2022).
- 11) Ian G. & Jean P.-A. & Mehdi M. & Bing X. & David W.-F. & Sherjil O. & Aaron C. & Yoshua B.: Generative adversarial nets, *Advances in neural information processing systems*, 27, (2014).
- 12) Gabriel P. & Marco C.: Computational Optimal Transport: With Applications to Data Science (Foundations and Trends in Machine Learning), (2019).
- 13) Pierre-L. D.-D. & Nathan K.: Quantum generative adversarial networks, *Physical Review A*, 98(1), (2018), 012324.
- 14) Kosuke M. & Makoto N. & Masahiro K. & Keisuke F.: Quantum circuit learning, *Physical Review A*, 98(3), (2018), 032309.
- 15) Maria S. & Ville B. & Christian G. & Josh I. & Nathan K.: Evaluating analytic gradients on quantum hardware, *Physical Review A*, 99(3), (2019), 032331.
- 16) Danilo R. & Shakir M.: Variational inference with normalizing flows, *In International conference on machine learning*, (2015), 1530–1538.
- 17) Gabriel P. & Marco C.: Computational Optimal Transport: With Applications to Data Science (Foundations and Trends in Machine Learning), (2019).
- 18) Yann O. & Pajot H. & Cedric V. editors.: Optimal Transport: Theory and Applications. London Mathematical Society Lecture Note Series, (Cambridge University Press, 2014).
- 19) Filippo S.: Optimal transport for applied mathematicians, *Birkauer, NY*, 55(58-63), (2015), 94.
- 20) Gabriel P. & Marco C.: Information and Inference special issue on optimal transport in data sciences. Information and Inference, *A Journal of the IMA*, 8(4) (2019), 655–656, 12.
- 21) Gregoire M. & Klaus-R. M. & Marco C.: Wasserstein training of restricted boltzmann machines, *Advances in Neural Information Processing Systems*, 29, (2016).
- 22) Espen B. & Pierre E J. & Mathieu G. & Christian P R.: Inference in generative models using the Wasserstein distance. *arXiv:1701.05146*, (2017).
- 23) Martin A. & Soumith C. & Leon B.: Wasserstein generative adversarial networks. *In International conference on machine learning*, (2017), 214–223.
- 24) Ilya T. & Olivier B. & Sylvain G. & Bernhard S.: Wasserstein auto-encoders. *arXiv:1711.01558*, (2017).
- 25) Aude G. & Gabriel P. & Marco C.: GAN and VAE from an optimal transport point of view, *arXiv:1706.01807*, (2017).
- 26) Olivier B. & Sylvain G. & Ilya T. & Carl-J. S.-G. & Bernhard S.: From optimal transport to generative modeling: the VEGAN cookbook. *arXiv:1705.07642*, (2017).
- 27) Leonid V K.: On the translocation of masses, *In Dokl. Akad. Nauk. USSR (NS)*, 37, (1942), 199–201.
- 28) Cedric V.: Optimal transport: old and new, 338, (Springer, 2009).
- 29) Richard M. D.: The speed of mean Glivenko-Cantelli convergence. *The Annals of Mathematical Statistics*, 40(1), (1969), 40–50.
- 30) Jonathan W. & Francis B.: Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance, *Bernoulli*, 25(4A), (2019), 2620–2648.
- 31) Michael A. N. & Isaac L. C.: Quantum Computation and Quantum Information *Cambridge Series on Information and the Natural Sciences*, (Cambridge University Press, 2000).
- 32) Harry B. & Richard C. & John W. & Ronald D. W.: Quantum fingerprinting, *Physical Review Letters*, 87(16), (2001), 167902.
- 33) Vojtech H. & Antonio D C. & Kristan T. & Aram W H. & Abhinav K. & Jerry M C. & Jay M G.: Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747), (2019), 209–212
- 34) Kouhei N. & Naoki Y.: Expressibility of the alternating layered ansatz for quantum computation, *Quantum*,

- 5:434, (2021).
- 35) Marco C. & Akira S. & Tyler V. & Lukasz C. & Patrick J. C.: Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nature communications*, 12(1), (2021), 1–12.
- 36) Sumeet K. & Ryan L. & Alexander P. & Lukasz C. & Andrew T. S. & Patrick J. C.: Quantum-assisted quantum compiling. *Quantum*, 3, (2019), 140.
- 37) Kunal S. & Sumeet K. & Marco C. & Patrick J. C.: Noise resilience of variational quantum compiling. *New Journal of Physics*, 22(4), (2020), 043006.
- 38) Esma A. & Gilles B. & Sebastien G.: Machine learning in a quantum world. *In Conference of the Canadian Society for Computational Studies of Intelligence*, (Springer, 2006), 431–442.
- 39) Varun C. & Arindam B. & Vipin K.: Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3), (2009), 1–58.
- 40) Satoshi H. & Takafumi O. & Ryo O. & Takashi W. & Shigeki T.: Anomaly detection in reconstructed quantum states using a machine-learning technique, *Physical Review A*, 89(2), (2014), 022104.
- 41) Satoshi H. & Takafumi O. & Ryo O. & Takashi W. & Shigeki T.: Quantum-state anomaly detection for arbitrary errors using a machine-learning technique. *Physical Review A*, 94(4), (2016), 042341.
- 42) Tsuyoshi I.: Introduction to anomaly detection using machine learning – a practical guide with R (in Japanese), (Corona Publishing, 2015).
- 43) Thomas S. & Philipp S. & Sebastian M. W. & Ursula S.-E. & Georg L.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. *In International conference on information processing in medical imaging*, (Springer, 2017), 146–157.
- 44) Nic E. & Elliott M. B. & Aliza U. S. & Mark M. W. & Andrew T. S. & Patrick J. C. & Zoë H.: Quantum Mixed State Compiling, arXiv:2209.00528, (2022).
- 45) Marco C.: Sinkhorn distances: Lightspeed computation of optimal transport, *Advances in neural information processing systems*, 26, (2013), 2292–2300.
- 46) Jean F. & Thibault S. & Francois-X. V. & Shun-ichi A. & Alain T. & Gabriel P.: Interpolating between optimal transport and MMD using Sinkhorn divergences, arXiv:1810.08278, (2018).
- 47) Aude G. & Lenaic C. & Francis B. & Marco C. & Gabriel P.: Sample complexity of Sinkhorn divergences, *In The 22nd International Conference on Artificial Intelligence and Statistics*, (2019), 1574–1583.
- 48) Shun-ichi A. & Naotsugu T. & Masafumi O.: Geometry of information integration. *In Information Geometry and its Applications IV*, (Springer, 2016), 3–17.
- 49) Shun-ichi A. & Ryo K. & Masafumi O.: Information geometry connecting Wasserstein distance and Kullback–Leibler divergence via the entropy-relaxed transportation problem, *Information Geometry*, 1(1), (2018), 13–37.
- 50) Kohei N. et al.: Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators, *Phys. Rev. Research* 4, (2022), 023136.