Using The Quantum Wasserstein Distance of Order 1 for Variationally Learning Quantum Error Correcting Codes

Quynh T. Nguyen

Department of Physics & Department of EECS,

Massachusetts Institute of Technology,

77 Massachusetts Ave., Cambridge, MA 02139

(Dated: May 19, 2021)

Abstract

We attempt to use the quantum Wasserstein distance of order 1 (quantum W1) as a loss metrics for variationally learning quantum error correcting codes. The quantum W1 distance metrics is recently proposed by De Palma et al. [1] and is a generalization of the classical W1 distance. It holds some desirable properties that make it a promising candidate loss metrics in mitigating barren plateaus in variational algorithms. First, the quantum W1 distance metrics recovers the classical Hamming distance when states are pure and in canonical basis. Second, as opposed to quantum fidelity which is unitarily invariant, the quantum W1 distance is continuous and non unitarily invariant. Third, it is additive under composition. The advantages of quantum W1 distance have been exploited Quantum Wasserstein Generative Adversarial Networks (QWGAN) [2], where a generator circuit learns to generate a target state and to fool a discriminator who tries to distinguish between the generated state and the actual state. Our work focuses on extending this QWGAN model into the problem of finding quantum codes. First, we review the classical and quantum Wassertstein distances and propose an efficient method to estimate the quantum W1 distance. We then demonstrate that given a sufficiently powerful ansatz encoding circuit, the $[[5, 1, 3]]$ code can be learned by via maximizing the quantum W1 distance between orthogonal encoded states. Next, we discuss the failures of our model in a more generic case and investigate the flaws in our proposed model. Finally, we outline future approaches that we want to take in order to improve our model.

I. INTRODUCTION

Many proposed algorithms for noisy intermediate-scale quantum (NISQ) devices involve training a parameterized quantum circuit via classical optimization [3–5], where the trainable quantum circuits are often chosen as intial guesses to explore the Hilbert space of interest. These variational hybrid algorithms, however, suffer from a trainability issue called barren plateaus, that is, the gradient of objective functions vanishes exponentially with the number of qubits. The barren plateaus emerge from at least two places: 1. the 2-design characteristics of common ansatz circuits [6], and 2. the inefficient choice of objective function in terms of global observables [7]. The latter can be interpreted as that the currently common metrics for measuring distance between two density matrices, which include trace distance, fidelity, and quantum relative entropy, are not capable of capturing the local difference. For example, in certain cases, it is desirable to quantify $|0^n\rangle$ to be closer to $|0^{n-1}\rangle \otimes |1\rangle$ than to $|1^n\rangle$, whereas all the aforementioned metrics are unitarily invariant and thus the distance between any orthogonal states is always maximal.

A promising candidate to mitigate this problem is the quantum Wasserstein distance of order 1 (quantum W1 distance) [1], which holds two main desirable local properties: it recovers them Hamming distance when states are in canonical basis and it is additive under system composition. Moreover, the quantum W1 distance is also a quantum generalization of the classical Wasserstein-1 distance, which is commonly known in machine learning community as the earth mover distance. Indeed, the earth mover distance has proved advantageous in improving the training stability of machine learning generative models such as the Wasserstein Generative Adversarial Networks (WGAN) [8]. The quantum counterpart of this model, QWGAN, has also been proposed recently by Kiani et al. [2], by which our work is greatly inspired. Particularly, we utilize the property of recovering Hamming distance in the task of learning quantum codes.

In section II, we review the classical Wassertstein distance. Next, we summarize the formulation of the quantum W1 distance and propose a method of approximating it in section III. In section IV, we demonstrate that given a sufficiently powerful ansatz encoding circuit, the $[[5, 1, 3]]$ code can be learned by via maximizing the quantum W1 distance between orthogonal encoded states. Then, we discuss the failures of our model in a more generic case and investigate the flaws in our proposed model. Finally, we outline future

approaches that we want to take in order to improve our model in section V.

II. THE CLASSICAL WASSERTSTEIN DISTANCE

The Wasserstein distance is a method of measuring the similarity between two probability distributions that have emerged recently as a prominent distance metrics in unsupervised machine learning models such as GANs [9]. They arise from transportation theory and have proved more advantageous than other metrics such as Kullback-Leibler divergence (relative entropy), Jensen-Shannon divergence.

Definition 1. *(Wasserstein-α distance, primal formulation) Given a metric space $(\mathcal{X}, D)$ and two probability distributions μ,ν over $\mathcal{X}$. For any $\alpha \geq 1$, the Wasserstein-α distance or the Wasserstein distance of order α between μ and ν is defined as*

$$W_\alpha(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{X}} D(x,y)^\alpha \, \mathrm{d}\gamma(x,y) \right)^{1/\alpha}, \tag{1}$$

where $\Gamma(\mu, \nu)$ denotes the set of all joint distributions over $\mathcal{X} \times \mathcal{X}$ whose marginal distribution are μ and ν.

Intuitively, for $\alpha = 1$, the Wasserstein-1 (W1) distance can be understood as the optimal cost of transporting mass in order to transform μ to ν, hence it is also referred to as the Earth mover's distance and is widely used in machine learning. However, this primal formulation is highly intractable. In practice, one often utilizes its dual formulation to approximate the W1 distance via a family of 1-Lipschitz functions.

Definition 2. *(Wassertstein-1 distance dual formulation (Kantorovich-Rubinstein theorem)) If $(\mathcal{X}, D)$ is compact, we have*

$$W_1(\mu, \nu) = \sup \left\{ \int_{\mathcal{X}} f \, \mathrm{d}\mu - \int_{\mathcal{X}} f \, \mathrm{d}v : \|f\|_L \leq 1 \right\}. \tag{2}$$

It is infeasible to calculate exactly W1 distance using the above definition. Fortunately, one heuristically find that a good parameterized family of f whose Lipschitz constant is upperbounded by constant is sufficient to estimate the W1 distance as well as enables the backprop algorithm.

III. THE QUANTUM WASSERTSTEIN-1 DISTANCE

Defining the quantum counterpart of W1 distance is of interest because density matrices are nothing but probability distributions over quantum states. Thus, it is reasonable to conjecture that the quantum W1 distance can help alleviate these issues as does its classical counterpart. In this section we summarize the formalism of the quantum W1 distance that is recently proposed by De Palma et al. [1].

Let $\{|1\rangle, \ldots, |d\rangle\}$ be the canonical basis of $\mathbb{C}^d$, and $\mathcal{H}_n = \left(\mathbb{C}^d\right)^{\otimes n}$ be the Hilbert space of n qudits. We denote by $\mathcal{O}_n$ the set of the self-adjoint linear operators on $\mathcal{H}_n$, by $\mathcal{O}_n^T \subset \mathcal{O}_n$ the subset of the traceless self-adjoint linear operators on $\mathcal{H}_n$, by $\mathcal{S}_n \in \mathcal{O}_n$ the set of the quantum states of $\mathcal{H}_n$. For any $X \in \mathcal{O}_n$, let $\|X\|_p$ be its Schatten p-norm.

In short, the quantum W1 distance is defined such that the distance between any pair of neighboring states is at most one.

Definition 3. *(Neighboring quantum states) Two states ρ, $\sigma \in \mathcal{S}_n$ are neighboring if they coincide after discarding one qudit, i.e. $\mathrm{Tr}_i \rho = \mathrm{Tr}_i \sigma$ for some $i \in [n]$. The set of differences between pair of neighboring states $\mathcal{N} \in \mathcal{O}^T$ is defined as*

$$\mathcal{N}_n = \bigcup_{i=1}^{n} \mathcal{N}_n^{(i)}, \quad \mathcal{N}_n^{(i)} = \{\rho - \sigma : \rho, \sigma \in \mathcal{S}_n, \mathrm{Tr}_i \rho = \mathrm{Tr}_i \sigma\}, \quad i \in [n]. \tag{3}$$

Definition 4. *(Quantum W1 norm) For any $X \in \mathcal{O}_n^T$, the quantum W1 norm, denoted by $\|\cdot\|_{W_1}$, is defined as*

$$\|X\|_{W_1} = \frac{1}{2} \min \left(\sum_{i=1}^{n} \left\|X^{(i)}\right\|_1 : X^{(i)} \in \mathcal{O}_n^T, \mathrm{Tr}_i X^{(i)} = 0, X = \sum_{i=1}^{n} X^{(i)} \right). \tag{4}$$

The full derivation can be found in Definitions (5), (6), Propositions (1), (6), and Lemma (1) in the original paper. In a nutshell, the convex hull of $\mathcal{N}_n$ is shown to be a ball which is subsequently defined to be the unit ball of a unique norm called the quantum W1 norm. This will enforce that the W1 distance between any two neighboring states is at most one. Indeed, in Eq.(4), if we take $X = \rho - \sigma$ where $\mathrm{Tr}_i \rho = \mathrm{Tr}_i \sigma$, then

$$\|X\|_{W_1} \leq \frac{1}{2} \|\rho - \sigma\|_1 \leq \frac{1}{2} \left(\|\rho\|_1 + \|\sigma\|_1\right) = 1. \tag{5}$$

The quantum W1 distance is naturally induced from (4) and, similarly to its classical counterpart, has its primal and dual formulations.

Definition 5. *(Quantum W1 distance, primal formulation) For any pair of n-qudit density matrices ρ, σ, the quantum W1 distance between them is*

$$\begin{aligned} W_1(\rho,\sigma) &= \|\rho-\sigma\|_{W_1} \\ &= \min\left(\sum_{i=1}^{n} c_i : c_i \geq 0, \rho-\sigma = \sum_{i=1}^{n} c_i\left(\rho^{(i)}-\sigma^{(i)}\right), \rho^{(i)},\sigma^{(i)} \in \mathcal{S}_n, \mathrm{Tr}_i\,\rho^{(i)} = \mathrm{Tr}_i\,\sigma^{(i)}\right). \end{aligned} \tag{6}$$

The quantum W1 distance exhibits two desirable properties as discussed in section I. First, they are additive under composition ([1], Proposition 3). Given a composite system AB, we have

$$\left\|\rho^A\otimes\rho^B-\sigma^A\otimes\sigma^B\right\|_{W_1} = \left\|\rho^A-\sigma^A\right\|_{W_1} + \left\|\rho^B-\sigma^B\right\|_{W_1}. \tag{7}$$

Second, it recovers the Hamming distance when ρ,σ are pure states in computational basis ([1], Proposition 5), i.e.

$$W_1(|x\rangle\langle x|, |y\rangle\langle y|) = h(x,y) \qquad \text{for } x,y\in[d]^n. \tag{8}$$

These two propeties are key in mitigating barren plateaus. Thus, if efficiently implemented, quantum W1 distance can be a better distance metrics in problems where one care about local changes in states. However, this formulation is an intractable semidefinite program and we, again, need to approximate the quantum W1 distance via its dual formulation, where one can utilize the Lipschitz operators (now quantum), as done in the classical case. De Palma et al. have derived the dual program as follows.

Definition 6. *(Quantum W1 distance, dual formulation)*

$$W_1(\rho,\sigma) = \max\left(\mathrm{Tr}[H(\rho-\sigma)] : H\in\mathcal{O}_n, \|H\|_L \leq 1\right), \tag{9}$$

where $\|\cdot\|_L$ denotes the quantum Lipschitz constant.

Definition 7. *(Quantum Lipschitz constant)*

$$\begin{aligned} \|H\|_L &= \max\left(\mathrm{Tr}[HX] : X\in\mathcal{N}_n\right) \\ &= \max_{i\in[n]}\left(\max\left(\mathrm{Tr}[H(\rho-\sigma)] : \rho,\sigma\in\mathcal{S}_n, \mathrm{Tr}_i\,\rho = \mathrm{Tr}_i\,\sigma\right)\right). \end{aligned} \tag{10}$$

Definitions 6 and 7 together allow one to approximate the quantum W1 distance via a family of parameterized H whose quantum Lipschitz constant is upperbounded by one. Naturally, H can be chosen to be a weighted sum of local Pauli operators, i.e. each Pauli acts on up to some number k of qubits, where the weights are constrained so as $\|H\|_L \leq 1$. We propose a way to approximating the quantum W1 distance, along same lines with [2], utilizing Proposition (10) in [1] which goes as follows.

Proposition 1. *(Upperbound to the quantum Lipschitz constant in terms of local structure). Let $H = \sum_{\mathcal{I}\subseteq[n]} H_{\mathcal{I}}$, where $H_{\mathcal{I}}$ acts non trivially on qubits in $\mathcal{I}$, then*

$$\|H\|_L \leq 2 \max_{i\in[n]} \left\| \sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_{\mathcal{I}} \right\|_\infty \tag{11}$$

Proof. For any $\rho, \sigma \in \mathcal{S}_n$ such that $\mathrm{Tr}_i\, \rho = \mathrm{Tr}_i\, \sigma$ we have

$$\mathrm{Tr}[H(\rho-\sigma)] = \mathrm{Tr}\left[\sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_{\mathcal{I}} (\rho-\sigma) \right] \leq 2 \left\| \sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_{\mathcal{I}} \right\|_\infty , \tag{12}$$

where the first equality follows from $\mathrm{Tr}_i\, \rho = \mathrm{Tr}_i\, \sigma$, and the inequality are essentially an application of Holder's inequality, after noting that $\|\rho-\sigma\|_1 \leq 2$. Thus the claim holds due to Definition 7. □

The operator norm on the RHS of (12) can be further upperbounded by grouping terms that act X, Y, Z on the same qubit,

$$\begin{aligned}
\left\| \sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_{\mathcal{I}} \right\|_\infty &= \left\| \sum_{P_{\bar{i}}} P_{\bar{i}} \otimes \left(c^X_{P_{\bar{i}}} X^i + c^Y_{P_{\bar{i}}} Y^i + c^Z_{P_{\bar{i}}} Z^i \right) \right\|_\infty \\
&\leq \sum_{P_{\bar{i}}} \left\| P_{\bar{i}} \otimes \left(c^X_{P_{\bar{i}}} X^i + c^Y_{P_{\bar{i}}} Y^i + c^Z_{P_{\bar{i}}} Z^i \right) \right\|_\infty \\
&= \sum_{P_{\bar{i}}} \|P_{\bar{i}}\|_\infty \left\| \left(c^X_{P_{\bar{i}}} X^i + c^Y_{P_{\bar{i}}} Y^i + c^Z_{P_{\bar{i}}} Z^i \right) \right\|_\infty \\
&= \sum_{P_{\bar{i}}} \sqrt{ \left(c^X_{P_{\bar{i}}} \right)^2 + \left(c^Y_{P_{\bar{i}}} \right)^2 + \left(c^Z_{P_{\bar{i}}} \right)^2 },
\end{aligned} \tag{13}$$

where the sum is over $P_{\bar{i}}$ who does not act on the i-th qubit. The final equality holds because $\|P_{\bar{i}}\|_\infty = 1$ as it is a Pauli operator, and the equality $\|c_X\sigma_X + c_Y\sigma_Y + c_Z\sigma_Z\|_\infty =$

$\sqrt{c_X^2 + c_Y^2 + c_Z^2}$. In summary, one performs the following steps in order to estimate the quantum W1 distance between two states ρ and σ:

1. Measure the expectations $\mathrm{Tr}\left[(\rho - \sigma)H_{\mathcal{I}}\right]$ for any $H_{\mathcal{I}}$ in the family (for example, all Pauli operators that act on up to k qubits)
2. Maximize $\sum_{\mathcal{I}} c_{\mathcal{I}} \mathrm{Tr}\left[(\rho - \sigma)H_{\mathcal{I}}\right]$, subject to $\sum_{P_{\bar{i}}} \sqrt{\left(c^X_{P_{\bar{i}}}\right)^2 + \left(c^Y_{P_{\bar{i}}}\right)^2 + \left(c^Z_{P_{\bar{i}}}\right)^2} \leq 1/2$ for any qubit i. This guarantees $\|H\|_L \leq 1$ due to (13) and Proposition 1.

Note that Kiani et al.[2] estimate the quantum W1 distance via a linear program, which they obtain by directly applying a triangle inequality on the entire RHS of (12), while our approach leads to a quadratic (convex) program. [2] also go one step further by using "two-local" Pauli operators, i.e. decomposing H into a weighted sum of Pauli operators whose support are only on up to two qubits

$$H = \sum_{\mathcal{I} \subseteq [n]} H_{\mathcal{I}} = \sum_{\mathcal{I} \subseteq \{0,1,2,3\}^n,\ h(\mathcal{I}) \leq 2} c_{\mathcal{I}} \sigma_{\mathcal{I}}, \tag{14}$$

and propose quantum Wasserstein GANs that optimize variational circuits to generate a target state, where they define the ansatz circuit as the generator and the parameterized weighted set of two-local Pauli operators as the discriminator. The generator and discriminator play a minimax game where in each turn the generator tries to adjust its parameters (via backprop) to minimize the W1 distance, while the discriminator adjust its weights (via solving a linear program) to maximize the distance. They also empirically demonstrate, with their particular set of examples, that the gradient of the (estimated) quantum W1 distance is robust up to at least 14 qubits, while that of fidelity distance is exponentially decaying.

IV. LEARNING CODES WITH TARGET RATE USING W1 DISTANCE

We demonstrate the usage of quantum W1 distance in a rather different setup than in [2]. Given $[[n, k]]$, our goal is to find a code of optimal distance d that protects against the Pauli errors. First, let us note that the learned codes in this section are not likely to be practical since we do not take into account correctability. However, we believe that our model can also be used in learning a decoding circuit. The correctability is taken into account only with a fully end-to-end encoder-noise-decoder model as done in QVECTOR algorithm [5].

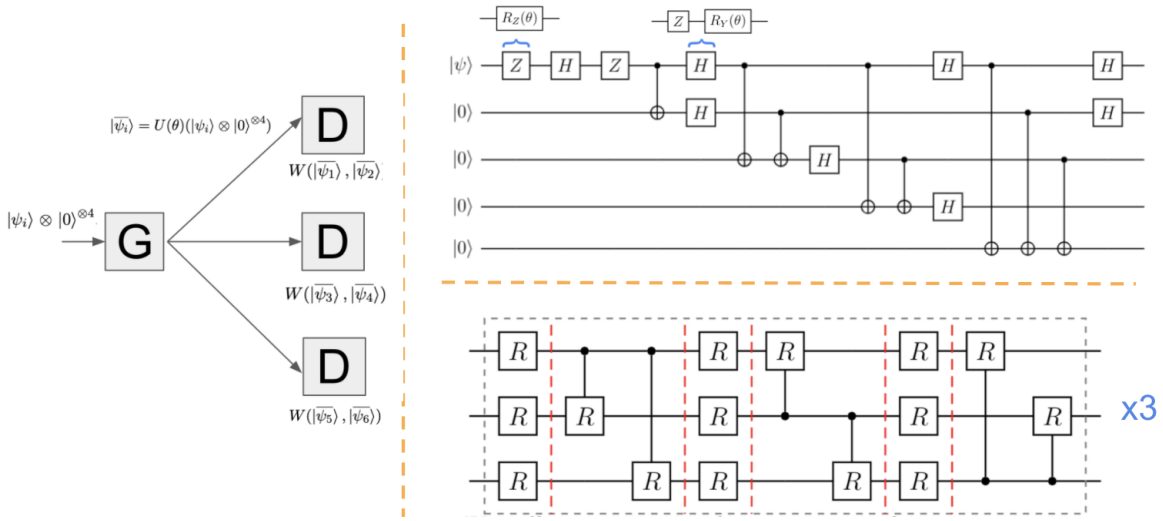

FIG. 1: Left: schematic diagram of our setup. The encoder/generator (G) takes in the qubit, which comes in three bases X, Y, Z, and four additional ancilla qubits, while the three discriminators (D) distinguish between orthogonal pairs of output states of each basis. Top right: the proper encoding circuit of $[[5, 1, 3]]$ code, in our toy experiment, we paramerterize the Z gates and decompose each H into a Z gate followed by a R_Y rotation gate. Bottom right: the generic ansatz circuit we adapt from [5], consisting of alternating general 1-qubit rotations and controlled 1-qubit rotations.

Particularly, in this section we want to train a parameterized circuit (encoder) to generate a code that represents 1 logical qubits. We consider the case $n = 5$ and recall that the optimal distance for $[[5, 1]]$ code is 3 and that an encoding circuit for a $[[5, 1, 3]]$ code is given in Figure 1. Thanks to the small number of qubits, we are able to calculate the exact quantum W1 distance between codewords of the $[[5, 1, 3]]$ by solving the semidefinite program in its primal formulation (Definition 5). We find that the distance between *any* pair of orthogonal states in the codespace is indeed three. In the next paragraphs, we attempt to answer the reverse question: armed with the quantum W1 distance, is it possible to search for a code that has a (nearly) optimal distance 3 that corrects 1 arbitrary Pauli errors?

The encoder circuit, denoted as $U(\theta)$, is tasked with searching for a 1-logical-qubit codespace, in which any pair of orthogonal states, in *any basis*, is 3 apart according to quantum W1 metrics. Inspired by [2], we will refer to the set of Pauli measurements needed for estimating quantum W1 distance as the discriminator. The task of discriminators is to adjust their weights to find an operator H whose Lipschitz constant is bounded by 1 that

gives the maximum estimated quantum W1 distance: $\max\{\mathrm{Tr}[H(\rho-\sigma)], \|H\|_L \leq 1\}$. To enforce that the distance between *any* orthogonal logical states is 3, it is natural to consider the logical states in three orthogonal bases X, Y, Z. Let

$$|\psi_1\rangle = |0\rangle, |\psi_2\rangle = |1\rangle, |\psi_3\rangle = |+\rangle, |\psi_4\rangle = |-\rangle, |\psi_5\rangle = \frac{|0\rangle + i|1\rangle}{\sqrt{2}}, |\psi_6\rangle = \frac{|0\rangle - i|1\rangle}{\sqrt{2}}, \tag{15}$$

and

$$\left|\overline{\psi_i}\right\rangle = U(\theta)(|\psi\rangle \otimes |0\rangle^{\otimes 4}). \tag{16}$$

Our objective is to find θ^* the maximizes the objective function

$$J(\theta) = \min\{W_1(\left|\overline{\psi_1}\right\rangle, \left|\overline{\psi_2}\right\rangle), W_1(\left|\overline{\psi_3}\right\rangle, \left|\overline{\psi_4}\right\rangle), W_1(\left|\overline{\psi_5}\right\rangle, \left|\overline{\psi_6}\right\rangle)\}. \tag{17}$$

If the estimated W1 distance is good enough and the objective function converges, we should expect $J(\theta^*) \approx 3$ at the end. Figure 1 depicts a schematic diagram of our setup as well as the circuits we use. Note that there are 3 discriminators because we are estimating 3 quantum W1 distances and for each of them the optimal weights of local Pauli operators are not necessarily the same. At each step, our algorithm does the following:

1. The generator outputs 6 states $\left|\overline{\psi_i}\right\rangle$ for $i = 1, ..., 6$.

2. The discriminators find the approximated distances of the 3 pairs.

3. The generator adjusts its parameters to improve the minimum distance.

After training, we check if the code distance is 3 by verifying the Knill-Laflamme conditions. That is,

$$\left\langle \overline{\psi_i} \right| \sigma_\alpha \sigma_\beta \left| \overline{\psi_j} \right\rangle \tag{18}$$

is small for any $\alpha, \beta \in [n]$, $\sigma \in I, X, Y, Z$, and any orthogonal $\left|\overline{\psi_i}\right\rangle$, $\left|\overline{\psi_j}\right\rangle$.

Before we run the simulations, we need to choose a family of Lipschitz operators in order to approximate the W1 distance. Kiani et al. [2] use two-local Pauli operators, which would not work in our case. Indeed, the K-L condititions (18) imply that, on any pair of orthogonal states $(\left|\overline{\phi_i}\right\rangle, \left|\overline{\phi_j}\right\rangle)$ in the optimal codespace, we have $\mathrm{Tr}[\sigma_\alpha\sigma_\beta(\left|\overline{\phi_i}\right\rangle \left\langle\overline{\phi_i}\right| - \left|\overline{\phi_j}\right\rangle \left\langle\overline{\phi_j}\right|)]$. This can be easily seen by requiring (18) to hold for any pair of orthogonal states spanned by $\{\left|\overline{\phi_i}\right\rangle, \left|\overline{\phi_j}\right\rangle\}$. Therefore, the distance estimation on perfect codes would be zero if we use only 2-local Pauli operators, and it would be infeasible to hope that maximizing (17) could

lead to a perfect code. We generalize this observation and conclude that in order to estimate distance in a codespace of distance d, one has to use at least d-local Pauli operators. Thus, knowing that the target code has distance 3, we estimate the quantum W1 distances using 3-local Pauli operators in our simulations.

The ansatz circuits we use are depicted in Figure (1). We first start with a simply parameterized version of the original 5-qubit encoding circuit (top right of Fig. (1)). Throughout this paper, we perform simulations of our algorithm using PennyLane, a cross-platform Python library for differentiable programming of quantum computers [10]. For each simulation, our algorithm runs for 800 training steps, using TensorFlow's Adam optimizer. We performed 100 runs, each with random initialization of parameters, and found that the algorithm succeeds in finding a perfect $[[5, 1, 3]]$ code in about 20 times, where we define success as when the K-L conditions are satisfied after training. A successful training is shown in Figure (2), where we obtain a code that satisfies the K-L conditions after training (depicted in the right plot). We also verify that the exact W1 distance between any learned orthogonal codewords is 3. Our result once again confirms that the W1 distance in all X, Y, Z bases being optimal is a *necessary* condition for K-L criteria. However, we note that our objective function, the estimated W1 distance shown on the left of Fig. (2), converges to a value who is much lower than 3. This clearly shows that our method is highly underestimating the W1 distances. Unfortunately, we were not able to improve our estimation due to time limitations of this project, we note here three main steps in our method that need improvement in our future work. First, the family of 1-Lipschitz operators could be improved, this is perhaps straight yet most difficult improvement to make and will require a more thorough understanding of the quantum 1-Lipschitz constant space. Second, in Proposition (1), we utilize Holder's inequality and triangle inequality on $(\rho - \sigma)$, this can be improved by using a tighter bound. Third, we use a giant triangle inequality in (13) to upperbound the operator norm $\left\| \sum_{\mathcal{I} \subseteq [n]: i \in \mathcal{I}} H_{\mathcal{I}} \right\|_{\infty}$ (note that this step exists only because we use Holder's inequality, improvements on it might not be needed if we use some other bounds that do not involve operator norm).

The above list of limitations are not exhaustive, indeed, the results above are not actually a "learning" procedure because we have done nothing but parameterizing the original encoding circuit itself. In general, we need a generic ansatz circuit that can be generalized to arbitrary number of qubits. In our next simulations, we use a circuit depicted

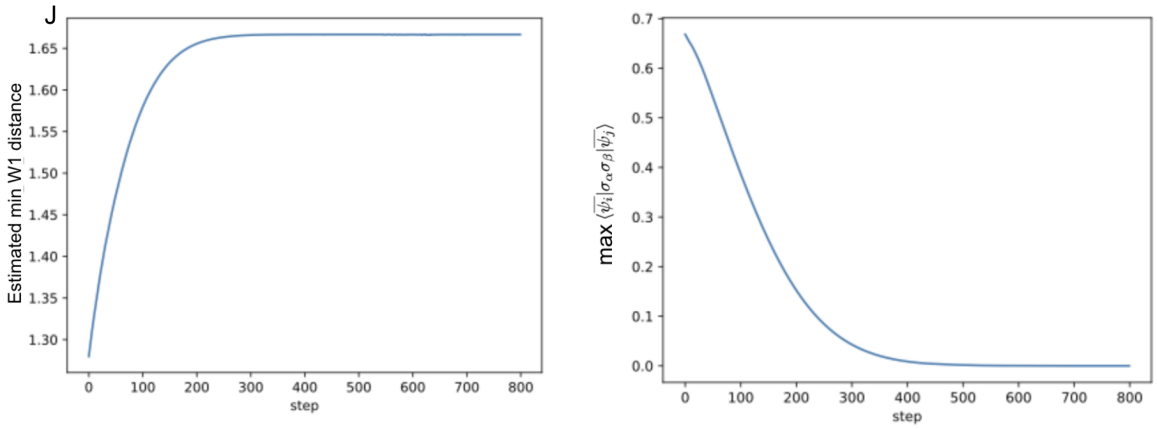

FIG. 2: Training curves in learning the $[[5,1,3]]$ code using the simply parameterized circuit (top right of Fig. (1)). The left plot shows the convergence of the estimated W1 distance. The right plot verifies that the Knill-Laflamme conditions (18) are satisified as the (estimated) W1 distances converge to their optimal value.

in the bottom right of Fig. (1), which consists of alternating general 1-qubit rotations and controlled 1-qubit rotations. We have verified that this circuit is indeed capable of generating the $[[5,1,3,]]$ code via our simulations. However, we found a more interesting results. That is, there are circuit parameters such that all the (exact) distances $W_1(|\overline{\psi_1}\rangle, |\overline{\psi_2}\rangle), W_1(|\overline{\psi_3}\rangle, |\overline{\psi_4}\rangle), W_1(|\overline{\psi_5}\rangle, |\overline{\psi_6}\rangle)$ are greater than 3 and yet the K-L conditions are unsatisfied. This means that the condition of W1 distances of these three pairs of orthogonal states being optimal is not sufficient to guarantee the distance is also 3 in any other basis. Therefore, we conclude that our proposed objective function does not suffice to learn perfect codes. We do not know which/how many bases on which the optimal distance condition is enforced are sufficient, or, even worse, whether or not such set of bases exists. In order to answer these questions, further work on the geometry of the quantum W1 norm is required. In the discussion section, we instead propose a method via which one can make the conditions in only 3 bases X, Y, Z sufficient, via the complementary channel "frame."

V. DISCUSSION

In summary, we review the quantum Wasserstein-1 distance metrics and its two most significant advantages, including its additive property under composition and the recovery

of classical Hamming distance on pure states in computational basis. We then propose a method to approximate it via its dual formulation. We point out places that need improvement in our method of estimating the quantum W1 distance, namely the Lipschitz constant bounding step. Finally, we propose an algorithm utilizing it that learns a code of optimal distance, given the number of physical qubits and logical qubits. Our algorithm seems to work in our toy five-qubit code simulation, but its objective function turns out to fail to guarantee K-L conditions in a more general case. Furthermore, our discussion in section IV implies that these difficulties are not likely to be overcome until a more thorough understanding of the quantum W1 distance is available. However, as we wrap up this project report, we notice that the way we have formulated our question as a maximization problem might not be ideal. In our future work, we plan to redefine our question into a minimization problem for the following reasons. First, most of the classical problems where one take advantages of using the classical W1 distance are minimization problems. In quantum settings, the qWGAN algorithm by [2] is also a minimization problem. This is, as far as we can tell, because one always works on the (under)estimated W1 distance rather than the exact distance, and making the exact distance approach 0 (which is known to be minimum) seems to be easier than finding a "hanging" unknown maximum value. Second, making a set of objects close to each other is easier than making them apart. For instance, making A, B, C close can be done by making A and B close and making B and C close, whereas, the task of making them apart requires consideration of any pair of two objects. Third, we have at least two ideas of formulating our problem as a minimization problem, which are: 1. Switching to the complementary channel of the Pauli error channel, 2. Constructing an end-to-end encoder-noise-decoder model. For the first approach, we consider the complementary channel $\tilde{\Phi}(\rho) = \sum_{P,Q\in\mathcal{P}_r} \sqrt{\pi_P \pi_Q}\,\mathrm{Tr}\left[P\rho Q\right]|P\rangle\langle Q|$, where $\mathcal{P}_r$ denotes the set of Pauli operators whose supports are on upto $r = \lfloor \frac{d-1}{2} \rfloor$ qubits. We have done initial calculations and found that requiring $\tilde{\Phi}(|\overline{\psi_i}\rangle) = \tilde{\Phi}(|\overline{\psi_j}\rangle)$ for $(i,j) = (1,2),(3,4),(5,6)$ is sufficient to guarantee the K-L criteria. Thus, in the complementary channel frame, our task will be to minimize $W_1\left(\tilde{\Phi}(|\overline{\psi_i}\rangle), \tilde{\Phi}(|\overline{\psi_j}\rangle)\right)$. For the second approach, we do not learn the optimal distance directly but via training an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$ simultaneously in order to protect information against the noise process $\mathcal{N}$, i.e. to make $\mathcal{D}\cdot\mathcal{N}\cdot\mathcal{E} \approx \mathcal{I}$. This approach

is indeed the core idea in [5].

---

[1] G. D. Palma, M. Marvian, D. Trevisan, and S. Lloyd, "The quantum wasserstein distance of order 1," 2020.

[2] B. T. Kiani, G. D. Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, "Quantum earth mover's distance: A new approach to learning quantum data," 2021.

[3] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014.

[4] J. Romero, J. P. Olson, and A. Aspuru-Guzik, "Quantum autoencoders for efficient compression of quantum data," *Quantum Science and Technology*, vol. 2, p. 045001, Aug 2017.

[5] P. D. Johnson, J. Romero, J. Olson, Y. Cao, and A. Aspuru-Guzik, "Qvector: an algorithm for device-tailored quantum error correction," 2017.

[6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, p. 4812, Nov. 2018.

[7] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature Communications*, vol. 12, no. 1, 2021.

[8] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," 2017.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[10] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, M. S. Alam, S. Ahmed, J. M. Arrazola, C. Blank, A. Delgado, S. Jahangiri, K. McKiernan, J. J. Meyer, Z. Niu, A. Száva, and N. Killoran, "Pennylane: Automatic differentiation of hybrid quantum-classical computations," 2020.