Quantum Wasserstein GANs
Shouvanik Chakrabarti, Yiming Huang, Tongyang Li, Soheil Feizi, and Xiaodi Wu
Department of Computer Science, Institute for Advanced Computer Studies, and Joint Center for Quantum Information and Computer Science, University of Maryland

Abstract
Inspired by previous studies on the adversarial training of classical and quantum generative models, we propose the first design of quantum Wasserstein Generative Adversarial Networks (WGANs), which has been shown to improve the robustness and the scalability of the adversarial training of quantum generative models even on noisy quantum hardware. Specifically, we propose a definition of the Wasserstein semimetric between quantum data, which inherits a few key theoretical merits of its classical counterpart. We also demonstrate how to turn the quantum Wasserstein semimetric into a concrete design of quantum WGANs that can be efficiently implemented on quantum machines.

Motivation
- The first practical applications of quantum computers are expected to be for training parameterized quantum circuits, which are networks of classical quantum gates controlled by classical parameters.
- These circuits can be used as a parameterized representation of functions as an alternate to neural networks, often called quantum neural networks.
- An important application is to learn the characteristics of quantum processes in physics, or as unknown quantum states. These measurements are quantities that have been shown to improve the training of classical GANs.
- Different quantum distributions can provide the same distance is continuous in those parameters as well, unlike classical Wasserstein Distance.

Quantum vs Classical Distributions
- Quantum distributions over a space $\Gamma$ are described by a density operator that is a positive semi-definite matrix over a space $\mathcal{X} \otimes \mathcal{Y}$ with trace 1.
- Generalization of a classical random process: Classical distributions are characterized by diagonal density operators.
- Quantum Measurements are described by Hermitian matrices over $\mathcal{X}$. The outcome of a measurement $H$ on a state $\rho$ is a random variable with expectation $\text{Tr}(\rho H)$.
- Classical Measurements are characterized by diagonal Hermitian Matrices.
- A quantum distribution over a system with components $\mathcal{X}, \mathcal{Y}$ is described by a quantum operator in the Kronecker product space $\mathcal{X} \otimes \mathcal{Y}$. The marginal distributions are obtained via the partial trace $\text{Tr}_{\mathcal{Y}}$.
- Different quantum distributions can provide the same distribution of classical outcomes: eg. the uniform classical distribution $\frac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the quantum state $\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Wasserstein Distance
- Classical Wasserstein Distance: The minimum cost of transporting a probability distribution to another given the cost $c(x,y)$ of transporting unit mass between given points.
- The transport plan is encoded as a joint distribution with its marginals as the two input distributions.
- Given classical distributions $P, Q$ the distance can be formulated in matrix form as $\max_{\pi} \text{Tr}(\pi C)$ s.t. $\pi \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})$, $\text{Tr}_{\mathcal{Y}}(\pi) = \text{diag}(Q)$, $\text{Tr}_{\mathcal{X}}(\pi) = \text{diag}(P)$ where $C = c(x,y)$.
- Advantages over other distances: If one of the input distributions is generated by applying a parameterized function that is continuous in the parameters, the Wasserstein distance is continuous in those parameters as well, unlike measures such as the KL-divergence and JS-divergence.
- The Wasserstein distance with $c(x,y) = \|x-y\|$ has been used to formulate Wasserstein GANs that have been shown to have advantages over vanilla GANs in training.

Quantum Wasserstein Semi-metric
- Relax the requirement that $\phi, \psi$ are diagonal. $\mathcal{W}(P, Q) = \max_{\pi} \text{Tr}(\pi C)$ s.t. $\pi \in \text{Pos}(\mathcal{X} \otimes \mathcal{Y})$, $\text{Tr}_{\mathcal{Y}}(\pi) = Q$, $\text{Tr}_{\mathcal{X}}(\pi) = P$.
- To ensure that $\mathcal{W}(P, P) = 0$ for all quantum states, $C$ fixed to be $\frac{1}{2}(1_{\mathcal{X} \otimes \mathcal{Y}} - \text{SWAP})$, where the SWAP operator swaps the subsystems $\mathcal{X}, \mathcal{Y}$ of a composite system (leaves $P \otimes P$ unchanged).
- Symmetric and positive, but does not satisfy the triangle inequality.

Quantum Wasserstein GAN
- The adversarial optimization problem for a GAN is formulated using the dual form of the semi-metric with an entropic regularization to remove the hard constraints.
- The fake state is generated using a parameterized quantum circuit $G$. We use a well-studied model for parameterized circuits [2].
- The discriminator $\phi, \psi$ and regularizer $\xi_G$ are quantities measured on the real and fake states. These measurements can be also implemented using quantum circuits as a linear combination of simpler measurements.

Structure of quantum Wasserstein GAN
- The system is scalable: the loss function and gradients can be evaluated efficiently on a quantum computer (using basic gates polynomial in the number of parameters.)

Numerical Evaluation
- The QWGAN can be used to compress quantum circuits used for quantum algorithms by finding a smaller circuit that approximates the output of the algorithm. A smaller parameterized quantum circuit is trained to generate a quantum state that encodes the quantum process. This results in a small circuit that closely matches the output on a large fraction of the inputs.

Compressing Quantum Circuits
- Compressing a 3-qubit circuit for Hamiltonian Simulation (reducing 11,900 gates to 52).

References

Implementation Details
- If the state being learned is pure (a single state) the generator in a single circuit applied to a known state. If the state is mixed (a distribution over states) we apply one of the set $G_1, \ldots, G_p$ of states with probability $p_1, \ldots, p_p$.
- The quantum relative entropy $\lambda Tr(\log(\rho) - \log(\text{fake} \otimes \text{real}))$ is added to the primal which results in a regularizer $\xi_G = \frac{1}{2} \exp(-C - \phi \otimes \psi - 1_P \otimes 1_P)/\lambda$.

Image