The Quantum Wasserstein Distance of Order 1

Giacomo De Palma

GdP, Milad Marvian, Dario Trevisan, Seth Lloyd

IEEE Transactions on Information Theory 67(10), 6627-6643 (2021)

GdP, Milad Marvian, Cambyse Rouzé, Daniel Stilck França,

arXiv:2204.03455
The classical Wasserstein distance

- Probability distributions as distributions of unit amount of mass
- Moving unit mass from $x$ to $y$ has cost $d(x,y)$
- $W_1(p,q)$: minimum cost to transport $p$ onto $q$
- Recovers $d$ for Dirac deltas
- Induced by a norm
- Countless applications in geometric analysis, probability, information theory, machine learning
- For bit strings, $d =$ Hamming distance
Quantum $W_1$ distance: Why?

- Hamming distance ubiquitous in classical probability, information theory, machine learning
- Yet no quantum version for qubits!!
- Bit flip small change wrt Hamming distance, but can generate orthogonal state
- Orthogonal states maximally far for any unitarily invariant distance
- Desired properties:
  - Recovery of Hamming distance for canonical basis states
  - Robust wrt one-qubit operations
  - Global quantities (e.g., entropy) continuous
Quantum $W_1$ norm

- Neighboring states: coincide after discarding one qudit
- Require: neighboring states have distance at most one, i.e., differences between neighboring states belong to unit ball
- Quantum $W_1$ norm: maximum norm that assigns distance at most one to any couple of neighboring states
- Unit ball: convex hull of differences between neighboring states
- Semidefinite program!
Properties

• Recovers classical $W_1$ distance for states diagonal in canonical basis
• Recovers Hamming distance for canonical basis states
• Extensivity

\[ \| \rho_{AB} - \sigma_{AB} \|_{W_1} \geq \| \rho_A - \sigma_A \|_{W_1} + \| \rho_B - \sigma_B \|_{W_1} \]
• Relation with trace distance

\[ \frac{1}{2} \| \rho - \sigma \|_1 \leq \| \rho - \sigma \|_{W_1} \leq \frac{n}{2} \| \rho - \sigma \|_1 \]
• Robust wrt local operations

\[ \text{Tr}_A \rho = \text{Tr}_A \sigma \quad \Rightarrow \quad \| \rho - \sigma \|_{W_1} \leq 2 |A| \]
Shallow quantum circuits

- Expand $W_1$ distance by at most twice the size of the largest light-cone of a qudit

$$\left\| U \rho U^\dagger - U \sigma U^\dagger \right\|_{W_1} \leq 2B(U) \left\| \rho - \sigma \right\|_{W_1}$$
Continuity of the von Neumann entropy

• von Neumann entropy
  \[ S(\rho) = -\text{Tr} \left[ \rho \log \rho \right] \]

• Quantifies uncertainty

• Continuity bound wrt trace distance void for orthogonal states, but flipping one qudit can turn state into orthogonal state with entropy change at most \(2 \log d\)

• Continuity bound wrt quantum \(W_1\) distance

\[
\frac{1}{n} |S(\rho) - S(\sigma)| \leq h_2 \left( \frac{1}{n} \|\rho - \sigma\|_{W_1} \right) + \frac{1}{n} \|\rho - \sigma\|_{W_1} \log \left( d^2 - 1 \right)
\]

\[
h_2(p) = -p \log p - (1 - p) \log (1 - p)
\]

• Only intensive quantities!
Transportation-Cost Inequality (TCI)

- Quantum relative entropy
  \[ S(\rho \| \omega) = \text{Tr} [\rho (\ln \rho - \ln \omega)] \]

- Pinsker’s inequality
  \[ \frac{1}{2} \| \rho - \omega \|_1 \leq \sqrt{\frac{1}{2} S(\rho \| \omega)} \]

- Quantum TCI for product states
  \[ \frac{1}{n} \| \rho - \omega_1 \otimes \ldots \otimes \omega_n \|_{W_1} \leq \sqrt{\frac{1}{2n} S(\rho \| \omega_1 \otimes \ldots \otimes \omega_n)} \]
The quantum Lipschitz constant

- Lipschitz constant
  \[ \| f \|_L = \max_{x, y} \frac{|f(x) - f(y)|}{d(x, y)} \]

- Quantum generalization
  \[ \| H \|_L = 2 \max_{i \in [n]} \min_{H_{ic}} \| H - \mathbb{1}_i \otimes H_{ic} \|_\infty \]

- Recovers classical Lipschitz constant for operators diagonal in canonical basis

- Provides dual SDP for quantum $W_1$ distance
  \[ \| \rho - \sigma \|_{W_1} = \max_{\| H \|_L \leq 1} \text{Tr} [(\rho - \sigma) H] \]
Gaussian concentration for maximally mixed state

- In high dimension, smooth functions are essentially constant

- Upper bound on partition function
  \[
  \frac{1}{n} \ln \operatorname{Tr} e^{H} \leq \ln d + \frac{1}{8} \| H \|_{L}^{2} \quad \quad \operatorname{Tr} H = 0
  \]

- Spectrum of \( H \) lies in interval with size \( O \left( \sqrt{n} \| H \|_{L} \right) \)
  \[
  \frac{1}{d^{n}} \dim \left( H \geq n \delta \right) \leq e^{-\frac{2n\delta^{2}}{\| H \|_{L}^{2}}}
  \]
Quadratic concentration for product states

• \( \omega \) product state

\[
\text{Var}_\omega H \leq n \left\| H \right\|_L^2
\]

• \( \rho \) output of quantum circuit with blow-up \( B \)

\[
\text{Var}_\rho H \leq 4n B^2 \left\| H \right\|_L^2
\]
Combinatorial optimization

- Goal: find bit string that maximizes cost function $C$
- Local cost: sum of functions each depending on $O(1)$ bits
- Efficient classical algorithms usually achieve

$$C = a C_{\text{max}} \quad 0 < a \leq 1$$

- **Example:** maximum cut problem, i.e., find the bipartition of a graph that maximizes the # of edges connecting the two parts
- Associate one bit to each vertex, set to 1 bits in second half of bipartition
- NP complete!
Variational quantum algorithms

- Associate one qubit to each bit, quantum Hamiltonian to cost function

\[ H = \sum_{x \in \{0,1\}^n} C(x) \ket{x} \bra{x} \]

- Train parametric quantum circuit to generate high-energy states

- **Example:** Quantum Alternating Operator Ansatz (QAOA)

- Alternate time evolution with \( H \) and mixing Hamiltonian

\[
\left( \prod_{k=1}^{P} e^{-i\gamma_k} \sum_{i=1}^{n} X_i e^{-i\beta_k H} \right) \ket{+} \otimes n
\]
Limitations of QAOA for MaxCut

- Toy model: $D$-regular bipartite graph ($\text{maxcut} = n \frac{D}{2}$)
- Technical assumption:

\[
C(x) \geq \left( \frac{D}{2} - \sqrt{D - 1} \right) \min \{ |x|, n - |x| \} \quad \forall x \in \{0, 1\}^n
\]

- Satisfied by Ramanujan expander graphs with $D \geq 3$ and for large $n$ by random $D$-regular graphs with high probability
- Observation [Bravyi et al., PRL 125, 260505 (2020)]: QAOA circuit commutes with $X \otimes n$
- Probability distribution of output measurement symmetric wrt flipping all bits and cannot be concentrated on single string
Limitations of QAOA for MaxCut

• Result: if

$$\text{Tr} [\rho H] \geq C_{\text{max}} \left( \frac{5}{6} + \frac{\sqrt{D-1}}{3D} \right)$$

then the quadratic concentration inequality implies

$$P \geq \frac{1}{2 \log (D+1)} \log \frac{n}{576} = \Omega(\log n)$$

• Holds for any circuit and initial state commuting with $X^\otimes n$

• Improves Bravyi et al.

$$P \geq \frac{1}{3 (D + 1)} \log_2 \frac{n}{4096}$$
Limitations of noisy quantum circuits

• Goal: generate high-energy state of traceless local Hamiltonian $H$

• Model: $L$ layers of 2-qubit gates with depolarizing noise on each qubit after each layer

• Result: output energy exponentially concentrated about 0

$$\mathbb{P} (|H| \geq n \delta) \leq \exp \left( -\frac{n}{2} \left( \frac{\delta^2}{\|H\|_L^2} - (1 - p)^{2L} \right) \right)$$

• Quantum advantage exponentially unlikely for

$$L > \left| \frac{\log \frac{\delta}{\|H\|_L}}{\log (1 - p)} \right| = O \left( \frac{1}{p} \right)$$
Proof idea

- 2-Rényi divergence wrt maximally mixed state $\omega$ decreases exponentially with $L$

$$D_2(\rho \| \omega) = \log \text{Tr} \left( \omega^{-\frac{1}{4}} \rho \omega^{-\frac{1}{4}} \right)^2 \leq n (1 - p)^{2L}$$

- Gaussian concentration for $\omega$ implies Gaussian concentration for $\rho$

$$\mathbb{P}_\rho (|H| \geq n \delta) \leq \sqrt{2} \exp \left( \frac{1}{2} D_2(\rho \| \omega) - \frac{n \delta^2}{\|H\|^2_L} \right)$$
Perspectives

- Robustness of quantum algorithms for machine learning with quantum input
- Design of quantum error correcting codes
- Quantum rate-distortion theory