The quantum Wasserstein distance of order 1

Giacomo De Palma

GdP, Milad Marvian, Dario Trevisan, Seth Lloyd
arXiv:2009.04469

Bobak Kiani, GdP, Milad Marvian, Zi-Wen Liu, Seth Lloyd
arXiv:2101.03037

Motivations

- Hamming distance ubiquitous in classical probability, information theory, machine learning
- Yet no quantum version for qudits!!
- 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-qudit operations
  - Global quantities (*e.g.*, entropy) continuous

The classical Wasserstein distance

- Probability distributions on metric space interpreted 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
- Induced by a norm
- Countless applications in geometric analysis, probability, information theory, machine learning
- We consider bit strings (strings of symbols from finite alphabet) with Hamming distance

Quantum W_1 norm

- Neighboring states: coincide after discarding one qudit
- Require: neighboring states have distance at most one
- Minimum unit ball: convex hull of differences between neighboring states

$$\mathcal{B}_n = \left\{ \sum_{i=1}^{n} p_i \left(\rho^{(i)} - \sigma^{(i)} \right) : \mathrm{Tr}_i \rho^{(i)} = \mathrm{Tr}_i \sigma^{(i)} \right\}$$

- Maximum norm

$$\left\| \rho - \sigma \right\|_{W_1} = \min \left(t \geq 0 : \rho - \sigma \in t\, \mathcal{B}_n \right)$$

Properties

- Recovers classical W_1 distance for states diagonal in canonical basis, Hamming distance for canonical basis states
- Symmetries: local unitaries, qudit permutations
- Contractive wrt one-qudit quantum channels
- Additive wrt tensor product

$$\|\rho \otimes \rho' - \sigma \otimes \sigma'\|_{W_1} = \|\rho - \sigma\|_{W_1} + \|\rho' - \sigma'\|_{W_1}$$

- 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: if Φ acts on *k* qudits,

$$\|\Phi(\rho) - \rho\|_{W_1} \leq 2\,k$$

Contraction coefficient

- Contraction of trace distance

$$\eta(\Phi) = \max_{\rho \neq \sigma \in \mathcal{S}_n} \frac{\left\|\Phi(\rho) - \Phi(\sigma)\right\|_1}{\left\|\rho - \sigma\right\|_1}$$

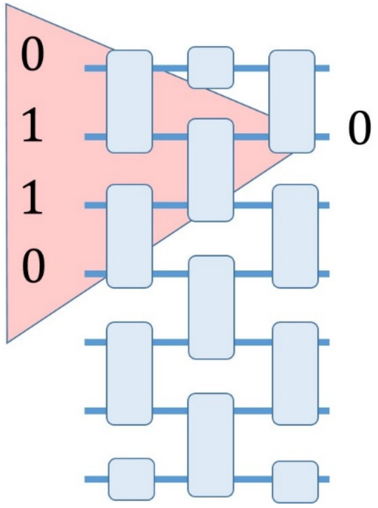
- Trivial for $n \to \infty$ for tensor power channels

$$\Phi(\rho) \neq \Phi(\sigma) \quad \Longrightarrow \quad \frac{1}{2} \lim_{n\to\infty} \left\|\Phi^{\otimes n}\left(\rho^{\otimes n}\right) - \Phi^{\otimes n}\left(\sigma^{\otimes n}\right)\right\|_1 = \lim_{n\to\infty} \eta\left(\Phi^{\otimes n}\right) = 1$$

- Contraction of quantum W_1 distance

$$\left\|\Phi\right\|_{W_1 \to W_1} = \max_{\rho \neq \sigma \in \mathcal{S}_n} \frac{\left\|\Phi(\rho) - \Phi(\sigma)\right\|_{W_1}}{\left\|\rho - \sigma\right\|_{W_1}}$$

Contraction coefficient



- At most twice the size of the largest light-cone of a qudit
- Tensor power channels: Φ one-qudit channel with fixed point ω, $\mathcal{E}$ replaces input with ω

$$\left\|\Phi^{\otimes n}\right\|_{W_1 \to W_1} \leq \left\|\Phi - \mathcal{E}\right\|_{\diamond}$$

- Amplitude damping channel (decay probability *1-p*)

$$\left\|\Phi_p^{\otimes n}\right\|_{W_1 \to W_1} \leq 2\sqrt{\frac{p}{1-p}}$$

- Depolarizing channel (input replaced with probability *1-p*)

$$\left\|\mathcal{E}_p^{\otimes n}\right\|_{W_1 \to W_1} = p$$

Continuity of the von Neumann entropy

- Continuity bounds wrt trace distance / fidelity void for orthogonal states, but flipping one qudit can turn state into orthogonal state with entropy change at most 2 ln *d*
- Continuity bound wrt quantum W_1 distance

$$|S(\rho) - S(\sigma)| \leq g\left(\|\rho - \sigma\|_{W_1}\right) + \|\rho - \sigma\|_{W_1} \ln\left(d^2\, n\right)$$

$$g(t) = (t+1)\ln(t+1) - t\ln t \leq \ln(t+1) + 1$$

$$\|\rho - \sigma\|_{W_1} = o\left(\frac{n}{\ln n}\right) \quad \Longrightarrow \quad |S(\rho) - S(\sigma)| = o(n)$$

- ln *n* not present for classical W_1 distance

Relation with relative entropy

- Pinsker's inequality

$$\frac{1}{2}\|\rho-\sigma\|_1 \leq \sqrt{\frac{1}{2}S(\rho\|\sigma)}$$

- Improvement to quantum Marton's transportation inequality

$$\|\rho-\sigma_1\otimes\ldots\otimes\sigma_n\|_{W_1} \leq \sqrt{\frac{n}{2}S(\rho\|\sigma_1\otimes\ldots\otimes\sigma_n)}$$

The quantum Lipschitz constant

- Quantum Lipschitz constant

$$\|H\|_L = 2\max_{i\in[n]}\min\left(\|H - H_{\bar{i}}\|_\infty : H_{\bar{i}} \text{ does not act on } i\text{-th qudit}\right)$$

- Recovers classical Lipschitz constant for operators diagonal in canonical basis
- Quantum W_1 distance as SDP

$$\|\rho - \sigma\|_{W_1} = \max_{\|H\|_L\leq 1} \mathrm{Tr}\left[H\left(\rho - \sigma\right)\right]$$

Quantum Gaussian concentration inequality

- In high dimension, smooth functions are essentially constant
- Bound on partition function

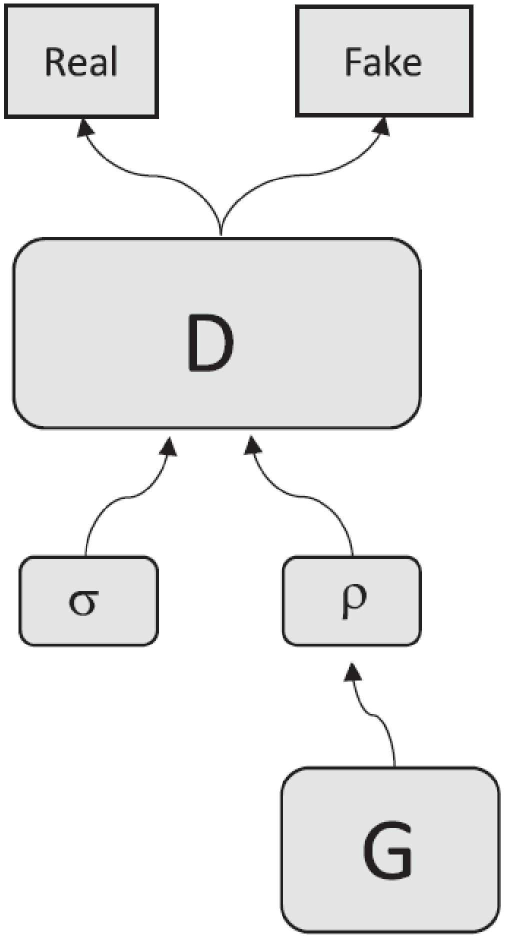
$$\frac{1}{n}\ln \mathrm{Tr}\,\mathrm{e}^{tH} \le \ln d + \frac{t^2}{8}\left\|H\right\|_L^2 \qquad \mathrm{Tr}\,H = 0$$

- Spectrum of *H* lies in interval with size $O\left(\sqrt{n}\left\|H\right\|_L\right)$

$$\dim\left(H \ge \left(\frac{\mathrm{Tr}\,H}{d^n} + \delta\sqrt{n}\left\|H\right\|_L\right)\mathbb{I}\right) \le d^n\,\mathrm{e}^{-2\delta^2}$$

Quantum Generative Adversarial Networks

- Algorithm to learn target quantum state from samples via parameterized quantum circuit
- Generator and discriminator trained against each other
- Generator generates quantum state as close as possible to target
- Discriminator discriminates between target and generated state
- Unitarily invariant distances as cost function suffer from exponentially vanishing gradients



Quantum Wasserstein GANs

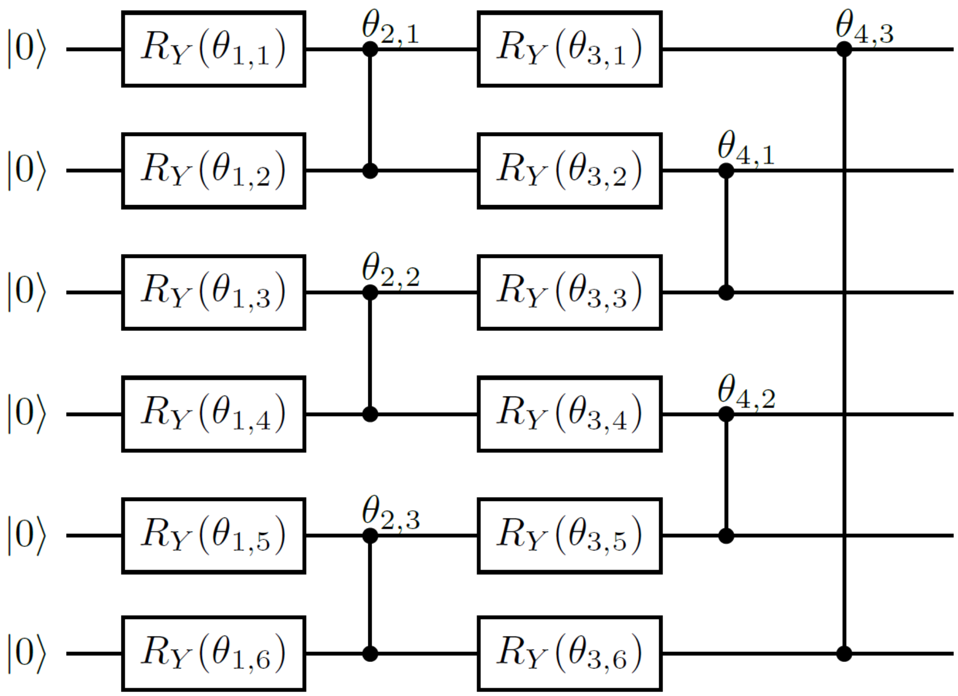
- Solution: quantum W_1 distance as cost function
- Target state σ, generated state ρ, cost function

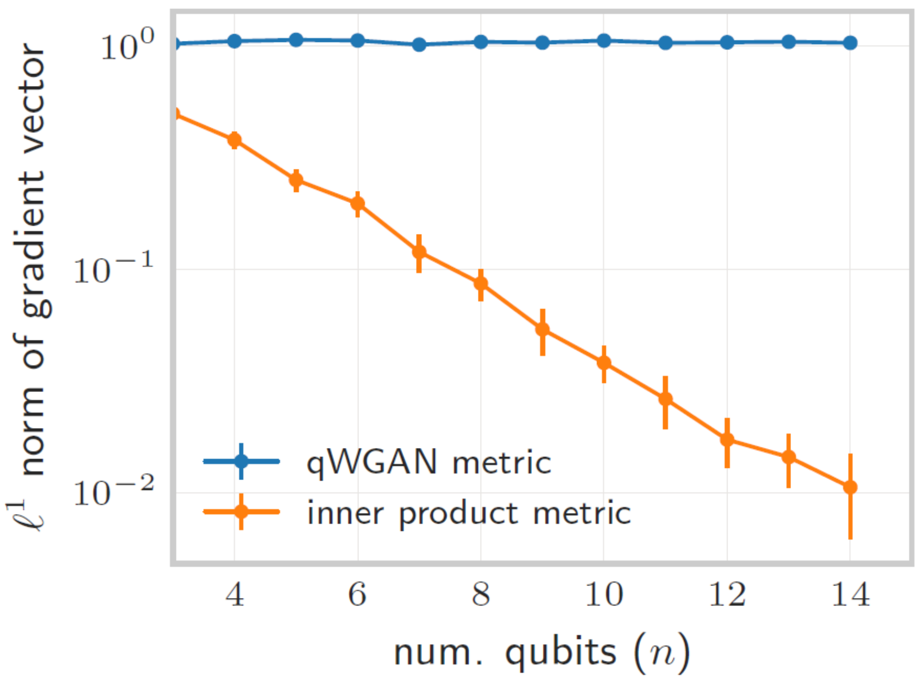
$$C = \left|\mathrm{Tr}\left[\left(\rho - \sigma\right) H\right]\right| \leq \left\|\rho - \sigma\right\|_{W_1} \qquad \left\|H\right\|_L \leq 1$$

- Generator tunes ρ with gradient descent to lower *C*
- Discriminator tunes *H* to raise *C*
- Simplification: *H* sum of few-qubits terms

Quantum Wasserstein GANs

- Parametric quantum circuit in teacher-student setup
- Size of gradient of cost function independent of number of qubits!





Perspectives

- Explore qWGAN!
- NISQ devices
- Robustness of quantum machine learning
- Computational capabilities of shallow quantum circuits
- Design of quantum error correcting codes
- Quantum rate-distortion theory
- Mixing time of quantum Markov semigroups
- Quantum spin systems on infinite lattices