The quantum Wasserstein distance of order 1

Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd

1 Motivations

The most prominent distinguishability measures between quantum states are the trace distance, the quantum fidelity and the quantum relative entropy, and they all have in common the property of being unitarily invariant [1–3]. A fundamental consequence of this property is that the distance between any couple of quantum states with orthogonal supports is always maximal. However, this property is not always desirable. For certain applications, it is natural to use a distance with respect to which the state $|0\rangle^{\otimes n}$ is much closer to $|1\rangle \otimes |0\rangle^{\otimes(n-1)}$ than to $|1\rangle^{\otimes n}$. Some desirable properties can be recovering the Hamming distance for vectors of the canonical basis, and more generally robustness against local perturbations on the input states. Such a distance may, for example, provide better continuity bounds for the von Neumann entropy since the von Neumann entropy is also robust against local perturbations. In particular, any operation on one qubit can change the entropy of a state by at most $\ln 4$, which does not depend on the number of qubits. Therefore, the entropy of an n-qubit state with initial entropy $O(n)$ remains $O(n)$ after such an operation. However, this continuity property cannot be captured by any unitarily invariant distinguishability measure, since a one-qubit operation can bring the initial state into an orthogonal state, resulting in a maximum possible change in the unitarily invariant measure.

In the setting of classical probability distributions on a metric space, the distances originating from the theory of optimal mass transport have emerged as prominent distances with the properties above. Their exploration has led to the creation of an extremely fruitful field in mathematical analysis, with applications ranging from differential geometry and partial differential equations to machine learning [4–6]. Given two mass or probability distributions on a metric space and given the cost of moving a unit mass between each couple of points of the metric space, optimal mass transport theory assigns a cost to each plan that transports the first distribution to the second one. The minimum cost over all the possible transport plans defines the optimal transport distance between the distributions [4]. One of the most prominent choices for the cost function is the distance on the metric space, giving rise to the Wasserstein distance of order 1, or W_1 distance.

2 Our contribution

We propose a generalization of the W_1 distance to the set of the quantum states of n qudits. The proposed quantum W_1 distance is based on the notion of neighboring states.

Definition 1 (Quantum W_1 distance)**.** Two quantum states of n qudits are neighboring if they coincide after a suitable qudit is discarded. The quantum W_1 distance is the maximum distance that is induced by a norm that assigns distance at most one to any couple of neighboring states.

We prove several properties of the proposed quantum W_1 distance:

- It is invariant with respect to permutations of the qudits and unitary operations acting on one qudit and additive with respect to the tensor product. Moreover, the W_1 distance between two quantum states which coincide after discarding k qudits is at most $2k$. In particular, any quantum operation on k qudits can displace the initial quantum state by at most $2k$ in the proposed distance.
- It recovers the Hamming distance for vectors of the canonical basis, and more generally the classical W_1 distance for quantum states diagonal in the canonical basis.
- Its ratio with the trace distance lies between 1 and n.

We define a generalization to quantum observables of the Lipschitz constant of real-valued functions on a metric space. We prove that, as in the classical case, the proposed quantum W_1 distance between two quantum states is equal to the maximum difference between the expectation values of the two states with respect to an observable with Lipschitz constant at most one. This dual formulation provides a recipe to calculate the proposed quantum W_1 distance using a semidefinite program.

Our main result is a continuity bound for the von Neumann entropy with respect to the proposed quantum W_1 distance. In the limit $n \to \infty$ this bound implies that, if two quantum states have distance $o(n/\ln n)$, their entropies can differ by at most $o(n)$. The von Neumann entropy is intimately linked to the entanglement properties of a quantum state, and our bound implies that the entanglement of a quantum state is robust against perturbations with size $o(n/\ln n)$ in the quantum W_1 distance.

We explore the relation between the quantum W_1 distance and the quantum relative entropy. In particular, we prove a quantum generalization of Marton's trasportation inequality, stating that the square root of the relative entropy between a generic quantum state and a product quantum state provides an upper bound to their quantum W_1 distance. We apply the quantum Marton's inequality to prove an upper bound to the partition function of a quantum Hamiltonian in terms of its quantum Lipschitz constant. A fundamental consequence of this result is a quantum Gaussian concentration inequality, stating that most of the eigenvalues of a quantum observable lie in a small interval whose size depends on its Lipschitz constant.

We study the contraction coefficient with respect to the proposed quantum W_1 distance of the n-th tensor power of a one-qudit quantum channel. While the contraction coefficient of these quantum channels with respect to the trace distance is trivial in the limit $n \to \infty$, we are able to prove an upper bound to the contraction coefficient for the proposed quantum W_1 distance which does not depend on n. Moreover, we prove that the contraction coefficient of a generic n-qudit quantum channel with respect to the proposed quantum W_1 distance is upper bounded by the size of the light-cones of the qudits.

3 Future perspectives

In the classical setting, the Wasserstein distances have a huge variety of applications ranging from mathematical analysis to machine learning and information theory. We expect the proposed quantum W_1 distance to be a powerful tool with a broad range of applications in quantum information, quantum computing and quantum machine learning. We propose a few of them in the following.

- **Robustness of quantum machine learning:** A fundamental desirable property of classical machine learning algorithms is the robustness with respect to small perturbations in the input [7], and the same property should be desirable also when the machine learning algorithm is quantum [8]. *Quantum input:* In the scenario with quantum input data, the size of the perturbations in the input has so far been measured with the trace distance or with the quantum fidelity [9], with respect to which any two perfectly distinguishable quantum states are maximally far. On the contrary, in the classical setting any two different inputs are perfectly distinguishable. The proposed quantum W_1 distance is a perfect candidate to measure the size of the perturbations for quantum algorithms for machine learning with a quantum input, and therefore provides a suitable quality factor for the robustness of the quantum algorithms for machine learning. *Classical input:* In the scenario with classical input data, choosing the right method to encode the input into quantum states is essential in the success of any quantum algorithm for machine learning [8,10]. In particular, it is reasonable to require the encoding to be robust with respect to small perturbations of the input. The proposed quantum W_1 distance provides a natural measure for the size of the input perturbations and hence for the robustness of the encoding.

- **Quantum Generative Adversarial Networks:** In analogy to classical GANs, quantum GANs [11] are a paradigm for quantum machine learning where a generator tries to produce quantum samples as close as possible to some true quantum data, and a discriminator tries to discriminate the generated from the true data. For classical GANs, the Wasserstein distances have turned out to be the right candidate for the loss function, since they solve the problem of the vanishing gradient in the training that plagued the GANs trained with the total variation distance or with the Jensen–Shannon divergence [12]. For this reason, quantum Wasserstein distances have been proposed as cost function for the quantum GANs [13]. The proposed quantum W_1 distance recovers the classical W_1 distance for states diagonal in the canonical basis and satisfies most of its properties, and is therefore a good candidate for the loss function of the quantum GANs.

- **Quantum rate distortion theory:** Rate-distortion theory addresses the problem of determining the maximum compression rate of a signal if a certain level of distortion in the recovered signal is allowed [14]. The measure employed to quantify the distortion plays a fundamental role, and for a discrete alphabet the most prominent distortion measure is the Hamming distance. Rate-distortion theory has been extended to the quantum setting in the iid regime [15–22] with a symbol-wise entanglement fidelity as distortion measure. The limitation to iid arises since such symbol-wise entanglement fidelity can be defined only when the quantum state to be encoded is a tensor product of one-qudit states. The proposed quantum W_1 distance does not have this limitation, and is therefore a candidate to extend quantum rate distortion theory beyond the iid regime.

- **Shallow quantum circuits:** The Hamming distance plays a key role in the study of the computational capabilities of quantum circuits [23, 24]. The proposed quantum W_1 distance recovers the Hamming distance for vectors of the canonical basis and is stable with respect to the action of local shallow quantum circuits. Therefore, the proposed quantum W_1 distance can be useful in characterizing the states generated by constant depth circuits and therefore in extending the current results on their computational capabilities.

References

[1] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[2] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017.

[3] Alexander S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. Texts and Monographs in Theoretical Physics. De Gruyter, 2019.

[4] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[6] Gabriel Peyré and Marco Cuturi. Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[7] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[8] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

[9] Nana Liu and Peter Wittek. Vulnerability of quantum classification to adversarial perturbations. *Physical Review A*, 101(6):062331, 2020.

[10] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. Quantum embeddings for machine learning. *arXiv preprint arXiv:2001.03622*, 2020.

[11] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Physical Review Letters*, 121(4):040502, 2018.

[12] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 214–223, 2017.

[13] Shouvanik Chakrabarti, Huang Yiming, Tongyang Li, Soheil Feizi, and Xiaodi Wu. Quantum Wasserstein generative adversarial networks. In *Advances in Neural Information Processing Systems*, pages 6781–6792, 2019.

[14] Robert M. Gray. *Entropy and Information Theory*. Springer New York, 2013.

[15] Howard Barnum. Quantum rate-distortion coding. *Physical Review A*, 62(4):042309, 2000.

[16] Igor Devetak and Toby Berger. Quantum rate-distortion theory for iid sources. In *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No. 01CH37252)*, page 276. IEEE, 2001.

[17] Igor Devetak and Toby Berger. Quantum rate-distortion theory for memoryless sources. *IEEE Transactions on Information Theory*, 48(6):1580–1589, 2002.

[18] Xiao-Yu Chen and Wei-Ming Wang. Entanglement information rate distortion of a quantum gaussian source. *IEEE Transactions on Information Theory*, 54(2):743–748, 2008.

[19] Nilanjana Datta, Min-Hsiu Hsieh, and Mark M Wilde. Quantum rate distortion, reverse Shannon theorems, and source-channel separation. *IEEE Transactions on Information Theory*, 59(1):615–630, 2013.

[20] Nilanjana Datta, Min-Hsiu Hsieh, Mark M Wilde, and Andreas Winter. Quantum-to-classical rate distortion coding. *Journal of Mathematical Physics*, 54(4):042201, 2013.

[21] Mark M Wilde, Nilanjana Datta, Min-Hsiu Hsieh, and Andreas Winter. Quantum rate-distortion coding with auxiliary resources. *IEEE Transactions on Information Theory*, 59(10):6755–6773, 2013.

[22] Sina Salek, Daniela Cadamuro, Philipp Kammerlander, and Karoline Wiesner. Quantum rate-distortion coding of relevant information. *IEEE Transactions on Information Theory*, 65(4):2603–2613, 2018.

[23] Lior Eldar and Aram W Harrow. Local Hamiltonians whose ground states are hard to approximate. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 427–438. IEEE, 2017.

[24] Sergey Bravyi, Alexander Kliesch, Robert Koenig, and Eugene Tang. Obstacles to state preparation and variational optimization from symmetry protection. *arXiv preprint arXiv:1910.08980*, 2019.