# The quantum Wasserstein distance of order 1

Giacomo De Palma[1], Milad Marvian[1,2], Dario Trevisan[3], and Seth Lloyd[1]

[1]Massachusetts Institute of Technology, Department of Mechanical Engineering and Research Laboratory of Electronics, 02139 Cambridge MA, USA
[2]Department of Electrical and Computer Engineering, Center for Quantum Information and Control (CQuIC), University of New Mexico, 87131 Albuquerque NM, USA
[3]Università degli Studi di Pisa, Dipartimento di Matematica, 56127 Pisa, Italy

September 11, 2020

**Abstract**

We propose a generalization of the Wasserstein distance of order 1 to the quantum states of $n$ qudits. The proposal recovers the Hamming distance for the vectors of the canonical basis, and more generally the classical Wasserstein distance for quantum states diagonal in the canonical basis. The proposed distance is invariant with respect to permutations of the qudits and unitary operations acting on one qudit and is additive with respect to the tensor product. Our main result is a continuity bound for the von Neumann entropy with respect to the proposed distance, which significantly strengthens the best continuity bound with respect to the trace distance. We also propose a generalization of the Lipschitz constant to quantum observables. The notion of quantum Lipschitz constant allows us to compute the proposed distance with a semidefinite program. We prove a quantum version of Marton's transportation inequality and a quantum Gaussian concentration inequality for the spectrum of quantum Lipschitz observables. Moreover, we derive bounds on the contraction coefficients of shallow quantum circuits and of the tensor product of one-qudit quantum channels with respect to the proposed distance. We discuss other possible applications in quantum machine learning, quantum Shannon theory, and quantum many-body systems.

# 1 Introduction

## 1.1 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.

## 1.2 The classical Wasserstein distances

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 a finite set $\mathcal{X}$, any distance $D$ on $\mathcal{X}$ induces a transport distance on the set of the probability distributions on $\mathcal{X}$, where the distance between the probability distributions $p$ and $q$ is the minimum of the mean distance over joint probability distributions on $\mathcal{X}^2$ with marginals $p$ and $q$. More precisely, we have the following definitions:

**Definition 1** (Coupling)**.** A coupling between the probability distributions $p$ and $q$ on $\mathcal{X}$ is a probability distribution $\pi$ on $\mathcal{X}^2$ with marginals $p$ and $q$, i.e., such that

$$p(x) = \sum_{y\in\mathcal{X}} \pi(x,y)\,, \qquad q(y) = \sum_{x\in\mathcal{X}} \pi(x,y)\,, \qquad x,\, y \in \mathcal{X}\,. \tag{1}$$

We denote with $\mathcal{C}(p,q)$ the set of the couplings between $p$ and $q$.

**Definition 2** (Classical $W_\alpha$ distances)**.** For any $\alpha \geq 1$, the $W_\alpha$ distance or Wasserstein distance of order $\alpha$ between the probability distributions $p$ and $q$ on $\mathcal{X}$ is

$$W_\alpha(p,q) = \left(\min_{\pi\in\mathcal{C}(p,q)} \sum_{x,\,y\in\mathcal{X}} D(x,y)^\alpha\, \pi(x,y)\right)^{\frac{1}{\alpha}}\,. \tag{2}$$

Although many properties of the $W_\alpha$ distances do not depend on the choice of $\alpha$, in recent years the distances $W_1$ and $W_2$ are playing a prominent role. The $W_1$ distance is also called Monge–Kantorovich distance, after the foundational works of Monge and Kantorovich [7,8]. In particular, Kantorovich noticed that the $W_1$ distance is in fact induced by a norm, and introduced the transport problem (2) as a linear programming problem (see [9] for a detailed historical account).

In many cases the set $\mathcal{X}$ is already endowed with a distance, e.g., when dealing with subsets of Riemannian manifolds or weighted graphs. However, one can always consider the trivial distance

$$D(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases} , \tag{3}$$

and the induced $W_1$ distance coincides with the total variation distance

$$W_1(p, q) = \frac{1}{2} \|p - q\|_1 . \tag{4}$$

The Hamming distance provides a natural choice when $\mathcal{X}$ is a set of finite strings over an alphabet:

**Definition 3** (Hamming distance)**.** For any $k \in \mathbb{N}$, let $[k] = \{1, \ldots, k\}$. The Hamming distance between $x, y \in [d]^n$ is the number of different components:

$$h(x, y) = |\{i \in [n] : x_i \neq y_i\}| . \tag{5}$$

The classical $W_1$ distance with respect to the Hamming distance is called Ornstein's $\bar{d}$ distance and was first considered in [10], together with its extension to stationary stochastic processes. It has found many applications in ergodic theory and information theory, such as coding theorems for a large class of discrete noisy channels with memory and rate distortion theory [11].

Finally, there has been a surge of interest towards applications of transportation distances in machine learning in the context of Generative Adversarial Networks (GANs) [12]. GANs [13] provide a useful algorithm to learn an unknown probability distribution using a neural network. The learning is performed by training a generator trying to produce samples of the unknown distribution against a discriminator trying to distinguish the true from the generated samples. The training process is a minimax game that converges to a Nash equilibrium. The choice of the loss functions for the discriminator plays a crucial role to ensure convergence in the training procedure. Employing the Wasserstein distances as loss function of GANs alleviates the problem of the vanishing gradient in the training, which plagued the original version with the Jensen–Shannon divergence (the symmetrized relative entropy) [12]. Suitable variants of the Wasserstein distances which further improve the efficiency of the training have also been proposed [14, 15].

### 1.3 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. We anticipate here an informal definition and refer to section 3 for the details.

**Definition 4** (Quantum $W_1$ distance, informal)**.** 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.

In section 4, 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 (subsection 4.1) and additive with respect to the tensor product (subsection 4.2). Moreover, the $W_1$ distance between two quantum states which coincide after discarding $k$ qudits is at most $2k$ (subsection 4.3). 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 (subsection 4.4).

- Its ratio with the trace distance lies between 1 and $n$ (subsection 4.5).

In section 5, 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 (section 6). 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.

In section 7, 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. In section 8, 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.

In section 9, 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.

We conclude in section 10 by discussing other possible applications of the defined quantum $W_1$ distance in quantum machine learning, quantum information, and quantum many-body systems.

## 1.4 Related works

Several quantum generalizations of the Wasserstein distances have been proposed. One line of research by Carlen, Maas, Datta and Rouzé [16–21] defines a quantum $W_2$ distance built on the definition of a quantum differential structure and on the equivalent dynamical formulation of the $W_2$ distance provided by Benamou and Brenier [22], which assigns a length to each path of probability distributions that connects the source with the target. The key property of this proposal is that the resulting quantum distance is induced by a Riemannian metric on the manifold of quantum states, and the quantum generalization of the heat semigroup is the gradient flow of the von Neumann entropy with respect to this metric. This quantum generalization of the $W_2$ distance has been shown to be intimately linked to both entropy and Fisher information [20], and has led to determine the rate of convergence of the quantum Ornstein-Uhlenbeck semigroup [17, 23]. Exploiting their quantum differential structure, Refs. [18, 19] also define a quantum generalization of the Lipschitz constant and the $W_1$ distance, and prove that it satisfies a Talagrand inequality, which also implies some concentration inequalities. Alternative definitions of quantum $W_1$ distances based on a quantum differential structure are proposed in Refs. [24–27]. Refs. [28–30] propose quantum $W_1$ distances based on a distance between the vectors of the canonical basis.

Another line of research by Golse, Mouhot, Paul and Caglioti [31–36] arose in the context of the study of the semiclassical limit of quantum mechanics and defines a quantum $W_2$ distance built on a quantum generalization of the couplings. This distance was the key element to prove that the mean-field limit of quantum mechanics is uniform in the semiclassical limit [31], and has been employed as a cost function to train the quantum counterpart of deep generative adversarial networks [37, 38]. Ref. [39] proposes another quantum $W_2$ distance based on quantum couplings, with the property that each quantum coupling is associated to a quantum channel. The relation between quantum couplings and quantum channels in the framework of von Neumann algebras has been explored in [40]. The problem of defining a quantum $W_1$ distance through quantum couplings has been explored in Ref. [41].

The quantum $W_\alpha$ distance between two quantum states can be defined as the classical $W_\alpha$ distance between the probability distributions of the outcomes of an informationally complete measurement performed on the states, which is a measurement whose probability distribution completely determines the state. This definition has been explored for Gaussian quantum systems with the heterodyne measurement in Refs. [42–44].

Notions of quantum Hamming ball of a subspace have been defined in Refs. [45, 46], who employ them to prove a Talagrand concentration inequality and a quantum generalization of de Finetti's theorem, respectively.

The Wasserstein distances have also been generalized to other noncommutative settings, such as noncommutative geometry [47, 48], free probability and random matrix theory [49].

## 2 Notation

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 $\mathcal{O}_n^+ \subset \mathcal{O}_n$ the subset of the positive semidefinite linear operators on $\mathcal{H}_n$, by $\mathcal{S}_n \subset \mathcal{O}_n^+$ the set of the quantum states of $\mathcal{H}_n$, and by $\mathcal{P}_n$ the set of the probability distributions on $[d]^n$. For any $\mathcal{I} \subseteq [n]$, let $\rho_{\mathcal{I}}$ be the marginal of $\rho \in \mathcal{S}_n$ over the qudits in $\mathcal{I}$. For any $X \in \mathcal{O}_n$, let $\|X\|_1$ be its trace norm, given by the sum of the absolute values of its eigenvalues.

## 3 The quantum $W_1$ distance

Our proposal for a quantum Wasserstein distance of order 1 is based on the following notion of neighboring quantum states:

**Definition 5** (Neighboring quantum states)**.** We say that $\rho$ and $\sigma \in \mathcal{S}_n$ are neighboring if they coincide after discarding one qudit, i.e., if $\mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma$ for some $i \in [n]$. We denote by $\mathcal{N}_n \subset \mathcal{O}_n^T$ the set of the differences between couples of neighboring quantum states:

$$\mathcal{N}_n = \bigcup_{i=1}^n \mathcal{N}_n^{(i)}\,, \qquad \mathcal{N}_n^{(i)} = \{\rho - \sigma : \rho,\, \sigma \in \mathcal{S}_n,\, \mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma\}\;, \qquad i \in [n]\,, \tag{6}$$

and with

$$\mathcal{B}_n = \left\{\sum_{i=1}^n p_i\left(\rho^{(i)} - \sigma^{(i)}\right) : p_i \geq 0,\ \sum_{i=1}^n p_i = 1,\ \rho^{(i)},\, \sigma^{(i)} \in \mathcal{S}_n,\ \mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)}\right\} \tag{7}$$

the convex hull of $\mathcal{N}_n$.

*Remark* 1. Other equivalent definitions of neighboring quantum states are possible, see Appendix A for details.

**Proposition 1.** *$\mathcal{B}_n$ is a bounded, closed, centrally symmetric (i.e., $-\mathcal{B}_n = \mathcal{B}_n$) and convex subset of $\mathcal{O}_n^T$ with nonempty interior.*

*Proof.* The only nontrivial property is the nonempty interior, which follows from Proposition 6, where we will prove that $\mathcal{B}_n$ is the unit ball of a norm which is upper bounded by $n/2$ times the trace norm. □

The classical $W_1$ distance is induced by a norm, and the distance between any couple of neighboring probability distributions on $[d]^n$ is at most one (Lemma 6 of Appendix C). Therefore, we look for a distance between quantum states that is induced by a norm that assigns distance at most one to each couple of neighboring quantum states, i.e., $\mathcal{N}_n$ should be contained in the unit ball of the norm. Since the unit ball of any norm is convex, also $\mathcal{B}_n$ should be contained in the unit ball of the norm. Any norm is completely determined by its unit ball, and any bounded, closed, centrally symmetric and convex set with nonempty interior is the unit ball of some norm. Therefore, we define the quantum $W_1$ norm as the unique norm on $\mathcal{O}_n^T$ whose unit ball is $\mathcal{B}_n$:

**Definition 6** (Quantum $W_1$ norm)**.** We define the quantum $W_1$ norm on $\mathcal{O}_n^T$ as the unique norm with unit ball $\mathcal{B}_n$, i.e., for any $X \in \mathcal{O}_n^T$,

$$\begin{aligned}\|X\|_{W_1} &= \min\left(t \geq 0 : X \in t\,\mathcal{B}_n\right) \\ &= \frac{1}{2}\min\left(\sum_{i=1}^n \left\|X^{(i)}\right\|_1 : X^{(i)} \in \mathcal{O}_n^T,\ \mathrm{Tr}_i\, X^{(i)} = 0,\ X = \sum_{i=1}^n X^{(i)}\right). \qquad (8)\end{aligned}$$

The equivalence between the two expressions in (8) is proved in Lemma 1 of Appendix C.

The quantum $W_1$ norm is the maximum norm on $\mathcal{O}_n^T$ such that the difference between each couple of neighboring quantum states has norm at most one. We define the quantum $W_1$ distance as the distance induced by the quantum $W_1$ norm:

**Definition 7** (Quantum $W_1$ distance)**.** We define the quantum $W_1$ distance between the quantum states $\rho$ and $\sigma$ of $\mathcal{H}_n$ as

$$\begin{aligned}&W_1(\rho,\sigma) = \|\rho - \sigma\|_{W_1} \\ &= \min\left(\sum_{i=1}^n c_i : c_i \geq 0,\ \rho - \sigma = \sum_{i=1}^n c_i\left(\rho^{(i)} - \sigma^{(i)}\right),\ \rho^{(i)},\ \sigma^{(i)} \in \mathcal{S}_n,\ \mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)}\right). \quad (9)\end{aligned}$$

The equivalence between the two expressions in (9) can be proved along the same lines of Lemma 1 of Appendix C.

For the sake of a simpler notation, we state all our results in terms of the quantum $W_1$ norm. Their counterparts for the quantum $W_1$ distance trivially follow.

# 4 Properties of the quantum $W_1$ distance

## 4.1 Symmetries

The classical $W_1$ distance on probability distributions on $[d]^n$ is invariant with respect to permutations of the $n$ subsystems and to permutations of the $d$ elements of one subsystem. The following Proposition 2 states that the quantum $W_1$ norm keeps all the symmetries of the classical case, and the permutations of the $d$ elements of one subsystem get enhanced to unitary operations acting on one qudit.

**Proposition 2** (Symmetries of the quantum $W_1$ norm)**.** *The quantum $W_1$ norm is invariant with respect to permutations of the qudits and unitary operations acting on one qudit, and non-increasing with respect to quantum channels acting on one qudit.*

*Proof.* The claim follows since all the transformations above send $\mathcal{N}_n$ to itself. $\square$

## 4.2 Tensorization

In the following Proposition 3, we prove that the quantum $W_1$ distance is additive with respect to the tensor product as its classical counterpart. This property is fundamental for distortion measures in rate distortion theory [11, Chapter 5], and it is not satisfied by the trace distance. On the other hand, the quantum relative entropy and the logarithm of the inverse of the quantum fidelity are additive, but they are not proper distances since they do not satisfy the triangle inequality.

**Proposition 3** (Tensorization)**.** *For any $X \in \mathcal{O}_{m+n}^T$,*

$$\|X\|_{W_1} \geq \|\mathrm{Tr}_{m+1\ldots m+n}X\|_{W_1} + \|\mathrm{Tr}_{1\ldots m}X\|_{W_1} \,, \tag{10}$$

*and for any $\rho$, $\sigma \in \mathcal{S}_{m+n}$,*

$$\|\rho - \sigma\|_{W_1} \geq \|\rho_{1\ldots m} - \sigma_{1\ldots m}\|_{W_1} + \|\rho_{m+1\ldots m+n} - \sigma_{m+1\ldots m+n}\|_{W_1} \,. \tag{11}$$

*Moreover, for any $\rho'$, $\sigma' \in \mathcal{S}_m$ and any $\rho''$, $\sigma'' \in \mathcal{S}_n$,*

$$\|\rho' \otimes \rho'' - \sigma' \otimes \sigma''\|_{W_1} = \|\rho' - \sigma'\|_{W_1} + \|\rho'' - \sigma''\|_{W_1} \,. \tag{12}$$

*Proof.* Let $X^{(1)}, \ldots, X^{(m+n)} \in \mathcal{O}_{m+n}^T$ be such that

$$\mathrm{Tr}_i X^{(i)} = 0 \quad \forall\, i \in [m+n] \,, \qquad X = \sum_{i=1}^{m+n} X^{(i)} \,. \tag{13}$$

We have $\mathrm{Tr}_{m+1\ldots m+n}X^{(i)} = 0$ for any $i = m+1, \ldots, m+n$ and $\mathrm{Tr}_{1\ldots m}X^{(i)} = 0$ for any $i \in [m]$, therefore

$$\mathrm{Tr}_{m+1\ldots m+n}X = \sum_{i=1}^{m} \mathrm{Tr}_{m+1\ldots m+n}X^{(i)} \,, \qquad \mathrm{Tr}_{1\ldots m}X = \sum_{i=m+1}^{m+n} \mathrm{Tr}_{1\ldots m}X^{(i)} \,, \tag{14}$$

then

$$\begin{aligned}\|\mathrm{Tr}_{m+1\ldots m+n}X\|_{W_1} + \|\mathrm{Tr}_{1\ldots m}X\|_{W_1} &\leq \frac{1}{2}\sum_{i=1}^{m} \left\|\mathrm{Tr}_{m+1\ldots m+n}X^{(i)}\right\|_1 + \frac{1}{2}\sum_{i=m+1}^{m+n} \left\|\mathrm{Tr}_{1\ldots m}X^{(i)}\right\|_1 \\ &\leq \frac{1}{2}\sum_{i=1}^{m+n} \left\|X^{(i)}\right\|_1 \,, \end{aligned} \tag{15}$$

and the claim (10) follows.

On the one hand, we have from (10)

$$\|\rho'\otimes\rho''-\sigma'\otimes\sigma''\|_{W_1}\ge\|\rho'-\sigma'\|_{W_1}+\|\rho''-\sigma''\|_{W_1}\ . \tag{16}$$

On the other hand, we get with the help of Lemma 2 of Appendix C

$$\begin{aligned}\|\rho'\otimes\rho''-\sigma'\otimes\sigma''\|_{W_1}&\le\|(\rho'-\sigma')\otimes\rho''\|_{W_1}+\|\sigma'\otimes(\rho''-\sigma'')\|_{W_1}\\&\le\|\rho'-\sigma'\|_{W_1}+\|\rho''-\sigma''\|_{W_1}\ ,\end{aligned} \tag{17}$$

and the claim (12) follows. □

**Corollary 1.** *For any $\rho,\,\sigma\in\mathcal{S}_n$,*

$$\|\rho-\sigma\|_{W_1}\ge\frac{1}{2}\sum_{i=1}^n\|\rho_i-\sigma_i\|_1\ , \tag{18}$$

*and equality holds whenever both $\rho$ and $\sigma$ are product states.*

### 4.3 Local operations

The quantum $W_1$ distance between two quantum states that coincide after discarding one qudit is at most one. In the following Proposition 4, we consider the case of quantum states that coincide after discarding $k$ qudits, and we prove that their distance is at most $2k$.

**Proposition 4.** *Let $\mathcal{I}\subseteq[n]$, and let $X\in\mathcal{O}_n^T$ such that $\mathrm{Tr}_{\mathcal{I}}X=0$. Then,*

$$\|X\|_{W_1}\le|\mathcal{I}|\,\frac{d^2-1}{d^2}\,\|X\|_1\ , \tag{19}$$

*and for any $\rho,\,\sigma\in\mathcal{S}_n$ such that $\rho_{\mathcal{I}}=\sigma_{\mathcal{I}}$,*

$$\|\rho-\sigma\|_{W_1}\le|\mathcal{I}|\,\frac{d^2-1}{d^2}\,\|\rho-\sigma\|_1\ . \tag{20}$$

*Proof.* Without loss of generality, we can assume that $\mathcal{I}=[k]$ for some $k\in[n]$. For any $i\in[k]$, let

$$X^{(i)}=\frac{\mathbb{I}_d^{\otimes(i-1)}}{d^{i-1}}\otimes\mathrm{Tr}_{1\ldots i-1}X-\frac{\mathbb{I}_d^{\otimes i}}{d^i}\otimes\mathrm{Tr}_{1\ldots i}X\,, \tag{21}$$

such that

$$\mathrm{Tr}_iX^{(i)}=0\,,\qquad X=\sum_{i=1}^kX^{(i)}\,. \tag{22}$$

We have with the help of Lemma 3 of Appendix C

$$\|X\|_{W_1}\le\frac{1}{2}\sum_{i=1}^k\left\|X^{(i)}\right\|_1\le\frac{d^2-1}{d^2}\sum_{i=1}^k\|\mathrm{Tr}_{1\ldots i-1}X\|_1\le|\mathcal{I}|\,\frac{d^2-1}{d^2}\,\|X\|_1\ , \tag{23}$$

and the claim follows. □

An important consequence of Proposition 4 is that the $W_1$ distance is continuous with respect to local operations, in the sense that any operation performed on $k$ qudits can displace the initial quantum state by at most $2k$ in the distance:

**Corollary 2.** *Let $\Phi$ be a quantum channel on $\mathcal{H}_n$ that acts on at most $k$ qudits. Then, for any $\rho \in \mathcal{S}_n$,*

$$\|\Phi(\rho) - \rho\|_{W_1} \leq 2\,k\,\frac{d^2-1}{d^2}\,. \tag{24}$$

*Proof.* Let $\mathcal{I} \subseteq [n]$ be the set of qudits on which $\Phi$ acts. Then, $\mathrm{Tr}_{\mathcal{I}}\left[\Phi(\rho) - \rho\right] = 0$, and the claim follows from Proposition 4. □

## 4.4 Recovery of the classical $W_1$ distance

The following Proposition 5 states that for quantum states diagonal in the canonical basis, the quantum $W_1$ distance recovers the classical $W_1$ distance.

**Proposition 5.** *Let $p,\, q \in \mathcal{P}_n$, and let*

$$\rho = \sum_{x\in[d]^n} p(x)\,|x\rangle\langle x|\,, \qquad \sigma = \sum_{y\in[d]^n} q(y)\,|y\rangle\langle y|\,. \tag{25}$$

*Then,*

$$\|\rho - \sigma\|_{W_1} = W_1(p,q)\,. \tag{26}$$

*In particular, the quantum $W_1$ distance between vectors of the canonical basis coincides with the Hamming distance:*

$$\left\||x\rangle\langle x| - |y\rangle\langle y|\right\|_{W_1} = h(x,y)\,, \qquad x,\, y \in [d]^n\,. \tag{27}$$

*Proof.* Let $x,\, y \in [d]^n$. We get from (18)

$$\left\||x\rangle\langle x| - |y\rangle\langle y|\right\|_{W_1} = \frac{1}{2}\sum_{i=1}^n \left\||x_i\rangle\langle x_i| - |y_i\rangle\langle y_i|\right\|_1 = h(x,y)\,, \tag{28}$$

and the claim (27) follows.

On the one hand, let $\pi \in \mathcal{C}(p,q)$. We have

$$\begin{aligned}\|\rho - \sigma\|_{W_1} &= \left\|\sum_{x,\,y\in[d]^n} \pi(x,y)\left(|x\rangle\langle x| - |y\rangle\langle y|\right)\right\|_{W_1} \leq \sum_{x,\,y\in[d]^n} \pi(x,y)\left\||x\rangle\langle x| - |y\rangle\langle y|\right\|_{W_1} \\ &= \sum_{x,\,y\in[d]^n} h(x,y)\,\pi(x,y)\,,\end{aligned} \tag{29}$$

therefore

$$\|\rho - \sigma\|_{W_1} \leq W_1(p,q)\,. \tag{30}$$

On the other hand, there exist a probability distribution $r$ on $[n]$ and quantum states $\rho^{(1)}, \sigma^{(1)}, \ldots, \rho^{(n)}, \sigma^{(n)} \in \mathcal{S}_n$ such that

$$\mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)} \quad \forall\, i \in [n]\,, \qquad \rho - \sigma = \left\|\rho - \sigma\right\|_{W_1} \sum_{i=1}^n r_i \left(\rho^{(i)} - \sigma^{(i)}\right)\,. \tag{31}$$

We can assume that each $\rho^{(i)}$ and each $\sigma^{(i)}$ is diagonal in the canonical basis. Let $p^{(1)}, \ldots, p^{(n)}$ and $q^{(1)}, \ldots, q^{(n)}$ be the associated probability distributions on $[d]^n$, such that

$$p - q = \left\|\rho - \sigma\right\|_{W_1} \sum_{i=1}^n r_i \left(p^{(i)} - q^{(i)}\right)\,. \tag{32}$$

Since also the classical $W_1$ distance is induced by a norm, we have from Lemma 6 of Appendix C

$$W_1(p,q) \le \left\|\rho - \sigma\right\|_{W_1} \sum_{i=1}^n r_i\, W_1\left(p^{(i)},\, q^{(i)}\right) \le \left\|\rho - \sigma\right\|_{W_1}\,, \tag{33}$$

and the claim (26) follows. □

### 4.5 Relation with the trace distance

The following Proposition 6 states that the quantum $W_1$ norm keeps the same upper and lower bounds in terms of the trace norm as its classical counterpart.

**Proposition 6** (Relation with the trace norm)**.** *For any $X \in \mathcal{O}_n^T$,*

$$\frac{1}{2}\left\|X\right\|_1 \le \left\|X\right\|_{W_1} \le \frac{n}{2}\left\|X\right\|_1\,. \tag{34}$$

*Moreover, if $\mathrm{Tr}_i X = 0$ for some $i \in [n]$, and in particular if $n = 1$,*

$$\left\|X\right\|_{W_1} = \frac{1}{2}\left\|X\right\|_1\,, \tag{35}$$

*i.e., for any $\rho,\, \sigma \in \mathcal{S}_n$ such that $\mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma$ for some $i \in [n]$,*

$$\left\|\rho - \sigma\right\|_{W_1} = \frac{1}{2}\left\|\rho - \sigma\right\|_1\,. \tag{36}$$

*Proof.* On the one hand, let $X^{(1)}, \ldots, X^{(n)}$ be as in (8). We have

$$\left\|X\right\|_1 \le \sum_{i=1}^n \left\|X^{(i)}\right\|_1\,, \tag{37}$$

therefore

$$\frac{1}{2}\left\|X\right\|_1 \le \left\|X\right\|_{W_1}\,. \tag{38}$$

On the other hand, let

$$X = X^+ - X^- \,, \tag{39}$$

where $X^+$ and $X^-$ are positive semidefinite with orthogonal supports and satisfy

$$\mathrm{Tr}\, X^\pm = \frac{1}{2} \left\|X\right\|_1 \,. \tag{40}$$

We can choose in (8)

$$X^{(i)} = \frac{2}{\left\|X\right\|_1} \left(\mathrm{Tr}_{i\ldots n} X^- \otimes \mathrm{Tr}_{1\ldots i-1} X^+ - \mathrm{Tr}_{i+1\ldots n} X^- \otimes \mathrm{Tr}_{1\ldots i} X^+\right) \,, \tag{41}$$

such that

$$\left\|X^{(i)}\right\|_1 \le \left\|X\right\|_1 \,, \tag{42}$$

therefore

$$\left\|X\right\|_{W_1} \le \frac{1}{2} \sum_{i=1}^{n} \left\|X^{(i)}\right\|_1 \le \frac{n}{2} \left\|X\right\|_1 \,, \tag{43}$$

and the claim (34) follows.

Let us now assume that $\mathrm{Tr}_i X = 0$. On the one hand, we have already proved that

$$\frac{1}{2} \left\|X\right\|_1 \le \left\|X\right\|_{W_1} \,. \tag{44}$$

On the other hand, choosing in (8)

$$X^{(i)} = X \,, \qquad X^{(1)} = \ldots = X^{(i-1)} = X^{(i+1)} = \ldots = X^{(n)} = 0 \,, \tag{45}$$

we get

$$\left\|X\right\|_{W_1} \le \frac{1}{2} \left\|X\right\|_1 \,, \tag{46}$$

and the claim (35) follows. □

# 5 The quantum Lipschitz constant and the dual formulation of the quantum $W_1$ distance

The classical $W_1$ distance between the probability distributions $p$ and $q$ on the metric space $\mathcal{X}$ admits a dual formulation as maximum difference between the expectation values of a Lipschitz function on $p$ and $q$:

$$W_1(p,q) = \max\left(\sum_{x\in\mathcal{X}} f(x)\,(p(x) - q(x)) : f \in \mathbb{R}^{\mathcal{X}},\ \left\|f\right\|_L \le 1\right) \,, \tag{47}$$

where

$$\left\|f\right\|_L = \max_{x\neq y\in\mathcal{X}} \frac{|f(x) - f(y)|}{D(x,y)} \tag{48}$$

is the Lipschitz constant of $f$, and $D$ is the distance on $\mathcal{X}$. This dual formulation makes the computation of the classical $W_1$ distance a semidefinite program (actually, the same holds for all the $W_\alpha$ distances).

We prove in the following that the computation of the quantum $W_1$ norm is also a semidefinite program. First, we need to define a quantum generalization of the Lipschitz constant:

**Definition 8** (Quantum Lipschitz constant)**.** We define the quantum Lipschitz constant of $H \in \mathcal{O}_n$ as the dual norm of the quantum $W_1$ norm on $\mathcal{O}_n^T$:

$$\begin{aligned}\|H\|_L &= \max\left(\mathrm{Tr}\left[H\,X\right] : X \in \mathcal{O}_n^T,\ \|X\|_{W_1} \leq 1\right) = \max\left(\mathrm{Tr}\left[H\,X\right] : X \in \mathcal{N}_n\right)\\ &= \max_{i\in[n]}\left(\max\left(\mathrm{Tr}\left[H\left(\rho-\sigma\right)\right] : \rho,\,\sigma \in \mathcal{S}_n,\ \mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma\right)\right). \end{aligned} \tag{49}$$

The quantum Lipschitz constant recovers the classical Lipschitz constant for operators diagonal in the canonical basis:

**Proposition 7.** *Let $f : [d]^n \to \mathbb{R}$, and let*

$$F = \sum_{x\in[d]^n} f(x)\,|x\rangle\langle x|\,. \tag{50}$$

*Then,*

$$\|F\|_L = \|f\|_L\ . \tag{51}$$

*Proof.* Let $\mathcal{D}$ be the quantum channel on $\mathbb{C}^d$ that dephases the input state in the canonical basis:

$$\mathcal{D}(X) = \sum_{i=1}^{d}\langle i|X|i\rangle\,|i\rangle\langle i|\,,\qquad X \in \mathcal{S}_1\,. \tag{52}$$

From Proposition 2, we have for any $X \in \mathcal{O}_n^T$

$$\left\|\mathcal{D}^{\otimes n}(X)\right\|_{W_1} \leq \|X\|_{W_1}\ . \tag{53}$$

We then have with the help of Proposition 5

$$\begin{aligned}\|F\|_L &= \max\left(\sum_{x\in[d]^n} f(x)\left(\langle x|\rho|x\rangle - \langle x|\sigma|x\rangle\right) : \rho,\,\sigma\in\mathcal{S}_n\,,\quad \|\rho-\sigma\|_{W_1}\leq 1\right)\\ &= \max\left(\sum_{x\in[d]^n} f(x)\left(\langle x|\mathcal{D}^{\otimes n}(\rho)|x\rangle - \langle x|\mathcal{D}^{\otimes n}(\sigma)|x\rangle\right) : \rho,\,\sigma\in\mathcal{S}_n\,,\quad \left\|\mathcal{D}^{\otimes n}(\rho-\sigma)\right\|_{W_1}\leq 1\right)\\ &= \max\left(\sum_{x\in[d]^n} f(x)\left(p(x)-q(x)\right) : p,\,q\in\mathcal{P}_n\,,\quad W_1(p,q)\leq 1\right) = \|f\|_L\ ,\end{aligned} \tag{54}$$

and the claim follows. ∎

Proposition 15 of Appendix B provides an estimate of the quantum Lipschitz constant up to multiplicative error $\sqrt{2}$ that does not require any optimization. The following Proposition 8 provides a dual formulation of the quantum Lipschitz constant:

**Proposition 8.** *For any $H \in \mathcal{O}_n$,*

$$\|H\|_L = 2 \max_{i\in[n]} \min_{H^{(i)}\in\mathcal{O}_{n-1}} \left\| H - \mathbb{I}_d^{(i)} \otimes H^{(i)} \right\|_\infty , \tag{55}$$

*where for any $i \in [n]$, $\mathbb{I}_d^{(i)}$ is the identity operator on the $i$-th qudit and $H^{(i)}$ does not act on the $i$-th qudit.*

*Proof.* It is sufficient to prove that

$$\max\left(\mathrm{Tr}\left[H\left(\rho-\sigma\right)\right] : \rho,\, \sigma \in \mathcal{S}_n,\ \mathrm{Tr}_1\rho = \mathrm{Tr}_1\sigma\right) = 2 \min_{K\in\mathcal{O}_{n-1}} \|H - \mathbb{I}_d \otimes K\|_\infty . \tag{56}$$

Let $\Phi : \mathbb{R}\times\mathcal{O}_{n-1} \to \mathcal{O}_n^2$ be given by

$$\Phi(t,K) = \left(t\,\mathbb{I}_d^{\otimes n} + \mathbb{I}_d \otimes K\,,\ t\,\mathbb{I}_d^{\otimes n} - \mathbb{I}_d \otimes K\right) , \qquad t\in\mathbb{R}\,,\ K\in\mathcal{O}_{n-1}\,, \tag{57}$$

such that

$$2 \min_{K\in\mathcal{O}_{n-1}} \|H - \mathbb{I}_d\otimes K\|_\infty = 2\min\left(t\in\mathbb{R} : \exists\, K\in\mathcal{O}_{n-1} : \Phi(t,K) - (H,\,-H) \in \left(\mathcal{O}_n^+\right)^2\right) \tag{58}$$

is a semidefinite program with dual program

$$\begin{aligned} &\max\left(\mathrm{Tr}\left[H\left(\alpha-\beta\right)\right] : \alpha,\, \beta\in\mathcal{O}_n^+\,,\ \Phi^\dagger(\alpha,\beta) = (2,0)\right) \\ &= \max\left(\mathrm{Tr}\left[H\left(\rho-\sigma\right)\right] : \rho,\,\sigma\in\mathcal{S}_n\,,\ \mathrm{Tr}_1\rho = \mathrm{Tr}_1\sigma\right) . \end{aligned} \tag{59}$$

$\left(\mathcal{O}_n^+\right)^2$ and $\mathbb{R}\times\mathcal{O}_{n-1}$ are both convex cones. Moreover, for any $t > \|H\|_\infty$ we have

$$\Phi(t,0) - (H,\,-H) = \left(t\,\mathbb{I}_d^{\otimes n} - H,\, t\,\mathbb{I}_d^{\otimes n} + H\right) \in \mathrm{int}\left(\mathcal{O}_n^+\right)^2 . \tag{60}$$

Therefore, from [50, Corollary 5.3.6] there is no duality gap, and the claim follows. □

In finite dimension, the dual of the dual norm always coincides with the original norm. Therefore, the quantum $W_1$ norm is the dual norm of the quantum Lipschitz constant. Thanks to Proposition 8, this dual formulation of the quantum $W_1$ norm is the dual program of the semidefinite program (8):

**Proposition 9** (Duality)**.** *The optimization problem (8) is a semidefinite program with the following dual program: for any $X \in \mathcal{O}_n^T$,*

$$\begin{aligned} \|X\|_{W_1} &= \max\left(\mathrm{Tr}\left[H\,X\right] : H\in\mathcal{O}_n,\ \|H\|_L \le 1\right) \\ &= \max\left(\mathrm{Tr}\left[H\,X\right] : H\in\mathcal{O}_n : \forall\, i\in[n]\ \exists\, H^{(i)}\in\mathcal{O}_{n-1} : \left\| H - \mathbb{I}_d^{(i)}\otimes H^{(i)}\right\|_\infty \le \frac{1}{2}\right) . \end{aligned} \tag{61}$$

## 6 $W_1$ continuity of the von Neumann entropy

The von Neumann entropy of a quantum state [1–3]

$$S(\rho) = -\mathrm{Tr}\left[\rho \ln \rho\right] , \qquad \rho \in \mathcal{S}_n \tag{62}$$

quantifies the amount of uncertainty contained in the state and plays a key role in quantum information theory. The von Neumann entropy is not sensitive to operations performed on a small subsystem: From Lemma 5 of Appendix C, any operation performed on $k$ qudits can change the entropy of the state by at most $2k \ln d$. Since already an operation performed on one qudit can generate a quantum state orthogonal to the initial state, this robustness of the von Neumann entropy cannot be captured by any unitarily invariant distinguishability measure, such as the trace distance, the quantum fidelity or the quantum relative entropy. The situation for the proposed quantum $W_1$ distance is radically different, since it is robust with respect to local perturbations.

In the classical case, the $W_1$ distance provides the following continuity bound for the Shannon entropy:

**Theorem** ($W_1$ continuity of the Shannon entropy [51, Proposition 8]). *For any $p$, $q \in \mathcal{P}_n$,*

$$|S(p) - S(q)| \le n\, h_2\left(\frac{W_1(p,q)}{n}\right) + W_1(p,q) \ln\left(d-1\right) , \tag{63}$$

*where $h_2$ is the binary entropy function*

$$h_2(x) = -x \ln x - (1-x)\ln\left(1-x\right) , \qquad 0 \le x \le 1 . \tag{64}$$

*Proof.* The proof is based on couplings. For the sake of completeness, we report it in Appendix D. □

A natural question is whether the continuity bound (63) still holds without any modification for the quantum $W_1$ distance. The answer is negative. Indeed, the right-hand side of (63) has a unique maximum equal to $n \ln d$ achieved at $W_1(p,q) = n\left(d-1\right)/d$. Since $n \ln d$ is the entropy of the maximally mixed state of $\mathcal{H}_n$, the continuity bound (63) would imply that the $W_1$ distance between the maximally mixed state and any pure state is equal to $n\left(d-1\right)/d$. However, if $\gamma$ is a maximally entangled state acting on $\left(\mathbb{C}^d\right)^{\otimes 2}$, from Lemma 4 of Appendix C for any even $n$ we have

$$\left\|\gamma^{\otimes \frac{n}{2}} - \frac{\mathbb{I}_d^{\otimes n}}{d^n}\right\|_{W_1} = \frac{n}{2}\,\frac{d^2-1}{d^2} < n\,\frac{d-1}{d} , \tag{65}$$

hence the continuity bound (63) cannot hold without modifications in the quantum setting.

Nonetheless, the von Neumann entropy has good continuity properties with respect to the quantum $W_1$ distance. Indeed, the von Neumann entropy satisfies the following continuity bound, which is equivalent to the classical bound (63) up to a factor $\ln n$:

**Theorem 1** ($W_1$ continuity of the von Neumann entropy)**.** *For any $\rho$, $\sigma\in\mathcal{S}_n$,*

$$|S(\rho)-S(\sigma)|\le g\left(\|\rho-\sigma\|_{W_1}\right)+\|\rho-\sigma\|_{W_1}\ln\left(d^2\,n\right)\,, \tag{66}$$

*where for any $t\ge0$*

$$g(t)=(t+1)\ln\left(t+1\right)-t\ln t\,. \tag{67}$$

*Proof.* Let

$$t=\|\rho-\sigma\|_{W_1}\,. \tag{68}$$

There exist a probability distribution $p$ on $[n]$ and $\sigma^{(1)}$, $\rho^{(1)}$, $\ldots$, $\sigma^{(n)}$, $\rho^{(n)}\in\mathcal{S}_n$ such that

$$\mathrm{Tr}_i\sigma^{(i)}=\mathrm{Tr}_i\rho^{(i)}\quad\forall\, i\in[n]\,,\qquad\rho-\sigma=t\sum_{i=1}^n p_i\left(\rho^{(i)}-\sigma^{(i)}\right)\,. \tag{69}$$

Let $q$ be the probability distribution on $\{0,\,\ldots,\,n\}$ given by

$$q_0=\frac{1}{t+1}\,,\qquad q_i=\frac{t}{t+1}\,p_i\,,\quad i\in[n]\,, \tag{70}$$

such that

$$q_0\,\rho+\sum_{i=1}^n q_i\,\sigma^{(i)}=q_0\,\sigma+\sum_{i=1}^n q_i\,\rho^{(i)}=\tau\in\mathcal{S}_n\,. \tag{71}$$

We have

$$S(q)=h_2(q_0)+(1-q_0)\,S(p)\le h_2(q_0)+(1-q_0)\ln n\,. \tag{72}$$

Moreover, Lemma 5 of Appendix C implies for any $i\in[n]$

$$S\left(\rho^{(i)}\right)-S\left(\sigma^{(i)}\right)\le2\ln d\,. \tag{73}$$

On the one hand, we have from the concavity of the entropy

$$S(\tau)\ge q_0\,S(\rho)+\sum_{i=1}^n q_i\,S\left(\sigma^{(i)}\right)\,. \tag{74}$$

On the other hand, we have

$$S(\tau)\le q_0\,S(\sigma)+\sum_{i=1}^n q_i\,S\left(\rho^{(i)}\right)+S(q)\,. \tag{75}$$

Putting together (74), (75), (73) and (72) we get

$$\begin{aligned}S(\rho)-S(\sigma)&\le\frac{1}{q_0}\left(\sum_{i=1}^n q_i\left(S\left(\rho^{(i)}\right)-S\left(\sigma^{(i)}\right)\right)+S(q)\right)\le\frac{1-q_0}{q_0}\ln\left(d^2\,n\right)+\frac{h_2(q_0)}{q_0}\\&=t\ln\left(d^2\,n\right)+(t+1)\ln\left(t+1\right)-t\ln t\,,\end{aligned} \tag{76}$$

and the claim follows. ☐

Theorem 1 implies that in the limit of large $n$ with fixed $d$ and for any $\epsilon > 0$, if

$$\|\rho - \sigma\|_{W_1} \leq \frac{\epsilon\, n}{\ln\left(d^2\, n\right)}\,, \tag{77}$$

then

$$|S(\rho) - S(\sigma)| \leq \epsilon\, n + O(\ln n)\,. \tag{78}$$

Since the entropy is intimately linked with entanglement, a fundamental consequence of this result is that the entanglement properties of a quantum state are robust with respect to perturbations in the quantum $W_1$ distance with size $o(n/\ln n)$. For example, we consider a bipartite quantum system $AB$ with each subsystem consisting of $n$ qudits. Let $\rho_{AB}$ be a pure quantum state of $AB$ with entanglement entropy and distillable entanglement

$$E_D(\rho_{AB}) = S(\rho_A) = O(n)\,. \tag{79}$$

For any perturbation that degrades the quantum state $\rho_{AB}$ to some state $\rho'_{AB}$ such that

$$\|\rho_{AB} - \rho'_{AB}\|_{W_1} = o\left(\frac{n}{\ln n}\right)\,, \tag{80}$$

we have

$$|S(\rho_{AB}) - S(\rho'_{AB})| = o(n)\,, \qquad |S(\rho_B) - S(\rho'_B)| = o(n)\,, \tag{81}$$

and from [52, Theorem 3.1], the distillable entanglement of $\rho'_{AB}$ is at least

$$E_D(\rho'_{AB}) \geq S(\rho'_B) - S(\rho'_{AB}) = E_D(\rho_{AB}) - o(n)\,. \tag{82}$$

# 7 Quantum Marton's transportation inequality

The quantum relative entropy between two quantum states [1–3]

$$S(\rho\|\sigma) = \mathrm{Tr}\left[\rho\left(\ln\rho - \ln\sigma\right)\right]\,, \qquad \rho,\, \sigma \in \mathcal{S}_n\,, \tag{83}$$

generalizes the classical Kullback–Leibler divergence. As in the classical case, it is always nonnegative and equal to zero if and only if $\rho = \sigma$. It can be thought as a distance between quantum states, but it is not symmetric nor it satisfies the triangle inequality. The quantum Pinsker's inequality [2, Theorem 11.9.1], [44, Eq. (14.38)]

$$\|\rho - \sigma\|_1 \leq \sqrt{2\, S(\rho\|\sigma)} \tag{84}$$

provides an upper bound for the trace distance in terms of the quantum relative entropy. In the classical case, an inequality by Marton [53] extends Pinsker's inequality to a transportation cost — information inequality, by replacing the left hand side with the $W_1$ distance induced by the Hamming distance: if $p$, $q$ are probability distributions on $[d]^n$ and $q$ is a product distribution $q(x) = \prod_{i=1}^n q_i(x_i)$, then

$$W_1(p,q) \leq \sqrt{\frac{n}{2}\, S(p\|q)}\,. \tag{85}$$

Marton's inequality (85) improves the classical Pinsker's inequality whenever

$$W_1(p,q) \geq \frac{\sqrt{n}}{2}\,\|p-q\|_1\ , \tag{86}$$

and was later extended to a larger class of distributions in discrete and continuous settings [54,55]. Noncommutative versions of (85) and related functional concentration inequalities are proposed in [19, 45, 56], with different quantum generalizations of the Wasserstein distances. In the following Theorem 2, we prove that the proposed quantum $W_1$ distance satisfies the Marton's inequality (85):

**Theorem 2** (Quantum Marton's transportation inequality)**.** *For any $\rho,\sigma \in \mathcal{S}_n$, with $\sigma = \sigma_1 \otimes \ldots \otimes \sigma_n$ product state,*

$$\|\rho-\sigma\|_{W_1} \leq \sqrt{\frac{n}{2}\,S(\rho\|\sigma)}. \tag{87}$$

*Proof.* As in the proof of Proposition 6, we write

$$\rho-\sigma = \sum_{i=1}^{n}\left(\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n} - \rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\right), \tag{88}$$

so that

$$\|\rho-\sigma\|_{W_1} \leq \frac{1}{2}\sum_{i=1}^{n}\|\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n} - \rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\|_1\,. \tag{89}$$

We apply (84) for every $i = 1,\,\ldots,\,n$,

$$\|\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n} - \rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\|_1 \leq \sqrt{2\,S\left(\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n}\|\rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\right)} \tag{90}$$

and use the concavity of the square root to obtain

$$\|\rho-\sigma\|_{W_1} \leq \sqrt{\frac{n}{2}\sum_{i=1}^{n}S\left(\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n}\|\rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\right)}. \tag{91}$$

Using the identity

$$\begin{aligned} S\left(\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n}\|\rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\right) &= S\left(\rho_{1\ldots i}\|\rho_{1\ldots i-1}\otimes\sigma_i\right)\\ &= -S(\rho_{1\ldots i}) + S(\rho_{1\ldots i-1}) - \mathrm{Tr}\left[\rho_i\log\sigma_i\right] \end{aligned} \tag{92}$$

and telescopic summation, we conclude that

$$\begin{aligned} \sum_{i=1}^{n}S\left(\rho_{1\ldots i}\otimes\sigma_{i+1\ldots n}\|\rho_{1\ldots i-1}\otimes\sigma_{i\ldots n}\right) &= -S(\rho) - \sum_{i=1}^{n}\mathrm{Tr}\left[\rho_i\log\sigma_i\right]\\ &= -S(\rho) - \mathrm{Tr}\left[\rho\log\sigma\right]\\ &= S(\rho\|\sigma)\,, \end{aligned} \tag{93}$$

and the claim follows. ∎

*Remark* 2. As in the classical case, the quantum Marton's inequality (87) improves the quantum Pinsker's inequality (84) whenever

$$\|\rho-\sigma\|_{W_1} \geq \frac{\sqrt{n}}{2}\,\|\rho-\sigma\|_1 \,. \tag{94}$$

# 8 Quantum Gaussian concentration inequality

A fundamental consequence of the classical Marton's transportation inequality is Talagrand's inequality [57], which is a Gaussian measure concentration result for Lipschitz functions. Talgrand's inequality states that any function that depends smoothly on many independent random variables, but not too much on any of them, must be essentially constant. This is a far-reaching extension of the law of large numbers for sample means of independent random variables, allowing for functions whose dependence on the many variables are quite implicit and computations may not be performed directly. As illustrated in Refs. [58–60], Talagrand's inequality is a quite general and versatile theoretical tool, with applications ranging from random combinatorial optimization to statistical physics and information theory.

Our first result in the quantum setting is the following quantum Gaussian concentration inequality, which can be thought as an upper bound to the partition function of a quantum Hamiltonian in terms of its quantum Lipschitz constant:

**Theorem 3** (Quantum Gaussian concentration inequality)**.** *For any $H\in\mathcal{O}_n$ and any $t\in\mathbb{R}$,*

$$\frac{1}{d^n}\,\mathrm{Tr}\exp\left(t\left(H-\frac{\mathrm{Tr}\,H}{d^n}\,\mathbb{I}\right)\right) \leq \exp\frac{n\,t^2\,\|H\|_L^2}{8}\,. \tag{95}$$

*Proof.* Without loss of generality, we can assume that $\mathrm{Tr}\,H=0$ and $\|H\|_L=1$, such that the claim becomes

$$\mathrm{Tr}\,\mathrm{e}^{tH} \leq d^n \exp\frac{n\,t^2}{8}\,. \tag{96}$$

From Theorem 2 and Proposition 9, we have for any $\rho\in\mathcal{S}_n$

$$S\left(\rho\middle\|\frac{\mathbb{I}_d^{\otimes n}}{d^n}\right) \geq \frac{2}{n}\left\|\rho-\frac{\mathbb{I}_d^{\otimes n}}{d^n}\right\|_{W_1}^2 \geq \frac{2}{n}\left(\mathrm{Tr}\left[H\,\rho\right]\right)^2 \geq t\,\mathrm{Tr}\left[H\,\rho\right]-\frac{n\,t^2}{8}\,. \tag{97}$$

(97) can be recast as

$$S\left(\rho\middle\|\frac{\mathrm{e}^{tH}}{\mathrm{Tr}\,\mathrm{e}^{tH}}\right) - \ln\mathrm{Tr}\,\mathrm{e}^{tH} + n\ln d + \frac{n\,t^2}{8} \geq 0\,, \tag{98}$$

and the claim follows choosing

$$\rho = \frac{\mathrm{e}^{tH}}{\mathrm{Tr}\,\mathrm{e}^{tH}}\,. \tag{99}$$

□

The left-hand side of (95) can be interpreted as the moment generating function of the empirical distribution associated to the spectrum of $H$. The right-hand side of (95) is the moment generating function of a centered Gaussian distribution with standard deviation $\sqrt{n}\left\|H\right\|_L/2$. The inequality (95) implies that the tails of the distribution of the eigenvalues of $H$ decay at least as fast as those of a Gaussian, hence the term "Gaussian concentration inequality". This consequence of Theorem 3 leads to the following concentration inequality for the spectrum of $H$:

**Corollary 3.** *Most of the eigenvalues of $H \in \mathcal{O}_n$ lie in an interval with size $O\left(\sqrt{n}\left\|H\right\|_L\right)$, i.e., for any $\delta \geq 0$,*

$$\dim\left(H \geq \left(\frac{\operatorname{Tr} H}{d^n} + \delta\sqrt{n}\left\|H\right\|_L\right)\mathbb{I}\right) \leq d^n\,\mathrm{e}^{-2\delta^2}\,, \tag{100}$$

*where for any $X,\,Y \in \mathcal{O}_n$, $\dim\left(X \geq Y\right)$ denotes the number of nonnegative eigenvalues of $X - Y$.*

*Proof.* Without loss of generality, we can assume that $\operatorname{Tr} H = 0$ and $\left\|H\right\|_L = 1$, such that the claim becomes

$$\dim\left(H \geq \delta\sqrt{n}\,\mathbb{I}\right) \leq d^n\,\mathrm{e}^{-2\delta^2}\,. \tag{101}$$

From Theorem 3, we have for any $t \geq 0$

$$d^n \exp\frac{n\,t^2}{8} \geq \operatorname{Tr}\mathrm{e}^{tH} \geq \mathrm{e}^{t\delta\sqrt{n}}\dim\left(H \geq \delta\sqrt{n}\,\mathbb{I}\right)\,, \tag{102}$$

and the claim follows choosing

$$t = \frac{4\,\delta}{\sqrt{n}}\,. \tag{103}$$

$\square$

Theorem 3 and Corollary 3 can find application in the field of many-body quantum systems to determine properties of the spectrum of local Hamiltonians, whose quantum Lipschitz constant can be easily controlled:

**Proposition 10.** *The quantum Lipschitz constant of a local Hamiltonian is upper bounded by the maximum among the operator norms of the sum of the Hamiltonian terms associated to each qudit. Formally, let*

$$H = \sum_{\mathcal{I}\subseteq[n]} H_{\mathcal{I}}\,, \tag{104}$$

*where for every $\mathcal{I} \subseteq [n]$, $H_{\mathcal{I}} \in \mathcal{O}_n$ has support on the qudits in $\mathcal{I}$ (e.g., if the qudits are arranged in a one-dimensional chain with nearest-neighbors interactions, the only nonzero terms in the sum (104) are the $2n-1$ terms associated to the subsets of $[n]$ of the form $\{k\}$ or $\{k,\,k+1\}$). Then,*

$$\left\|H\right\|_L \leq 2\max_{i\in[n]}\left\|\sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_{\mathcal{I}}\right\|_\infty\,. \tag{105}$$

*Proof.* Let $X \in \mathcal{N}_n$, and let $i \in [n]$ such that $\mathrm{Tr}_i X = 0$. We have

$$\mathrm{Tr}\left[H\, X\right] = \sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} \mathrm{Tr}\left[H_\mathcal{I}\, X\right] \le 2 \left\| \sum_{\mathcal{I}\subseteq[n]:i\in\mathcal{I}} H_\mathcal{I} \right\|_\infty , \tag{106}$$

and the claim follows. □

# 9 Contraction coefficient

A fundamental property of the trace distance is that it is contractive with respect to the action of a quantum channel [1–3], i.e., for any quantum channel $\Phi : \mathcal{O}_n \to \mathcal{O}_m$ and any $\rho,\, \sigma \in \mathcal{S}_n$,

$$\left\|\Phi(\rho) - \Phi(\sigma)\right\|_1 \le \left\|\rho - \sigma\right\|_1 . \tag{107}$$

The inequality (107) can be sharpened to

$$\left\|\Phi(\rho) - \Phi(\sigma)\right\|_1 \le \eta(\Phi) \left\|\rho - \sigma\right\|_1 , \tag{108}$$

where

$$\eta(\Phi) = \max_{\rho\neq\sigma\in\mathcal{S}_n} \frac{\left\|\Phi(\rho) - \Phi(\sigma)\right\|_1}{\left\|\rho - \sigma\right\|_1} \tag{109}$$

is called contraction coefficient of $\Phi$ with respect to the trace distance [61,62], and is strictly smaller than one for any quantum channel with a unique fixed point.

In this section, we explore the contraction properties of the quantum $W_1$ distance. Since any quantum channel $\Phi$ is trace preserving, it sends $\mathcal{O}_n^T$ to $\mathcal{O}_m^T$. We denote by

$$\left\|\Phi\right\|_{W_1\to W_1} = \max\left(\left\|\Phi(X)\right\|_{W_1} : X \in \mathcal{O}_n^T,\, \left\|X\right\|_{W_1} \le 1\right) = \max_{X\in\mathcal{N}_n} \left\|\Phi(X)\right\|_{W_1} \tag{110}$$

the norm of $\Phi$ restricted to $\mathcal{O}_n^T$ with respect to the quantum $W_1$ norm, which can also be expressed as

$$\left\|\Phi\right\|_{W_1\to W_1} = \max_{\rho\neq\sigma\in\mathcal{S}_n} \frac{\left\|\Phi(\rho) - \Phi(\sigma)\right\|_{W_1}}{\left\|\rho - \sigma\right\|_{W_1}} , \tag{111}$$

and is therefore equal to the contraction coefficient of $\Phi$ with respect to the quantum $W_1$ distance.

## 9.1 Tensor power channels

From Proposition 2, any quantum operation acting on one qudit cannot expand the quantum $W_1$ distance. Therefore, for any quantum channel $\Phi$ on $\mathbb{C}^d$, the contraction coefficient of $\Phi^{\otimes n}$ with respect to the quantum $W_1$ distance is at most 1:

$$\left\|\Phi^{\otimes n}\right\|_{W_1\to W_1} \le 1 , \tag{112}$$

as the contraction coefficient with respect to the trace distance.

Assuming that the output of $\Phi$ is not independent of the input, in the limit $n\to\infty$ the contraction coefficient of $\Phi^{\otimes n}$ with respect to the trace distance is trivial:

$$\lim_{n\to\infty}\eta\left(\Phi^{\otimes n}\right)=1\,. \tag{113}$$

Indeed, for any $\rho,\,\sigma\in\mathcal{S}_1$ such that $\Phi(\rho)\neq\Phi(\sigma)$ we have

$$\lim_{n\to\infty}\left\|\Phi^{\otimes n}\left(\rho^{\otimes n}\right)-\Phi^{\otimes n}\left(\sigma^{\otimes n}\right)\right\|_1=2\,. \tag{114}$$

For the quantum $W_1$ distance, the situation is radically different. Indeed, the following Proposition 11 provides a nontrivial upper bound to the contraction coefficient of $\Phi^{\otimes n}$ which does not depend on $n$. When $\Phi$ is a quantum Markov semigroup, Proposition 11 bounds the worst-case convergence to the equilibrium state.

**Proposition 11.** *Let $\Phi$ be a quantum channel on $\mathbb{C}^d$, let $\omega\in\mathcal{S}_1$ be a fixed point of $\Phi$, and let $\mathcal{E}$ be the quantum channel on $\mathbb{C}^d$ that replaces the input state with $\omega$. Then,*

$$\frac{1}{2}\left\|\Phi-\mathcal{E}\right\|_{1\to1}\le\left\|\Phi^{\otimes n}\right\|_{W_1\to W_1}\le\left\|\Phi-\mathcal{E}\right\|_\diamond\,, \tag{115}$$

*where we recall that for any linear map $\mathcal{F}$ on $\mathcal{O}_1$,*

$$\left\|\mathcal{F}\right\|_{1\to1}=\max_{\rho\in\mathcal{S}_1}\left\|\mathcal{F}(\rho)\right\|_1\,,\qquad\left\|\mathcal{F}\right\|_\diamond=\max_{\rho\in\mathcal{S}_2}\left\|\left(\mathcal{F}\otimes\mathbb{I}_{\mathcal{O}_1}\right)(\rho)\right\|_1\,. \tag{116}$$

*Therefore, for any $\rho,\,\sigma\in\mathcal{S}_n$,*

$$\left\|\Phi^{\otimes n}(\rho)-\Phi^{\otimes n}(\sigma)\right\|_{W_1}\le\left\|\Phi-\mathcal{E}\right\|_\diamond\left\|\rho-\sigma\right\|_{W_1}\,. \tag{117}$$

*Proof.* Let $X\in\mathcal{N}_n$. Then, $\mathrm{Tr}_iX=0$ for some $i\in[n]$. Without loss of generality, we can assume that $i=1$. Since $\mathrm{Tr}_1\Phi^{\otimes n}(X)=0$, we have from (35)

$$\begin{aligned}\left\|\Phi^{\otimes n}(X)\right\|_{W_1}&=\frac{1}{2}\left\|\Phi^{\otimes n}(X)\right\|_1\le\frac{1}{2}\left\|\left(\Phi\otimes\mathbb{I}_{\mathcal{O}_{n-1}}\right)(X)\right\|_1=\frac{1}{2}\left\|\left((\Phi-\mathcal{E})\otimes\mathbb{I}_{\mathcal{O}_{n-1}}\right)(X)\right\|_1\\&\le\frac{1}{2}\left\|\Phi-\mathcal{E}\right\|_\diamond\left\|X\right\|_1\le\left\|\Phi-\mathcal{E}\right\|_\diamond\,,\end{aligned} \tag{118}$$

where we have also used that $\left(\mathcal{E}\otimes\mathbb{I}_{\mathcal{O}_{n-1}}\right)(X)=0$, therefore

$$\left\|\Phi^{\otimes n}\right\|_{W_1\to W_1}\le\left\|\Phi-\mathcal{E}\right\|_\diamond\,. \tag{119}$$

Let $\rho\in\mathcal{S}_1$, and let

$$X=(\rho-\omega)\otimes\omega^{\otimes(n-1)}\in\mathcal{N}_n\,. \tag{120}$$

We have

$$\left\|\Phi^{\otimes n}(X)\right\|_{W_1}=\frac{1}{2}\left\|\Phi^{\otimes n}(X)\right\|_1=\frac{1}{2}\left\|\Phi(\rho)-\omega\right\|_1=\frac{1}{2}\left\|\left(\Phi-\mathcal{E}\right)(\rho)\right\|_1\,, \tag{121}$$

therefore

$$\left\|\Phi^{\otimes n}\right\|_{W_1\to W_1}\ge\frac{1}{2}\left\|\Phi-\mathcal{E}\right\|_{1\to1}\,, \tag{122}$$

and the claim follows. ∎

We consider the quantum amplitude damping channel as example of application of Proposition 11:

**Example 1** (Amplitude damping channel)**.** Let $d = 2$, and for any $0 \leq p \leq 1$, let $\Phi_p$ be the quantum amplitude damping channel with decay probability $1 - p$ whose action on the Pauli matrices is

$$\Phi_p(\mathbb{I}_2) = \mathbb{I}_2 + (1-p)\,\sigma_z\,, \qquad \Phi_p(\sigma_x) = \sqrt{p}\,\sigma_x\,, \qquad \Phi_p(\sigma_y) = \sqrt{p}\,\sigma_y\,, \qquad \Phi_p(\sigma_z) = p\,\sigma_z\,. \tag{123}$$

Then, for any $0 \leq p \leq 1/5$,

$$\frac{1}{2}\sqrt{\frac{p}{1-p}} \leq \left\|\Phi_p^{\otimes n}\right\|_{W_1\to W_1} \leq 2\sqrt{\frac{p}{1-p}} \tag{124}$$

(For $1/5 \leq p \leq 1$, the upper bound of (124) is trivial), and for any $\rho,\,\sigma \in \mathcal{S}_n$,

$$\left\|\Phi_p^{\otimes n}(\rho) - \Phi_p^{\otimes n}(\sigma)\right\|_{W_1} \leq 2\sqrt{\frac{p}{1-p}}\,\|\rho-\sigma\|_{W_1}\,. \tag{125}$$

*Proof.* The only fixed quantum state of $\Phi_p$ is

$$\omega = \frac{\mathbb{I}_2 + \sigma_z}{2}\,. \tag{126}$$

We parameterize a pure state $\rho \in \mathcal{S}_1$ as

$$\rho = \frac{\mathbb{I}_2 + v_x\,\sigma_x + v_y\,\sigma_y + v_z\,\sigma_z}{2}\,, \tag{127}$$

where $v$ is a unit vector in $\mathbb{R}^3$. We have

$$\begin{aligned}\|\Phi_p(\rho) - \omega\|_1 &= \frac{\sqrt{p}}{2}\,\|v_x\,\sigma_x + v_y\,\sigma_y + \sqrt{p}\,(v_z-1)\,\sigma_z\|_1 = \sqrt{p\left(v_x^2 + v_y^2 + p\,(1-v_z)^2\right)}\\ &= \sqrt{p\,(1-v_z)\,(1+p+(1-p)\,v_z)} \leq \sqrt{\frac{p}{1-p}}\,,\end{aligned} \tag{128}$$

where we have used that $v^2 = 1$. Let $\mathcal{E}$ be the quantum channel on $\mathbb{C}^2$ that replaces the input state with $\omega$. We have

$$\|\Phi_p - \mathcal{E}\|_\diamond \leq 2\,\|\Phi_p - \mathcal{E}\|_{1\to1} = 2\sqrt{\frac{p}{1-p}}\,, \tag{129}$$

and the claim follows from Proposition 11. □

We can determine exactly the quantum coefficient of the quantum depolarizing channel with respect to the quantum $W_1$ distance:

**Proposition 12** (Quantum depolarizing channel)**.** *Let $\omega \in \mathcal{S}_1$, and for any $0 \le p \le 1$, let $\mathcal{E}_p$ be the quantum channel on $\mathbb{C}^d$ that is the identity with probability $p$ and replaces the input state with $\omega$ with probability $1-p$:*

$$\mathcal{E}_p(X) = p\,X + (1-p)\,\omega\,\mathrm{Tr}\,X\,, \qquad X \in \mathcal{O}_1\,. \tag{130}$$

*Then,*

$$\left\|\mathcal{E}_p^{\otimes n}\right\|_{W_1\to W_1} = p\,, \tag{131}$$

*and for any $\rho$, $\sigma \in \mathcal{S}_n$,*

$$\left\|\mathcal{E}_p^{\otimes n}(\rho) - \mathcal{E}_p^{\otimes n}(\sigma)\right\|_{W_1} \le p\,\|\rho-\sigma\|_{W_1}\,. \tag{132}$$

*Proof.* Let $X \in \mathcal{N}_n$, and let $i \in [n]$ be such that $\mathrm{Tr}_i X = 0$. Without loss of generality, we can assume that $i = 1$. We then have from (35)

$$\left\|\mathcal{E}_p^{\otimes n}(X)\right\|_{W_1} = \frac{1}{2}\left\|\mathcal{E}_p^{\otimes n}(X)\right\|_1 = \frac{p}{2}\left\|\left(\mathbb{I}_{\mathcal{O}_1}\otimes\mathcal{E}_p^{\otimes(n-1)}\right)(X)\right\|_1 \le \frac{p}{2}\|X\|_1 \le p\,, \tag{133}$$

therefore

$$\left\|\mathcal{E}_p^{\otimes n}\right\|_{W_1\to W_1} \le p\,. \tag{134}$$

On the other hand,

$$X = (|1\rangle\langle 1| - |2\rangle\langle 2|)\otimes\omega^{\otimes(n-1)} \tag{135}$$

achieves equality in (133), and the claim follows. □

## 9.2 Shallow quantum circuits

Quantum channels acting on multiple qudits can in general expand the quantum $W_1$ distance. In the following Proposition 13, we prove that the expansion factor is bounded by the size of the light-cones of the input qudits, which can be easily bounded if the channel can be implemented by a shallow local quantum circuit:

**Proposition 13.** *Let $\Phi : \mathcal{O}_n \to \mathcal{O}_m$ be a quantum channel. For any $i \in [n]$, let $\mathcal{I}_i \subseteq [m]$ be the light-cone of the $i$-th qudit, i.e., the minimum subset of qudits such that $\mathrm{Tr}_{\mathcal{I}_i}\Phi(X) = 0$ for any $X \in \mathcal{O}_n$ such that $\mathrm{Tr}_i X = 0$. Then,*

$$\|\Phi\|_{W_1\to W_1} \le 2\,\frac{d^2-1}{d^2}\max_{i\in[n]}|\mathcal{I}_i|\,, \tag{136}$$

*and for any $\rho$, $\sigma \in \mathcal{S}_n$*

$$\|\Phi(\rho)-\Phi(\sigma)\|_{W_1} \le 2\,\frac{d^2-1}{d^2}\max_{i\in[n]}|\mathcal{I}_i|\,\|\rho-\sigma\|_{W_1}\,. \tag{137}$$

*Proof.* Let $X \in \mathcal{N}_n$. Then, $\mathrm{Tr}_i X = 0$ for some $i \in [n]$, hence $\mathrm{Tr}_{\mathcal{I}_i}\Phi(X) = 0$. Proposition 4 implies

$$\|\Phi(X)\|_{W_1} \le |\mathcal{I}_i|\,\frac{d^2-1}{d^2}\,\|\Phi(X)\|_1 \le 2\,\frac{d^2-1}{d^2}\max_{i\in[n]}|\mathcal{I}_i|\,, \tag{138}$$

and the claim follows. □

# 10 Future perspectives

We have proposed a quantum generalization of the $W_1$ distance which recovers the classical $W_1$ distance as a special case and keeps most of its properties, among which the continuity of the entropy. 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.

- **Quantum state estimation**

  Estimating a quantum state of $n$ qudits up to $o(1)$ error in the trace distance is a notoriously difficult task, since the number of required copies of the state grows exponentially with $n$ [63]. Requiring instead the quantum $W_1$ distance between the true quantum state and its estimate to be $o(n)$ is a much weaker condition, and the number of required copies can be much smaller. Therefore, in all the situations where a precision guarantee in terms of the quantum $W_1$ distance is sufficient, employing this distance rather than the trace distance can lead to a significant improvement to the complexity of the estimate.

- **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 [64], and the same property should be desirable also when the machine learning algorithm is quantum [65].

  *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 [66], 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, and when the input is a bit string the size of the perturbations is measured with the Hamming distance. Since the proposed quantum $W_1$ distance recovers the Hamming distance for vectors of the canonical basis, it is a perfect candidate to measure the size of the perturbations for quantum algorithms for machine learning with a quantum input. Therefore, the proposed quantum $W_1$ distance 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 [65, 67]. In particular, it is reasonable to require the encoding to be robust with respect to small perturbations of the input. The trivial encoding maps each bit string to the corresponding computational basis state, and is not continuous with respect to any unitarily invariant distance, since any bit flip on the input transforms the quantum state into an orthogonal state. On the contrary, the trivial encoding is continuous with respect to the proposed quantum $W_1$ distance, since it recovers the Hamming distance for vectors of the canonical basis. Therefore, the quantum $W_1$ distance provides a natural measure for the size of the input perturbations and hence for the robustness of the encoding, favoring encodings that map classical inputs with small Hamming distance into quantum states with small quantum $W_1$ distance.

- **Quantum Generative Adversarial Networks**

  In analogy to classical GANs, quantum GANs [37] 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 best 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 [38, 68]. 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 [11]. 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 [69–76] 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 recovers the Hamming distance for vectors of the canonical basis, and is therefore a candidate to extend quantum rate distortion theory beyond the iid regime.

- **Quantum differential privacy**

  A quantum measurement is gentle if the pre- and post-measurement states are close. Ref. [77] defines a measurement of the state of $n$ qudits to be differentially private if the probability distributions of the outcome of the measurement performed on any couple of neighboring states are close, i.e., if the measurement cannot distinguish between any two neighboring states. For product states, the two properties above are intimately connected: any measurement that is gentle on product states is also differentially private and vice versa. The proposed quantum $W_1$ distance can be thought as a generalization of the notion of neighboring quantum states, and is therefore a candidate to extend beyond product states the connection between quantum differential privacy and gentleness.

- **Mixing time of quantum Markov semigroups**

  In section 9, we have determined upper bounds to the contraction coefficient of the $n$-th tensor power of a one-qudit quantum channel with respect to the proposed quantum $W_1$ distance and we have shown that, in contrast to the situation for the trace distance, such coefficient remains nontrivial in the limit $n \to \infty$. It is natural to generalize these observations and consider the mixing times of general quantum Markov semigroups with respect to the quantum $W_1$ distance. A nice property of this approach, in contrast to the bounds derived using the quantum relative entropy, is that the stationary state of the quantum Markov process does not need to have full rank.

- **Shallow quantum circuits**

  The Hamming distance plays a key role in the study of the computational capabilities of quantum circuits [78,79]. 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 distance might be useful in characterizing the states generated by constant depth circuits, and it may be able to extend the current results on their computational capabilities.

- **Quantum many-body Hamiltonians**

  In Proposition 10, we have proved that local quantum Hamiltonians have a small quantum Lipschitz constant. Therefore, the notion of quantum Lipschitz constant can provide a generalization of the notion of local Hamiltonian and lead to the consequent extension of the related properties.


# Acknowledgements

GdP, MM and SL have been supported by the USA Air Force Office of Scientific Research, the USA Army Research Office under the Blue Sky program, DOE and IARPA. DT has been partially supported by the University of Pisa, Project PRA 2018-49 and INdAM, GNAMPA Project 2020 “Problemi di ottimizzazione con vincoli via trasporto ottimo e incertezza”.


# A Alternative definition of neighboring quantum states

Ref. [77] defines the quantum states $\rho$, $\sigma \in \mathcal{S}_n$ to be neighboring if there is a quantum channel $\Phi$ on $\mathcal{H}_n$ that acts on only one qudit and such that either $\rho = \Phi(\sigma)$ or $\sigma = \Phi(\rho)$. Proceeding along the same lines of section 3, this alternative definition of neighboring quantum states induces an alternative quantum $W_1$ norm $\|\cdot\|_{\tilde{W}_1}$. In the following Proposition 14, we prove that the norms $\|\cdot\|_{\tilde{W}_1}$ and $\|\cdot\|_{W_1}$ are equivalent.

**Proposition 14.** *For any $X \in \mathcal{O}_n^T$,*

$$\|X\|_{W_1} \leq \|X\|_{\tilde{W}_1} \leq 2\,\frac{d^2-1}{d^2}\,\|X\|_{W_1}\,. \tag{139}$$

*Proof.* Let

$$\tilde{\mathcal{N}}_n = \{\pm\left(\rho - \Phi(\rho)\right) : \rho \in \mathcal{S}_n,\ \Phi\,\text{quantum channel acting on one qudit}\} \tag{140}$$

be the set of the differences between couples of neighboring states according to the alternative definition. Since $\tilde{\mathcal{N}}_n \subseteq \mathcal{N}_n$, we have

$$\|X\|_{W_1} \leq \|X\|_{\tilde{W}_1}\,. \tag{141}$$

Let $\rho,\,\sigma \in \mathcal{S}_n$ such that $\rho - \sigma \in \mathcal{N}_n$. Then, there is $i \in [n]$ such that $\mathrm{Tr}_i\rho = \mathrm{Tr}_i\sigma$. Without loss of generality, we can assume that $i = 1$. We have

$$\|\rho - \sigma\|_{\tilde{W}_1} \leq \left\|\rho - \frac{\mathbb{I}_d}{d}\otimes\mathrm{Tr}_1\rho\right\|_{\tilde{W}_1} + \left\|\frac{\mathbb{I}_d}{d}\otimes\mathrm{Tr}_1\sigma - \sigma\right\|_{\tilde{W}_1}\,. \tag{142}$$

Let $U^{(1)},\,\ldots,\,U^{\left(d^2\right)}$ be as in (166). Then,

$$\begin{aligned}\left\|\rho - \frac{\mathbb{I}_d}{d}\otimes\mathrm{Tr}_1\rho\right\|_{\tilde{W}_1} &= \frac{1}{d^2}\left\|\sum_{i=2}^{d^2}\left(\rho - \left(U^{(i)}\otimes\mathbb{I}_d^{\otimes(n-1)}\right)\rho\left(U^{(i)\dagger}\otimes\mathbb{I}_d^{\otimes(n-1)}\right)\right)\right\|_{\tilde{W}_1}\\ &\leq \frac{1}{d^2}\sum_{i=2}^{d^2}\left\|\rho - \left(U^{(i)}\otimes\mathbb{I}_d^{\otimes(n-1)}\right)\rho\left(U^{(i)\dagger}\otimes\mathbb{I}_d^{\otimes(n-1)}\right)\right\|_{\tilde{W}_1} \leq \frac{d^2-1}{d^2}\,,\end{aligned} \tag{143}$$

therefore

$$\|\rho - \sigma\|_{\tilde{W}_1} \leq 2\,\frac{d^2-1}{d^2}\,. \tag{144}$$

Then,

$$\|X\|_{\tilde{W}_1} \leq 2\,\frac{d^2-1}{d^2}\,\|X\|_{W_1}\,, \tag{145}$$

and the claim follows. □

# B Efficient estimation of the quantum Lipschitz constant

The following Proposition 15 provides an estimate of the quantum Lipschitz constant up to a multiplicative error $\sqrt{2}$ that does not require any optimization.

**Proposition 15** (Efficient estimation of the quantum Lipschitz constant)**.** *For any $i \in [n]$, let $\mathcal{E}_i$ be the quantum channel on $\mathcal{H}_n$ that replaces the state of the i-th qudit with the maximally mixed state. Then, for any $H \in \mathcal{O}_n$,*

$$\frac{d^2}{d^2-1}\max_{i\in[n]}\|H-\mathcal{E}_i(H)\|_\infty \le \|H\|_L \le 2\max_{i\in[n]}\|H-\mathcal{E}_i(H)\|_\infty \,. \tag{146}$$

*Proof.* Let $X \in \mathcal{N}_n$. Then, there exists $i \in [n]$ such that $\mathrm{Tr}_i X = 0$, and

$$0 = \mathrm{Tr}\left[H\,\mathcal{E}_i(X)\right] = \mathrm{Tr}\left[\mathcal{E}_i(H)\,X\right] . \tag{147}$$

We then have

$$\mathrm{Tr}\left[H\,X\right] = \mathrm{Tr}\left[\left(H-\mathcal{E}_i(H)\right)X\right] \le \|H-\mathcal{E}_i(H)\|_\infty\,\|X\|_1 \le 2\,\|H-\mathcal{E}_i(H)\|_\infty \,, \tag{148}$$

therefore

$$\|H\|_L \le 2\max_{i\in[n]}\|H-\mathcal{E}_i(H)\|_\infty \,. \tag{149}$$

Let $i \in [n]$, and let $\rho \in \mathcal{S}_n$ such that

$$\left|\mathrm{Tr}\left[H\left(\rho-\mathcal{E}_i(\rho)\right)\right]\right| = \left|\mathrm{Tr}\left[\left(H-\mathcal{E}_i(H)\right)\rho\right]\right| = \|H-\mathcal{E}_i(H)\|_\infty \,. \tag{150}$$

From (35) and Lemma 3 of Appendix C,

$$\|\rho-\mathcal{E}_i(\rho)\|_{W_1} = \frac{1}{2}\,\|\rho-\mathcal{E}_i(\rho)\|_1 \le \frac{d^2-1}{d^2}\,, \tag{151}$$

therefore

$$\|H\|_L \ge \frac{d^2}{d^2-1}\,\|H-\mathcal{E}_i(H)\|_\infty \,, \tag{152}$$

and the claim follows. □

# C Lemmas

**Lemma 1.** *For any $X \in \mathcal{O}_n^T$,*

$$\|X\|_{W_1} = \frac{1}{2}\min\left(\sum_{i=1}^n \left\|X^{(i)}\right\|_1 : X^{(i)} \in \mathcal{O}_n^T,\ \mathrm{Tr}_i\,X^{(i)} = 0,\ X = \sum_{i=1}^n X^{(i)}\right) . \tag{153}$$

*Proof.* Throughout this proof, $\|\cdot\|_{W_1}$ denotes the norm defined in (153). The optimization in (153) is performed over a compact set, therefore the minimum is achieved. To prove the claim, it is sufficient to prove that the unit ball of $\|\cdot\|_{W_1}$ coincides with $\mathcal{B}_n$.

On the one hand, let $X \in \mathcal{B}_n$. Since each $\mathcal{N}_n^{(i)}$ is convex, $X$ is a convex combination of $n$ elements, each belonging to the corresponding $\mathcal{N}_n^{(i)}$, i.e., there exists a probability distribution $p$ on $[n]$ such that

$$X = \sum_{i=1}^n p_i\left(\rho^{(i)}-\sigma^{(i)}\right) , \qquad \rho^{(i)},\, \sigma^{(i)} \in \mathcal{S}_n \,, \qquad \mathrm{Tr}_i\rho^{(i)} = \mathrm{Tr}_i\sigma^{(i)} \,. \tag{154}$$

Therefore, choosing in (153)

$$X^{(i)} = p_i \left(\rho^{(i)} - \sigma^{(i)}\right) , \tag{155}$$

we get

$$\|X\|_{W_1} \le \frac{1}{2}\sum_{i=1}^n p_i \left\|\rho^{(i)} - \sigma^{(i)}\right\|_1 \le 1 , \tag{156}$$

and $X$ belongs to the unit ball of $\|\cdot\|_{W_1}$.

On the other hand, let $X \in \mathcal{O}_n^T$ such that $\|X\|_{W_1} = 1$. Then, there exist $X^{(1)}, \ldots X^{(n)}$ as in (153) such that

$$\frac{1}{2}\sum_{i=1}^n \left\|X^{(i)}\right\|_1 = 1 . \tag{157}$$

For any $i \in [n]$, let

$$p_i = \frac{1}{2}\left\|X^{(i)}\right\|_1 , \tag{158}$$

such that $p$ is a probability distribution on $[n]$. We can express each $X^{(i)}$ as

$$X^{(i)} = p_i \left(\rho^{(i)} - \sigma^{(i)}\right) , \tag{159}$$

where $\rho^{(i)}$, $\sigma^{(i)} \in \mathcal{S}_n$ have orthogonal supports. Since $\mathrm{Tr}_i X^{(i)} = 0$, $\rho^{(i)}$ and $\sigma^{(i)}$ are neighboring, and $\rho^{(i)} - \sigma^{(i)} \in \mathcal{N}_n$. Since

$$X = \sum_{i=1}^n X^{(i)} = \sum_{i=1}^n p_i \left(\rho^{(i)} - \sigma^{(i)}\right) , \tag{160}$$

$X \in \mathcal{B}_n$, and the claim follows. □

**Lemma 2.** *For any $X \in \mathcal{O}_m^T$ and any $Y \in \mathcal{O}_n$,*

$$\|X \otimes Y\|_{W_1} \le \|X\|_{W_1} \|Y\|_1 . \tag{161}$$

*Proof.* Let $X^{(1)}, \ldots, X^{(m)} \in \mathcal{O}_m^T$ such that

$$\mathrm{Tr}_i X^{(i)} = 0 \quad \forall\, i \in [m] , \qquad X = \sum_{i=1}^m X^{(i)} . \tag{162}$$

Since

$$X \otimes Y = \sum_{i=1}^m X^{(i)} \otimes Y , \tag{163}$$

we get

$$\|X \otimes Y\|_{W_1} \le \frac{1}{2}\sum_{i=1}^m \left\|X^{(i)}\right\|_1 \|Y\|_1 , \tag{164}$$

and the claim follows. □

**Lemma 3.** *For any $X \in \mathcal{O}_n$,*

$$\left\| X - \frac{\mathbb{I}_d}{d} \otimes \mathrm{Tr}_1 X \right\|_1 \leq 2\, \frac{d^2-1}{d^2} \left\| X \right\|_1 . \tag{165}$$

*Proof.* Let $U^{(1)}, \ldots, U^{\left(d^2\right)}$ be a set of unitary operators on $\mathbb{C}^d$ such that

$$U^{(1)} = \mathbb{I}_d\,, \qquad \mathrm{Tr}\left[{U^{(i)}}^\dagger U^{(j)}\right] = d\,\delta_{ij}\,, \qquad i,\, j \in \left[d^2\right] . \tag{166}$$

Then,

$$\begin{aligned} \left\| X - \frac{\mathbb{I}_d}{d} \otimes \mathrm{Tr}_1 X \right\|_1 &= \frac{1}{d^2} \left\| \left(d^2-1\right) X - \sum_{i=2}^{d^2} \left(U^{(i)} \otimes \mathbb{I}_d^{\otimes(n-1)}\right) X \left({U^{(i)}}^\dagger \otimes \mathbb{I}_d^{\otimes(n-1)}\right) \right\|_1 \\ &\leq 2\, \frac{d^2-1}{d^2} \left\| X \right\|_1 , \end{aligned} \tag{167}$$

and the claim follows. $\square$

**Lemma 4.** *Let $\gamma$ be a maximally entangled state of $\left(\mathbb{C}^d\right)^{\otimes 2}$. Then, for any even $n$,*

$$\left\| \gamma^{\otimes \frac{n}{2}} - \frac{\mathbb{I}_d^{\otimes n}}{d^n} \right\|_{W_1} = \frac{n}{2}\, \frac{d^2-1}{d^2} . \tag{168}$$

*Proof.* From Proposition 3, it is sufficient to prove the claim for $n = 2$. Since

$$\mathrm{Tr}_1 \gamma = \frac{\mathbb{I}_d}{d}\,, \tag{169}$$

we have from (35)

$$\left\| \gamma - \frac{\mathbb{I}_d^{\otimes 2}}{d^2} \right\|_{W_1} = \frac{1}{2} \left\| \gamma - \frac{\mathbb{I}_d^{\otimes 2}}{d^2} \right\|_1 = \frac{d^2-1}{d^2}\,, \tag{170}$$

and the claim follows. $\square$

**Lemma 5.** *Let $\rho$ and $\sigma$ be quantum states of the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ such that $\rho_B = \sigma_B$. Then,*

$$\left| S(\rho) - S(\sigma) \right| \leq 2 \ln \dim \mathcal{H}_A\,. \tag{171}$$

*Proof.* We have

$$\begin{aligned} \left| S(\rho) - S(\sigma) \right| &= \left| S(\rho_B) + S(A|B)_\rho - S(\sigma_B) - S(A|B)_\sigma \right| = \left| S(A|B)_\rho - S(A|B)_\sigma \right| \\ &\leq \left| S(A|B)_\rho \right| + \left| S(A|B)_\sigma \right| \leq 2 \ln \dim \mathcal{H}_A\,, \end{aligned} \tag{172}$$

and the claim follows. $\square$

**Lemma 6.** *Let $p$ and $q$ be probability distributions on $[d]^n$ whose marginals over the first $n-1$ components coincide, i.e., such that $p(x_1 \ldots x_{n-1}) = q(x_1 \ldots x_{n-1})$ for any $x_1, \ldots, x_{n-1} \in [d]$. Then,*

$$W_1(p, q) \le 1 \,. \tag{173}$$

*Proof.* We have for any $x \in [d]^n$

$$\begin{aligned} p(x) &= p(x_1 \ldots x_{n-1})\, p(x_n|x_1 \ldots x_{n-1}) \,, \\ q(x) &= q(x_1 \ldots x_{n-1})\, q(x_n|x_1 \ldots x_{n-1}) = p(x_1 \ldots x_{n-1})\, q(x_n|x_1 \ldots x_{n-1}) \,. \end{aligned} \tag{174}$$

Since the $W_1$ distance is jointly convex, we have

$$W_1(p, q) \le \sum_{x_1,\, \ldots,\, x_{n-1} \in [d]} p(x_1 \ldots x_{n-1})\, W_1\left(p(\cdot|x_1 \ldots x_{n-1}),\, q(\cdot|x_1 \ldots x_{n-1})\right) \le 1 \,, \tag{175}$$

and the claim follows. □

# D Proof of the $W_1$ continuity of the Shannon entropy

Let $X$, $Y$ be random variables with values in $[d]^n$ whose joint probability distribution is the optimal coupling between $p$ and $q$. For any $i \in [n]$, let $p_i$ be the probability that $X_i \neq Y_i$, such that

$$W_1(p, q) = \sum_{i=1}^n p_i \,. \tag{176}$$

We have

$$\begin{aligned} S(X) - S(Y) &\overset{\text{(a)}}{\le} S(XY) - S(Y) = S(X|Y) \overset{\text{(b)}}{\le} \sum_{i=1}^n S(X_i|Y) \overset{\text{(c)}}{\le} \sum_{i=1}^n S(X_i|Y_i) \\ &\overset{\text{(d)}}{\le} \sum_{i=1}^n \left(h_2(p_i) + p_i \ln{(d-1)}\right) \overset{\text{(e)}}{\le} n\, h_2\left(\frac{1}{n}\sum_{i=1}^n p_i\right) + \ln{(d-1)} \sum_{i=1}^n p_i \\ &= n\, h_2\left(\frac{W_1(p,q)}{n}\right) + W_1(p,q) \ln{(d-1)} \ , \end{aligned} \tag{177}$$

where (a) follows from the monotonicity of the Shannon entropy, (b) and (c) follow from the strong subadditivity of the Shannon entropy, (d) follows from Fano's inequality and (e) follows from Jensen's inequality applied to the concave function $h_2$. The claim follows.

# 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] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.

[8] Leonid Vitalievich Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

[9] Anatoly Moiseevich Vershik. Long history of the Monge-Kantorovich transportation problem. *The Mathematical Intelligencer*, 35(4):1–9, 2013.

[10] Donald S Ornstein. An application of ergodic theory to probability theory. *The Annals of Probability*, 1(1):43–58, 1973.

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

[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] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[14] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of Wasserstein GANs. In *Advances in neural information processing systems*, pages 5767–5777, 2017.

[15] Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics*, pages 1608–1617, 2018.

[16] Eric A Carlen and Jan Maas. An analog of the 2-Wasserstein metric in non-commutative probability under which the Fermionic Fokker–Planck equation is gradient flow for the entropy. *Communications in Mathematical Physics*, 331(3):887–926, 2014.

[17] Eric A Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017.

[18] Eric A Carlen and Jan Maas. Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. *Journal of Statistical Physics*, 178(2):319–378, 2020.

[19] Cambyse Rouzé and Nilanjana Datta. Concentration of quantum states from quantum functional and transportation cost inequalities. *Journal of Mathematical Physics*, 60(1):012202, 2019.

[20] Nilanjana Datta and Cambyse Rouzé. Relating relative entropy, optimal transport and Fisher information: A quantum HWI inequality. *Annales Henri Poincaré*, 21:2115–2150, 2020.

[21] Tan Van Vu and Yoshihiko Hasegawa. Geometrical Bounds of the Irreversibility in Markovian Systems. *arXiv preprint arXiv:2005.02871*, 2020.

[22] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[23] Giacomo De Palma and Stefan Huber. The conditional Entropy Power Inequality for quantum additive noise channels. *Journal of Mathematical Physics*, 59(12):122201, 2018.

[24] Yongxin Chen, Tryphon T Georgiou, Lipeng Ning, and Allen Tannenbaum. Matricial Wasserstein-1 distance. *IEEE control systems letters*, 1(1):14–19, 2017.

[25] Ernest K Ryu, Yongxin Chen, Wuchen Li, and Stanley Osher. Vector and matrix optimal mass transport: theory, algorithm, and applications. *SIAM Journal on Scientific Computing*, 40(5):A3675–A3698, 2018.

[26] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Matrix optimal mass transport: a quantum mechanical approach. *IEEE Transactions on Automatic Control*, 63(8):2612–2619, 2018.

[27] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Wasserstein geometry of quantum states and optimal transport of matrix-valued measures. In *Emerging Applications of Control and Systems Theory*, pages 139–150. Springer, 2018.

[28] Julián Agredo. A Wasserstein-type distance to measure deviation from equilibrium of quantum Markov semigroups. *Open Systems & Information Dynamics*, 20(02):1350009, 2013.

[29] J Agredo. On exponential convergence of generic quantum Markov semigroups in a Wasserstein-type distance. *International Journal of Pure and Applied Mathematics*, 107(4):909–925, 2016.

[30] Kazuki Ikeda. Foundation of quantum optimal transport and applications. *Quantum Information Processing*, 19(1):25, 2020.

[31] François Golse, Clément Mouhot, and Thierry Paul. On the mean field and classical limits of quantum mechanics. *Communications in Mathematical Physics*, 343(1):165–205, 2016.

[32] Emanuele Caglioti, François Golse, and Thierry Paul. Towards optimal transport for quantum densities. preprint, Dec 2018.

[33] François Golse. The quantum N-body problem in the mean-field and semiclassical regime. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2118):20170229, 2018.

[34] François Golse and Thierry Paul. The Schrödinger equation in the mean-field and semiclassical regime. *Archive for Rational Mechanics and Analysis*, 223(1):57–94, 2017.

[35] François Golse and Thierry Paul. Wave packets and the quadratic Monge–Kantorovich distance in quantum mechanics. *Comptes Rendus Mathematique*, 356(2):177–197, 2018.

[36] Emanuele Caglioti, François Golse, and Thierry Paul. Quantum Optimal Transport is Cheaper. *Journal of Statistical Physics*, 2020.

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

[38] 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.

[39] Giacomo De Palma and Dario Trevisan. Quantum optimal transport with quantum channels. *arXiv preprint arXiv:1911.00803*, 2019.

[40] Rocco Duvenhage and Machiel Snyman. Balance between quantum Markov semigroups. *Annales Henri Poincaré*, 19(6):1747–1786, 2018.

[41] J Agredo and Franco Fagnola. On quantum versions of the classical Wasserstein distance. *Stochastics*, 89(6-7):910–922, 2017.

[42] Karol Zyczkowski and Wojeciech Slomczynski. The Monge distance between quantum states. *Journal of Physics A: Mathematical and General*, 31(45):9095, 1998.

[43] Karol Zyczkowski and Wojciech Slomczynski. The Monge metric on the sphere and geometry of quantum states. *Journal of Physics A: Mathematical and General*, 34(34):6689, 2001.

[44] Ingemar Bengtsson and Karol Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2017.

[45] Tobias Osborne and Andreas Winter. A quantum generalisation of Talagrand's inequality. *Tobias J. Osborne's research notes*, 2009.

[46] Thomas Vidick and Henry Yuen. A simple proof of Renner's exponential de Finetti theorem. *arXiv preprint arXiv:1608.04814*, 2016.

[47] Alain Connes and John Lott. The metric aspect of noncommutative geometry. In *New symmetry principles in quantum field theory*, pages 53–93. Springer, 1992.

[48] Pierre Martinetti. Connes distance and optimal transport. *Journal of Physics: Conference Series*, 968:012007, Feb 2018.

[49] Philippe Biane and Dan Voiculescu. A free probability analogue of the Wasserstein metric on the trace-state space. *Geometric & Functional Analysis GAFA*, 11(6):1125–1138, 2001.

[50] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS Books in Mathematics. Springer New York, 2013.

[51] Yury Polyanskiy and Yihong Wu. Wasserstein continuity of entropy and outer bounds for interference channels. *IEEE Transactions on Information Theory*, 62(7):3992–4002, 2016.

[52] Igor Devetak and Andreas Winter. Distillation of secret key and entanglement from quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and engineering sciences*, 461(2053):207–235, 2005.

[53] Katalin Marton. A simple proof of the blowing-up lemma (corresp.). *IEEE Transactions on Information Theory*, 32(3):445–446, 1986.

[54] Katalin Marton. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *The Annals of Probability*, 24(2):857–866, 1996.

[55] Michel Talagrand. Transportation cost for Gaussian and other product measures. *Geometric & Functional Analysis GAFA*, 6(3):587–600, 1996.

[56] Marius Junge and Qiang Zeng. Noncommutative martingale deviation and Poincaré type inequalities with applications. *Probability Theory and Related Fields*, 161(3-4):449–507, 2015.

[57] Michel Talagrand. A new look at independence. *The Annals of probability*, pages 1–34, 1996.

[58] N. Gozlan and C. Leonard. Transport Inequalities. A Survey. *Markov Processes and Related Fields*, 16(4):635–736, 2010.

[59] Maxim Raginsky and Igal Sason. Concentration of Measure Inequalities in Information Theory, Communications, and Coding. *Foundations and Trends in Communications and Information Theory*, 10(1-2):1–246, 2013.

[60] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[61] David Reeb, Michael J Kastoryano, and Michael M Wolf. Hilbert's projective metric in quantum information theory. *Journal of mathematical physics*, 52(8):082201, 2011.

[62] Michael J Kastoryano and Kristan Temme. Quantum logarithmic Sobolev inequalities and rapid mixing. *Journal of Mathematical Physics*, 54(5):052202, 2013.

[63] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, 2007.

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

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

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

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

[68] Simon Becker and Wuchen Li. Quantum statistical learning via Quantum Wasserstein natural gradient. *arXiv preprint arXiv:2008.11135*, 2020.

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

[70] 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.

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

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] Scott Aaronson and Guy N Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, 2019.

[78] 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.

[79] 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.