Quantum Monge-Kantorovich Problem and Transport Distance between Density Matrices

Shmuel Friedland,[1] Michał Eckstein,[2,*] Sam Cole,[3] and Karol Życzkowski[2,4]
[1]Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, Illinois 60607-7045, USA
[2]Institute of Theoretical Physics, Jagiellonian University and Mark Kac Center, ul. Łojasiewicza 11, 30–348 Kraków, Poland
[3]Department of Mathematics, University of Missouri, Columbia, Missouri 65211, USA
[4]Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland

(Received 23 February 2021; revised 1 August 2022; accepted 2 August 2022; published 7 September 2022)

A quantum version of the Monge-Kantorovich optimal transport problem is analyzed. The transport cost is minimized over the set of all bipartite coupling states ρ^{AB} such that both of its reduced density matrices ρ^A and ρ^B of dimension N are fixed. We show that, selecting the quantum cost matrix to be proportional to the projector on the antisymmetric subspace, the minimal transport cost leads to a semidistance between ρ^A and ρ^B, which is bounded from below by the rescaled Bures distance and from above by the root infidelity. In the single-qubit case, we provide a semianalytic expression for the optimal transport cost between any two states and prove that its square root satisfies the triangle inequality and yields an analog of the Wasserstein distance of the order of 2 on the set of density matrices. We introduce an associated measure of proximity of quantum states, called SWAP fidelity, and discuss its properties and applications in quantum machine learning.

DOI: 10.1103/PhysRevLett.129.110402

Introduction.—Remarkable progress in quantum technologies stimulates further research on foundations of quantum mechanics. In particular, one aims to improve our understanding of the structure of the set of quantum states [1,2]—the arena in which quantum information processing takes place. It is, therefore, important to analyze various distances in the space of quantum states and to describe their properties and diverse physical applications.

In the classical case, one considers several distances in the space of probability distributions. A prominent role is played by the Monge distance, directly linked to the famous mass transport problem [3], in which one minimizes the work against friction required to move a pile of earth of shape p^A into the final shape p^B. For continuous distributions, the problem is solved analytically in any 1D case [4] and in several particular 2D cases [5], while effective numerical algorithms can be applied for any discrete probability distributions.

More general formulations of Kantorovich [6,7] and Wasserstein [8], relying on joint probability distributions with marginals p^A and p^B, are explicitly symmetric with respect to given probability distributions. Because of numerous applications of the mass transport problem in operations research and economics and its relation to the assignment problem, it remains a subject of intensive mathematical research [9,10]. The transport problem was inspected from the perspective of free probability [11] and applied in the study of causality [12–14].

An attempt to generalize the notion of the Monge distance for quantum theory was pursued for the setup of infinite- [15] and finite-dimensional [1,16] Hilbert spaces. Such a distance between any two quantum states, defined by the Monge distance between the corresponding Husimi distributions, enjoys the semiclassical property: The distance between two coherent states, centered at points x and y in the classical phase space, is equal to the distance $|x-y|$ between the points at which both coherent states are concentrated. This property, crucial for studies on quantum analogs of the Lyapunov exponent [17], is also shared by the distance recently proposed in Ref. [18].

Any definition based on the notion of the Husimi function depends on the choice of the set of coherent states. It is, therefore, natural to look for a universal method to introduce the transport distance between quantum states directly by applying the Kantorovich-Wasserstein approach and performing optimization over the set of bipartite quantum states with fixed marginals [19–21]. In spite of recent vibrant activity in this field [22–28], this aim has not been fully achieved until now [29–32].

In parallel, quantum optimal transport has found numerous applications in quantum physics, in particular, in connection with the measures of proximity of quantum states [33–36]. The latter play a key role in quantum metrology [37–39] as well as in quantum machine learning [40–43].

In this Letter, we introduce a measure of proximity of quantum states, dubbed the "SWAP fidelity," as it is inspired by quantum optimal transport with a specific quantum cost matrix. It shares many properties with the standard

Uhlmann-Jozsa fidelity [44,45] and agrees with the latter if at least one of the states is pure. We prove that the square root of the associated quantum optimal transport cost yields a new distance on the space of qubits which is a quantum analog of the 2-Wasserstein distance. For larger dimensions, we show analytically that this quantity gives a semidistance and provide numerical evidence that it also satisfies the triangle inequality. Moreover, we prove that it is bounded from above by the root infidelity and from below by the rescaled Bures distance. We further discuss the general form of a quantum cost matrix and study the quantum-to-classical transition of the transport problem. The latter shows that quantum optimal transport is cheaper than the classical one, generalizing the results of Ref. [36]. Finally, we discuss an application of the new quantum metric in the context of quantum generative adversarial networks (QGANs).

Classical transport problem.—To formulate the mass transport problem in the setup of Kantorovich for any probability distributions p_i^A and p_j^B, one introduces the notion of a classical coupling—a joint probability distribution P_{ij}^{AB} with two specified marginals $p_i^A = \sum_j P_{ij}^{AB}$ and $p_j^B = \sum_i P_{ij}^{AB}$. In the case of two probability vectors of the order of N, p^A, $p^B \in \Delta_N$, any joint distribution $P^{AB} \in \Delta_{N^2}$, which determines a transport plan, is represented by a single vector P_μ, with $\mu = (i-1)N + j = 1, \ldots, N^2$. The set $\Gamma^{\mathrm{cl}}(p^A, p^B)$ of all admissible couplings forms a convex subset of the simplex Δ_{N^2}, with extreme points characterized in Ref. [46].

Consider a set of N points $X := \{x_i\}_{i=1}^N$ equipped with a distance function d. With the latter, we associate a symmetric $N \times N$ matrix $E_{ij} := d(x_i, x_j)$. Assuming that the transport cost of a unit of mass from point x_i to point x_j is equal to the distance E_{ij}, one can formulate the classical transport problem [5,26]. In order to study its quantum analog, it will be convenient to reshape the square distance matrix E_{ij} of dimension N into a distance vector D_μ of length N^2.

To generate a Wasserstein distance between probability distributions p^A and p^B, one can use a $N^2 \times N^2$ classical cost matrix C which is diagonal, $C_{\mu\nu} = D_\mu \delta_{\mu\nu}$ for $\mu, \nu = 1, \ldots, N^2$, and a diagonal density matrix $\rho_{\mu\nu}^{AB} = P_\mu \delta_{\mu\nu}$. For a given transport plan P^{AB}, the total transport cost is given by the scalar product $\hat{T}_C(P) := \sum_{\mu=1}^{N^2} D_\mu P_\mu = \mathrm{Tr} C\rho^{AB}$. The minimal transport cost $T_C^{\mathrm{cl}} := \min_P \hat{T}_C(P)$ leads to the family of Wasserstein distances $W_{C,p}^{\mathrm{cl}}$ defined for $p \geq 1$:

$$W_{C,p}^{\mathrm{cl}}(p^A, p^B) := [\min_{P^{AB} \in \Gamma^{\mathrm{cl}}} (\mathrm{Tr} C^p \rho^{AB})]^{1/p} = (T_{C^p}^{\mathrm{cl}})^{1/p}. \quad (1)$$

The minimum is taken over the set $\Gamma^{\mathrm{cl}}(p^A, p^B)$ of classical couplings [5]. If $d(x_i, x_j) = 1 - \delta_{ij}$, the space X has the geometry of an N-point simplex Δ_N. In this case, $C^p = C$ and $W_{C,p}^{\mathrm{cl}} = (T_C^{\mathrm{cl}})^{1/p}$ for any $p \geq 1$, so we shall abbreviate $T^{\mathrm{cl}} := T_C^{\mathrm{cl}}$ and denote the classical cost matrix by C^{cl}.

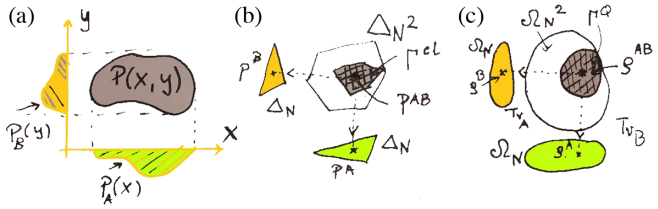

FIG. 1. Couplings between probability distributions used for Kantorovich distance: (a) continuous 1D probabilities $p_A(x)$ and $p_B(y)$ coupled by a joint distribution $P(x, y)$; (b) two N-point classical states p^A, $p^B \in \Delta_N$ coupled by a joint state $P^{AB} \in \Gamma^{\mathrm{cl}} \subset \Delta_{N^2}$ with adjusted marginals; (c) two quantum states $\rho^A, \rho^B \in \Omega_N$ coupled by a bipartite state $\rho^{AB} \in \Gamma^Q \subset \Omega_{N^2}$ such that $\mathrm{Tr}_A \rho^{AB} = \rho^B$ and $\mathrm{Tr}_B \rho^{AB} = \rho^A$.

Proximity of quantum states.—We now switch to the quantum setting and denote the set of $N \times N$ density matrices by $\Omega_N = \{\rho : \rho = \rho^\dagger, \rho \geq 0, \mathrm{Tr}\rho = 1\}$. To quantify the closeness of any two quantum states, one uses various distances on Ω_N—see Ref. [1]. The *trace distance*, singled out by the Helstrom theorem on optimal distinguishability [47], reads $D_{\mathrm{Tr}}(\rho^A, \rho^B) := \frac{1}{2}\mathrm{Tr}|\rho^A - \rho^B|$, where $|X| := \sqrt{XX^\dagger}$. Another way to characterize the proximity between two density matrices relies on Uhlmann-Jozsa *fidelity* [44,45] $F(\rho^A, \rho^B) := (\mathrm{Tr}|\sqrt{\rho^A}\sqrt{\rho^B}|)^2$. It leads to the following distances: the *root infidelity* [48] $I := \sqrt{1-F}$, the *Bures distance* [44,49] $B := \sqrt{2(1-\sqrt{F})}$, and the *Bures angle* $A := (2/\pi)\arccos\sqrt{F}$. Note that the Bures distance and other distances based on fidelity are closely related to statistical distinguishability and quantum Fisher information [37,38], so they have a direct interpretation in quantum metrology [39].

We shall now introduce a quantity analogous to fidelity, which is directly related to the quantum optimal transport problem and its applications in machine learning [42].

SWAP fidelity.—Consider two arbitrary states ρ^A, $\rho^B \in \Omega_N$. A composed (bipartite) density matrix ρ^{AB} of the order of N^2 is called a *coupling matrix* [50] between ρ^A and ρ^B if both partial traces agree, $\mathrm{Tr}_A \rho^{AB} = \rho^B$ and $\mathrm{Tr}_B \rho^{AB} = \rho^A$. The set of all possible quantum couplings matrices will be denoted by $\Gamma^Q(\rho^A, \rho^B) \subset \Omega_{N^2}$—see Fig. 1(c). The bipartite quantum states can be conveniently represented in the *Fano form*—see Supplemental Material [51], which includes the additional Refs. [52–70].

Let S denote the SWAP operator, $S(|x\rangle \otimes |y\rangle) := |y\rangle \otimes |x\rangle$ for any vectors $|x\rangle$ and $|y\rangle$. For any $\rho^A, \rho^B \in \Omega_N$ we introduce the SWAP *fidelity*:

$$F_S(\rho^A, \rho^B) := \max_{\rho^{AB} \in \Gamma^Q} (\mathrm{Tr} S\rho^{AB}), \quad (2)$$

where the maximum is taken over the set Γ^Q of all admissible coupling matrices ρ^{AB}.

Proposition 1.—For any dimension N, the SWAP fidelity F_S is a symmetric jointly concave function from $\Omega_N \times \Omega_N$ to the unit interval. Furthermore, $F_S(\rho^A, \rho^B) = 1$ iff $\rho^A = \rho^B$, $F_S(\rho^A, \rho^B) = 0$ iff $\mathrm{Tr}\rho^A\rho^B = 0$, and

$$F_S(\rho^A, \rho^B) = F_S(U\rho^A U^\dagger, U\rho^B U^\dagger), \quad \text{for } U \in U(N), \tag{3}$$

$$F(\rho^A, \rho^B) \le F_S(\rho^A, \rho^B) \le \sqrt{F(\rho^A, \rho^B)}, \tag{4}$$

$$F_S(\rho^A \otimes \sigma^A, \rho^B \otimes \sigma^B) \ge F_S(\rho^A, \rho^B) F_S(\sigma^A, \sigma^B). \tag{5}$$

The above result, proven in Supplemental Material [51], shows that, in analogy to the fidelity F, the SWAP fidelity F_S equals unity iff both states coincide and vanishes iff they are orthogonal. Furthermore, it interpolates between fidelity and root fidelity—see inequality (4)—with the first inequality saturated if at least one of the states is pure. Notably, the SWAP fidelity is *supermultiplicative* with respect to the tensor product, as it satisfies inequality (5) characteristic to superfidelity [71]. Note also that F_S is jointly concave, as is the root fidelity $\sqrt{F}$, while it has a probabilistic interpretation for pure states, $F_S(\phi, \psi) = F(\phi, \psi) = |\langle\phi, \psi\rangle|^2$—see Ref. [1]. The SWAP fidelity is shown below to be closely related to the quantum optimal transport and yields a novel metric on the Bloch ball.

Quantum cost matrix.—To study the transport problem between two quantum states of the order of N, we need to specify a quantum cost matrix C^Q of size N^2. Let $\{|i\rangle\}_{i=1}^N$ be the computational basis of an N-level quantum system and denote the maximally entangled singlet states in the subspace spanned by $|i\rangle$ and $|j\rangle$ by $|\psi_{ij}^-\rangle = (1/\sqrt{2})(|i, j\rangle - |j, i\rangle)$. In the case of the simplex geometry $E_{ij} = 1 - \delta_{ij}$, the quantum optimal transport enjoys several desirable features if one chooses the cost matrix C^Q to be the projector onto the antisymmetric subspace:

$$C^Q = \sum_{j>i=1}^{N} |\psi_{ij}^-\rangle\langle\psi_{ij}^-| = \frac{1}{2}(\mathbb{1}_{N^2} - S) = (C^Q)^2, \tag{6}$$

as advocated also in Refs. [28–30,42]. In particular, for the simplest, one-qubit problem, $N = 2$, the cost matrix reads

$$C^Q = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = |\psi^-\rangle\langle\psi^-|. \tag{7}$$

The above quantum cost matrix C^Q of size N^2 forms a *coherification* [72] of a classical cost matrix $C^{\mathrm{cl}} = \mathrm{diag}(C^Q)$ corresponding to the simplex geometry.

We are now going to look for the minimal quantum transport cost, which can be expressed using the SWAP fidelity:

$$T^Q(\rho^A, \rho^B) := \min_{\rho^{AB} \in \Gamma^Q} (\mathrm{Tr} C^Q \rho^{AB}) = \frac{1 - F_S(\rho^A, \rho^B)}{2}. \tag{8}$$

Proposition 1 directly implies that for any two states $\rho^A, \rho^B \in \Omega_N$ the optimal quantum transport cost T^Q is jointly convex, symmetric, non-negative and vanishes iff $\rho^A = \rho^B$. Furthermore, for C^Q given by Eq. (6), one has

$$T^Q(\rho^A, \rho^B) = T^Q(U\rho^A U^\dagger, U\rho^B U^\dagger), \tag{9}$$

for any unitary operator U on $\mathbb{C}^N$. Hence, T^Q forms a semidistance on Ω_N, as shown independently in Ref. [42]. In analogy with the classical definition (1), for any $p \ge 0$ we introduce a quantum analog of the p-Wasserstein distance, $W_p := (T^Q)^{1/p}$. As shown below, W_2 plays a distinguished role, so we will denote it simply by $W := W_2 = \sqrt{T^Q}$.

As an immediate corollary of inequality (4), proven in Supplemental Material [51] with the help of recent results by Yu *et al.* [29], we arrive at explicit bounds for the quantum transport cost and its square root W.

Corollary 2.—For any two states $\rho^A, \rho^B \in \Omega_N$ we have

$$\frac{1}{\sqrt{2}} B(\rho^A, \rho^B) \ge \frac{1}{\sqrt{2}} I(\rho^A, \rho^B) \ge W(\rho^A, \rho^B) \ge \frac{1}{2} B(\rho^A, \rho^B). \tag{10}$$

Furthermore, the second inequality is saturated if either ρ^A or ρ^B is pure.

Note that this corollary implies that W is a distance on the space of pure quantum states of any dimension N.

Single-qubit transport.—In the simplest case of $N = 2$, the quantum cost matrix is given by Eq. (7). As shown in Supplemental Material [51], the full solution of the quantum transport problem for the one-qubit case is equivalent to solving a polynomial equation of the order of six. The latter yields analytic expressions in several special cases.

For two diagonal matrices $\rho_r^{\mathrm{cl}} = \mathrm{diag}(r, 1 - r)$ and $\rho_s^{\mathrm{cl}} = \mathrm{diag}(s, 1 - s)$, we have

$$W(\rho_r^{\mathrm{cl}}, \rho_s^{\mathrm{cl}}) = \frac{1}{\sqrt{2}} \max\{|\sqrt{r} - \sqrt{s}|, |\sqrt{1 - r} - \sqrt{1 - s}|\}. \tag{11}$$

Furthermore, if one of the states is totally mixed, we obtain

$$W\left(\frac{1}{2}\mathbb{1}, \rho\right) = \frac{1}{2} \max\{|1 - \sqrt{2\lambda}|, |1 - \sqrt{2(1 - \lambda)}|\}, \tag{12}$$

where λ and $1-\lambda$ denote the eigenvalues of ρ. A surprisingly simple formula is available for two isospectral states:

$$W(\rho, U\rho U^\dagger) = \sqrt{\frac{1}{2} - \sqrt{\lambda(1-\lambda)}}\,|\sin(\theta/2)|, \quad (13)$$

where $\{\lambda, 1-\lambda\}$ is the common spectrum and θ is the U-dependent angle between Bloch vectors associated with the states—see Supplemental Material [51]. In the case of a single qubit, we obtain one of the key results of this Letter, proved in Supplemental Material [51].

Theorem 3.—For $N = 2$, the function W_p satisfies the triangle inequality iff $p \geq 2$: For any states $\rho^A, \rho^B, \rho^C \in \Omega_2$, one has $W_p(\rho^A, \rho^B) + W_p(\rho^B, \rho^C) \geq W_p(\rho^A, \rho^C)$. Thus, W_p generates a distance on the Bloch ball Ω_2, analogous to the classical p-Wasserstein distance (1), provided that $p \geq 2$.

Whereas the quantum transport cost itself, $T^Q = W_1$, is *not* a distance on Ω_2, its square root W_2 is—see Supplemental Material [51]. Note that, similarly, while the infidelity $1-F$ does not satisfy the triangle inequality, its square root $I = \sqrt{1-F}$ does [48]. An analogous property was recently proved for the quantum Jensen-Shannon divergence, the square root of which satisfies the triangle inequality and leads to the transmission metric [73,74].

Recall also that the Monge distance between quantum states defined by the Husimi distribution with respect to spin coherent states for $N = 2$ leads to the Hilbert-Schmidt distance and the Euclidean geometry on the Bloch ball, while the discrete Monge distance, describing movements

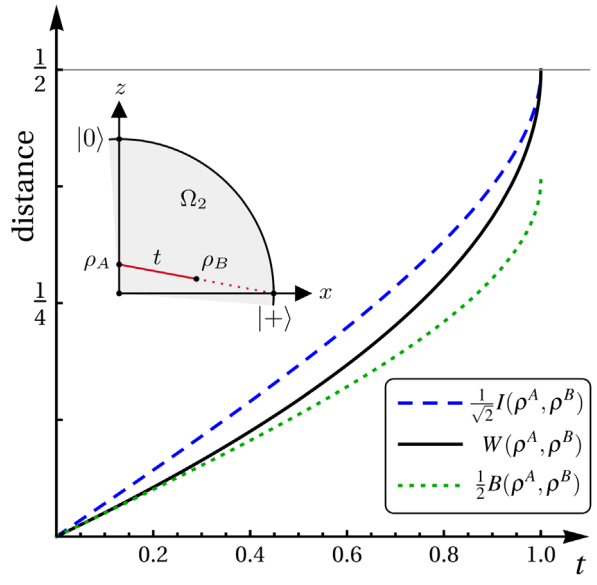

FIG. 2. Bounds (10) illustrated in the Bloch ball Ω_2. The distances between the state $\rho^A = (9/20)\mathbb{1} + (1/10)|0\rangle\langle 0|$ and $\rho^B = (1-t)\rho^A + t(|+\rangle\langle +|)$ are shown as a function of $t \in [0,1]$ varying along the Euclidean line (red line in the inset).

of the Majorana stars corresponding to pure states, gives geodesic distance on the Bloch sphere [16]. Whereas formula (10) implies that W is strongly equivalent to the Bures metric B and induces the same topology on the Bloch ball, the corresponding (curved) geometries are different. This is illustrated in Fig. 2 for a fixed mixed state ρ^A and ρ^B varying continuously from ρ^A to the pure state $|+\rangle$. We witness the validity of the bound (10), with $W(\rho^A, |+\rangle) = I(\rho^A, |+\rangle)$. Observe also that initially the transport distance curve closely follows that of the Bures distance. Using the Pauli matrices σ_i and the Bloch representation $\rho_\pm(\vec{\tau}) \coloneqq \frac{1}{2}(\mathbb{1} \pm \vec{\tau}\cdot\vec{\sigma})$ for $\|\vec{\tau}\| \in [0,1]$, we have

$$W[\rho_+(\vec{\tau}), \rho_-(\vec{\tau})] = \frac{1}{\sqrt{2}} B[\rho_+(\vec{\tau}), \rho_-(\vec{\tau})]. \quad (14)$$

Quantization of an arbitrary classical cost matrix.—In a more general setup, consider an arbitrary distance function on the N-point set X, determined by the matrix E_{ij}. With any such classical geometry of X, we associate the following quantum cost matrix:

$$C \coloneqq C_E^Q = \sum_{j>i=1}^{N} E_{ij} |\psi_{ij}^-\rangle\langle\psi_{ij}^-|. \quad (15)$$

Accordingly, for any $p \geq 1$, we define

$$T_C^Q(\rho^A, \rho^B) \coloneqq \min_{\rho^{AB} \in \Gamma^Q} (\mathrm{Tr} C \rho^{AB}), \qquad W_{C,p} \coloneqq (T_{C^p}^Q)^{1/p} \quad (16)$$

and prove the following result in Supplemental Material [51].

Proposition 4.—For any N, any $p \geq 1$, and any choice of classical geometry E, the pth root of the corresponding optimal quantum transport cost $W_{C,p}$ is a semidistance on Ω_N.

The quantum-to-classical transition.—It is instructive to compare the quantum transport problem with its classical counterpart. To this end, one embeds classical probability vectors in diagonal density matrices. The following result shows that the quantum transport cost between two classical states is always cheaper than the corresponding classical cost (see also Ref. [36]).

Proposition 5.—Let $\vec{r}$, $\vec{s}$ be two N-dimensional probability vectors and let $\rho_{\vec{r}}^{\mathrm{cl}}, \rho_{\vec{s}}^{\mathrm{cl}} \in \Omega_N$ be the corresponding quantum states defined as $(\rho_{\vec{r}}^{\mathrm{cl}})_{ij} \coloneqq r_i \delta_{ij}$. Then, for any quantum cost matrix C of the form (15), we have

$$T_C^Q(\rho_{\vec{r}}^{\mathrm{cl}}, \rho_{\vec{s}}^{\mathrm{cl}}) \leq T_{\mathrm{diag}(C)}^{\mathrm{cl}}(\vec{r}, \vec{s}). \quad (17)$$

This follows from the fact that the quantum optimization is performed over a strictly larger set of admissible couplings, $\Gamma^{\mathrm{cl}}(\vec{r}, \vec{s}) \subset \Gamma^Q(\rho_{\vec{r}}^{\mathrm{cl}}, \rho_{\vec{s}}^{\mathrm{cl}})$.

The quantum-to-classical transition of the transport problem can be interpreted in terms of decoherence caused by the interaction of the information processing device with

its environment. As a simple model (cf. [75]), one can assume that the quantum cost matrix is acted upon by a dephasing channel $\mathcal{E}_\alpha(C) = \alpha C + (1-\alpha)\mathrm{diag}(C)$, with the parameter $\alpha \in [0,1]$ proportional to the l_1 coherence [76]. One can then study the function $T^Q_{C,\alpha} := \min_{\rho^{AB}\in\Gamma^Q}[\mathrm{Tr}\mathcal{E}_\alpha(C)\rho^{AB}]$. In the single-qubit case, it is easy to show (see Supplemental Material [51]) that $\sqrt{T^Q_{C^Q,\alpha}}$ is a distance on the set of commuting density matrices of the order of 2. Moreover, it is a strictly decreasing function of α, provided that the two input states are different and none of them is pure.

Applications.—The introduced SWAP fidelity offers an original measure of proximity between quantum states and, thus, provides a new tool to quantify protocols of quantum information processing. Its most promising and straightforward application pertains to the QGANs [41,77]. This protocol of quantum machine learning [40] consists of a generator, which produces "fake" data, and a discriminator, which aims at distinguishing between the real and fake input data. The adversarial training reaches a fixed point when the generator produces data with true statistics and the discriminator's efficiency is 50%. Similar to with classical GANs, the choice of the distance between real and fake data is critical for the stability and performance of the training [39,42,78]. In Ref. [39], it was argued that problems with efficiency of quantum learning algorithms [79–82] arise because the employed measure of proximity diminishes exponentially with the number of qubits. Although the introduced SWAP fidelity suffers from the same drawback for pure states (as it equals to fidelity in this case), it might prove superior for mixed states because of its supermultiplicativity (5).

In fact, in Ref. [42], a QGAN based on the semidistance T^Q was shown to exhibit improved performance over other QGANs. Furthermore, it is noise tolerant and can be successfully used to approximate complicated quantum circuits with a limited number of quantum gates. Our results suggest that the choice $W = \sqrt{T^Q}$ is superior to T^Q, as it forms a genuine distance.

Conclusions and outlook.—We studied the quantum transport problem for density matrices of dimension N with a quantum cost matrix [83]. Inspired by the Wasserstein distance of the order of 2, we proved that the square root of the optimal transport cost, $W = \sqrt{T^Q}$, yields a distance on the Bloch ball, bounded by the rescaled Bures distance and the root infidelity. In the general problem of N-level systems and an arbitrary classical geometry, we showed that the analog of the p-Wasserstein distance, $W_{C,p}$, yields a semidistance on the full space of quantum states for any $p \geq 1$. Furthermore, extensive numerical studies (see Supplemental Material [51]) allow us to put forward the following conjectures.

Conjecture I.—The quantum analog of the 2-Wasserstein distance, $W = \sqrt{T^Q}$, enjoys the triangle inequality in any dimension N.

Conjecture II.—The quantity $W_{C,p}$ is a distance for any $p \geq 2$ and any C in some neighborhood of C^Q.

Moreover, recent analytical and numerical results established in Ref. [84] motivate the following.

Conjecture III.—The quantum analog of the 2-Wasserstein distance, W, is monotone under quantum channels.

Given the multifarious applications of the classical Wasserstein distances, we expect its quantum analog to play a pivotal role in diverse branches of quantum information processing and quantum machine learning [32–36,43]. Furthermore, the SWAP fidelity—a novel quantity introduced in this work—is likely to offer new advances in characterizing proximity between quantum states.

It is a pleasure to thank R. Bistroń, P. W. Lamberti, T. Miller, and W. Słomczyński for inspiring discussions and fruitful interaction. Financial support by Simons collaboration grant for mathematicians, Narodowe Centrum Nauki under the Maestro Grant No. DEC-2015/18/A/ST2/00274, and by Foundation for Polish Science under the Team-Net Project No. POIR.04.04.00-00-17C1/18-00 is gratefully acknowledged.

*Corresponding author.
michal.eckstein@uj.edu.pl

[1] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States*, 2nd ed. (Cambridge University Press, Cambridge, England, 2017).
[2] G. Aubrun and S. J. Szarek, *Alice and Bob Meet Banach* (AMS, Providence RI, 2017).
[3] G. Monge, *Mémoire sur la théorie des déblais et des remblais* (Histoire de l'Académie Royale des Sciences de Paris, Paris, 1781).
[4] T. Salvemini, Sul calcolo degli indici di concordanza tra due caratteri quantitativi, Atti della VI Riunione Soc. Ital. di Statistica, Roma (1943).
[5] S. Rachev and L. Rüschendorf, *Mass Transportation Problems*, Vol. I: Theory and Vol. II: Applications (Springer, New York, 1998).
[6] L. V. Kantorovich, On the translocation of masses, J. Math. Sci. **133**, 1381 (2006).
[7] L. V. Kantorovich, On a problem of Monge, J. Math. Sci. **133**, 1383 (2006).
[8] L. N. Vaserstein, Markov processes over denumerable products of spaces describing large systems of automata, Probl. Inf. Transm. **5**, 47 (1969).
[9] C. Villani, *Optimal Transport. Old and New* (Springer-Verlag, Berlin, 2009).
[10] A. M. Vershik, Long history of the Monge-Kantorovich transportation problem, Math. Intell. **35**, 1 (2013).
[11] P. Biane and D. Voiculescu, A free probability analogue of the Wasserstein metric on the trace-state space, Geom. Funct. Anal. **11**, 1125 (2001).
[12] M. Eckstein and T. Miller, Causality for nonlocal phenomena, Ann. Henri Poincaré **18**, 3049 (2017).

[13] M. Eckstein and T. Miller, Causal evolution of wave packets, Phys. Rev. A **95**, 032106 (2017).

[14] M. Eckstein, P. Horodecki, T. Miller, and R. Horodecki, Operational causality in spacetime, Phys. Rev. A **101**, 042128 (2020).

[15] K. Życzkowski and W. Słomczyński, Monge distance between quantum states, J. Phys. A **31**, 9095 (1998).

[16] K. Życzkowski and W. Słomczyński, The Monge metric on the sphere and geometry of quantum states, J. Phys. A **34**, 6689 (2001).

[17] K. Życzkowski, H. Wiedemann, and W. Słomczyński, How to generalize Lapunov exponent for quantum mechanics, Vistas Astron. **37**, 153 (1993).

[18] Z. Wang, Y. Wang, and B. Wu, Genuine quantum chaos and physical distance between quantum states, Phys. Rev. E **103**, 042209 (2021).

[19] J. Agredo and F. Fagnola, On quantum versions of the classical Wasserstein distance, Stochastics **89**, 910 (2017).

[20] F. Golse and T. Paul, Wave packets and the quadratic Monge-Kantorovich distance in quantum mechanics, C. R. Math. **356**, 177 (2018).

[21] Y. Chen, T. T. Georgiou, and A. Tannenbaum, Wasserstein geometry of quantum states and optimal transport of matrix-valued measures, in *Emerging Applications of Control and Systems Theory* (Springer, New York, 2018), p. 139.

[22] F. Golse, C. Mouhot, and T. Paul, On the mean field and classical limits of quantum mechanics, Commun. Math. Phys. **343**, 165 (2016).

[23] Y. Chen, T. T Georgiou, L. Ning, and A. Tannenbaum, Matricial Wasserstein-1 distance, IEEE Control Syst. Lett. **1**, 14 (2017).

[24] R. Bhatia, S. Gaubert, and T. Jain, Matrix versions of the Hellinger distance, Lett. Math. Phys. **109**, 1777 (2019).

[25] S. Friedland, J. Ge, and L. Zhi, Quantum Strassen's theorem, Inf. Dim. Anal. Quant. Prob. Rel. Topics **23**, 2050020 (2020).

[26] S. Friedland, Tensor optimal transport, distance between sets of measures and tensor scaling, arXiv:2005.00945.

[27] Y. Chen, W. Gangbo, T. T. Georgiou, and A. Tannenbaum, On the matrix Monge-Kantorovich problem, Eur. J. Appl. Math. **31**, 574 (2020).

[28] R. Duvenhage, Quadratic Wasserstein metrics for von Neumann algebras via transport plans, arXiv:2012.03564.

[29] N. Yu, L. Zhou, S. Ying, and M. Ying, Quantum earth mover's distance, no-go quantum Kantorovich-Rubinstein theorem, and quantum marginal problem, arXiv: 1803.02673v2.

[30] M. H. Reira, A transport approach to distances in quantum systems, Bachelor's Thesis, Universitat Autònoma de Barcelona, 2018.

[31] K. Ikeda, Foundation of quantum optimal transport and applications, Quantum Inf. Process. **19**, 25 (2020).

[32] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd, The quantum Wasserstein distance of order 1, IEEE Trans. Inf. Theory **67**, 6627 (2021).

[33] G. De Palma and D. Trevisan, Quantum optimal transport with quantum channels, Ann. Henri Poincaré **22**, 3199 (2021).

[34] E. A. Carlen and J. Maas, Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems, J. Stat. Phys. **178**, 319 (2020).

[35] N. Datta and C. Rouzé, Relating relative entropy, optimal transport and Fisher information: A quantum HWI inequality, Ann. Henri Poincaré **21**, 2115 (2020).

[36] E. Caglioti, F. Golse, and T. Paul, Quantum optimal transport is cheaper, J. Stat. Phys. **181**, 149 (2020).

[37] S. L. Braunstein and C. M. Caves, Statistical Distance and the Geometry of Quantum States, Phys. Rev. Lett. **72**, 3439 (1994).

[38] D. Šafránek, Discontinuities of the quantum Fisher information and the Bures metric, Phys. Rev. A **95**, 052320 (2017).

[39] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, Quantum Fisher information matrix and multiparameter estimation, J. Phys. A **53**, 023001 (2020).

[40] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[41] S. Lloyd and C. Weedbrook, Quantum Generative Adversarial Learning, Phys. Rev. Lett. **121**, 040502 (2018).

[42] S. Chakrabarti, Y. Huang, T. Li, S. Feizi, and X. Wu, Quantum Wasserstein generative adversarial networks, in *Advances in Neural Information Processing Systems* (2019).

[43] B. T. Kiani, G. De Palma, M. Marvian, Z.-W. Liu, and S. Lloyd, Quantum Earth Mover's distance: A new approach to learning quantum data, Quantum Sci. Technol. **7**, 045002 (2022).

[44] A. Uhlmann, The 'transition probability' in the state space of a *-algebra, Rep. Math. Phys. **9**, 273 (1976).

[45] R. Jozsa, Fidelity for mixed quantum states, J. Mod. Opt. **41**, 2315 (1994).

[46] K. R. Parthasarathy, Extreme points of the convex set of joint probability distributions with fixed marginals, Proc. Indian Acad. Sci. (Math. Sci.) **117**, 505 (2007).

[47] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[48] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Distance measures to compare real and ideal quantum processes, Phys. Rev. A **71**, 062310 (2005).

[49] A. Uhlmann, Geometric phases and related structures, Rep. Math. Phys. **36**, 461 (1995).

[50] A. Winter, Tight uniform continuity bounds for quantum entropies: Conditional entropy, relative entropy distance and energy constraints, Commun. Math. Phys. **347**, 291 (2016).

[51] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.129.110402, which contains the proofs of technical results stated in the paper, the discussion of numerical simulations and some auxiliary theoretical concepts.

[52] R. F. Werner, Quantum states with Einstein–Podolsky–Rosen correlations admitting a hidden-variable mode, Phys. Rev. A **40**, 4277 (1989).

[53] C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum mechanical states, IEEE Trans. Inf. Theory **45**, 1216 (1999).

[54] J. Preskill, *Lecture Notes for Quantum Information* (Caltech, Pasadena, 2018), Chap. 3.
[55] J. Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. **261**, 2250 (2011).
[56] U. Fano, Pairs of two-level systems, Rev. Mod. Phys. **55**, 855 (1983).
[57] P. Badziąg, M. Horodecki, P. Horodecki, and R. Horodecki, Local environment can enhance fidelity of quantum teleportation, Phys. Rev. A **62**, 012311 (2000).
[58] J. I. de Vincente, Separability criteria based on the Bloch representation of density matrices, Quantum Inf. Comput. **7**, 624 (2007).
[59] O. Rudolph, On extremal quantum states of composite systems with fixed marginals, J. Math. Phys. (N.Y.) **45**, 4035 (2004).
[60] K. R. Parthasarathy, Extremal quantum states in coupled systems, Ann. Inst. Henri Poincaré, Probab. Stat. **41**, 257 (2005).
[61] C.-K. Li, Y.-T. Poon, and X. Wang, Ranks and eigenvalues of states with prescribed reduced states, Electron. J. Linear Algebra **27**, 935 (2014).
[62] S. Bravyi, Requirments for compatibility between local and multipartite quantum states, Quantum Inf. Comput. **4**, 12 (2004).
[63] A. Klyachko, Quantum marginal problem and representations of the symmetric group, arXiv:quant-ph/0409113.
[64] C. King and M. B. Ruskai, Minimal entropy of states emerging from noisy quantum channels, IEEE Trans. Inf. Theory **47**, 192 (2001).
[65] S. Kimmel, M. P. da Silva, C. A. Ryan, B. R. Johnson, and T. Ohki, Robust Extraction of Tomographic Information via Randomized Benchmarking, Phys. Rev. X **4**, 011050 (2014).
[66] H. Ohno, Maximal rank of extremal marginal tracial states, J. Math. Phys. (N.Y.) **51**, 092101 (2010).
[67] R. Duvenhage, Optimal quantum channels, Phys. Rev. A **104**, 032604 (2021).
[68] K. Życzkowski, K. Penson, I. Nechita, and B. Collins, Generating random density matrices, J. Math. Phys. (N.Y.) **52**, 062201 (2011).
[69] M. Grant and S. Boyd, Graph implementations for nonsmooth convex programs, in *Recent Advances in Learning and Control* (Springer, New York, 2008), pp. 95–110.
[70] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.2, http://cvxr.com/cvx (2014).
[71] J. A. Miszczak, Z. Puchala, P. Horodecki, A. Uhlmann, and K. Życzkowski, Sub-and super-fidelity as bounds for quantum fidelity, Quantum Inf. Comput. **9**, 0103 (2009).
[72] K. Korzekwa, S. Czachórski, Z. Puchała, and K. Życzkowski, Coherifying quantum channels, New J. Phys. **20**, 043028 (2018).
[73] J. Briët and P. Harremoës, Properties of classical and quantum Jensen-Shannon divergence, Phys. Rev. A **79**, 052311 (2009).
[74] D. Virosztek, The metric property of the quantum Jensen–Shannon divergence, Adv. Math. **380**, 107595 (2021).
[75] W. H. Zurek, Decoherence, einselection, and the quantum origins of the classical, Rev. Mod. Phys. **75**, 715 (2003).
[76] T. Baumgratz, M. Cramer, and M. B. Plenio, Quantifying Coherence, Phys. Rev. Lett. **113**, 140401 (2014).
[77] P.-L. Dallaire-Demers and N. Killoran, Quantum generative adversarial networks, Phys. Rev. A **98**, 012324 (2018).
[78] M. Arjovsky, S. Chintala, and L. Bottou, Wasserstein generative adversarial networks, Proc. Mach. Learn. Res. **70**, 214 (2017).
[79] A. N. Pechen and D. J. Tannor, Are there Traps in Quantum Control Landscapes?, Phys. Rev. Lett. **106**, 120402 (2011).
[80] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).
[81] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).
[82] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost-function-dependent barren plateaus in shallow quantum neural networks, Nat. Commun. **12**, 1791 (2021).
[83] A more detailed mathematical study is presented in the companion paper: S. Cole, M. Eckstein, S. Friedland, and K. Życzkowski, Quantum optimal transport, arXiv:2105.06922.
[84] R. Bistroń, M. Eckstein, and K. Życzkowski, Monotonicity of the quantum 2-Wasserstein distance, arXiv:2204.07405.