Check for updates

Quantum optimal transport: an invitation

Dario Trevisan[1] 🔘

Abstract
The optimal mass transport problem was formulated centuries ago, but only recently there has been a surge in its applications, particularly in functional inequalities, geometry, stochastic analysis, and numerical solutions for partial differential equations. Quantum optimal transport aims to extend this success story to non-commutative systems, where density operators replace probability measures. This brief review paper aims to describe the latest approaches, highlighting their advantages, disadvantages, and open mathematical problems relevant to applications.

Keywords Optimal transport · Wasserstein distance · Quantum systems

1 Introduction

Since the pioneering works of Monge in the 18th century and Kantorovich in the 20th century, the problem of optimal mass transport has stimulated the development of tools in various areas of mathematics, such as the analysis of partial differential equations, convex analysis, operations research, as well as probability and statistics. In recent decades, there has been an explosion of applications of optimal transport, with significant implications in the study of functional inequalities, geometry and stochastic analysis, as well as numerical solution schemes for partial differential equations [1]. More recently, the increase in computing power combined with iterative algorithms has allowed solving instances of large-scale and high-dimensional optimal transport problems, with applications in machine learning and, more generally, providing useful tools for analyzing large amounts of data [2].

Quantum optimal transport is a generalization of classical optimal transport where probability measures are replaced by suitable density operators on Hilbert spaces, which represent the states of a quantum system. The simplest example concerns the process of transforming a qubit (a quantum system in a two-dimensional space) from one state to another optimally, minimizing a suitably defined transport cost. Currently, there are several approaches to formulate the mathematical problem of quantum optimal transport, with applications in quantum computing, quantum communication, and many body quantum systems.

The purpose of this review paper, following the author's exposition at the XXII congress of the Italian Mathematical Union, is to describe the most recent approaches, examining their advantages and disadvantages, with a particular emphasis on the main currently open

✉ Dario Trevisan
  dario.trevisan@unipi.it

[1] Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 10, Pisa 56127, PI, Italy

mathematical problems, and their relevance in applications. We intentionally try to keep the exposition short, at the price of possibly losing some generality, but still introducing the key concepts. In order to appeal to a broader audience, we provide a short introduction to classical optimal transport theory and quantum mechanics (for finite dimensional systems). We also refer to the upcoming monograph [3] for a collection of notes based on series of lectures by leading experts on the subject. We aimed to give a fairly complete review of all the main approaches to quantum optimal transport, although a natural focus is on those we are currently working on.

The paper is structured as follows. Section 2 describes the basic facts and notation for the classical optimal transport problem. Section 3 briefly recalls quantum systems and their operations. Section 4 discusses the main approaches to quantum optimal transport and their applications.

2 Classical optimal transport

In this section, we provide a brief overview of classical optimal transport theory, focusing on finite sets and discrete measures. We refer interested readers to comprehensive monographs such as [1, 2, 4–6] for a more in-depth treatment.

Monge

The roots of optimal transport theory can be traced back to Monge's memoir, published in 1781 [7]. He laid the basis for a mathematical framework to study the optimal transportation of goods or mass between locations. His key idea was to seek a transportation map that minimizes the total cost (in his case, the distance) required to move mass from one location to another. Although he considered only absolutely continuous distributions of mass (in modern terms, densities with respect to the Lebesgue measure), to keep technicalities at a minimum, we focus instead on the following discrete formulation of the optimal transport problem. Given

1. finite sets $\mathcal{X}$, $\mathcal{Y}$, representing the source and target locations of masses (one can also have $\mathcal{X} = \mathcal{Y}$),
2. a source distribution $\sigma = (\sigma(x))_{x \in \mathcal{X}}$ (with $\sigma(x) \geq 0$ for every $x \in \mathcal{X}$)
3. a target distribution $\rho = (\rho(y))_{y \in \mathcal{Y}}$, (with $\rho(y) \geq 0$ for every $y \in \mathcal{Y}$)
4. and a cost function for moving a unit of mass from $x \in \mathcal{X}$ to $y \in \mathcal{Y}$,

$$c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}, \quad (x, y) \mapsto c(x, y), \tag{1}$$

the solution to Monge's problem is a function $T : \mathcal{X} \to \mathcal{Y}$ that transports σ into ρ, minimizing the total transport cost, defined as

$$\sum_{x \in \mathcal{X}} c(x, T(x))\sigma(x). \tag{2}$$

The condition that T transports σ into ρ can be stated as the constraint

$$\sum_{x \in T^{-1}(y)} \sigma(x) = \rho(y), \quad \text{for every} y \in \mathcal{Y}. \tag{3}$$

By summation upon $y \in \mathcal{Y}$, the total masses of the two distributions must be equal. Therefore, after a simple scaling it is sufficient to consider only probability distributions.

Kantorovich

Although the original formulation of the problem in terms of maps seems intuitive, it has limitations. Trivial examples demonstrate that for general probability distributions σ and ρ, transport maps satisfying (3) may not exist, for one is forced to "split" the mass at one site and locate it to different target sites. To address this challenge, L. Kantorovich extended Monge's approach by introducing the concept of coupling π, i.e., a (joint) probability distribution on $\mathcal{X} \times \mathcal{Y}$ with marginals σ and ρ, or equivalently by a transport plan, where the "deterministic" image $T(x) \in \mathcal{Y}$ is replaced by a probability distribution $\pi(\cdot|x) = \pi(x, \cdot)/\sigma(x)$ over the target sites in $\mathcal{Y}$. The set of couplings (or plans) is a closed and convex polytope and the Kantorovich cost reads

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} c(x, y)\pi(x, y) = \sum_{x \in \mathcal{X}} \sigma(x) \sum_{y \in \mathcal{Y}} c(x, y)\pi(y|x), \tag{4}$$

that is linear with respect to the target variable π, hence the problem fits into the linear programming framework. In fact, it was the study of this and related problems that eventually lead to the birth of linear programming as a subject.

The Wasserstein distance

If $\mathcal{X} = \mathcal{Y}$ and the cost $c(x, y) = d(x, y)$ is a distance, Kantorovich's optimal transport cost

$$W_1(\sigma, \rho) = \min_{\pi} \sum_{x, y \in \mathcal{X}} d(x, y)\pi(x, y) \tag{5}$$

induces a distance between probability distributions over $\mathcal{X}$, sometimes referred to as the Earth Mover's distance, but more broadly called Wasserstein distance, although the role of Wasserstein is rather marginal [8]. It measures the minimum amount of distance needed to shape the distribution σ into ρ, where splitting of masses is also allowed. Actually, for every $p \geq 1$, one can define the Wasserstein distance of order p as

$$W_p(\sigma, \rho) = \min_{\pi \in \mathcal{C}(\sigma, \rho)} \left(\sum_{x, y \in \mathcal{X}} d(x, y)^p \pi(x, y) \right)^{1/p}, \tag{6}$$

which also induces a distance. The case $p = 2$ has become particularly relevant in recent years, starting from the seminal works by F. Otto [9] who, motivated by applications to evolution equations (e.g., of porous media type), used it to develop a Riemannian-like structure on the space of probabilities, also known as Otto's calculus.

Duality

The concept of duality in linear programming problems is fundamental. Roughly, the dual problem is obtained by taking the transpose of the matrix of coefficients defining the primal problem, and exchanging the roles of variables and constraints. In the case of the Wasserstein distance of order 1 its expression is rather simple:

$$W_1(\sigma, \rho) = \max \left\{ \sum_{x \in \mathcal{X}} f(x)(\sigma(x) - \rho(x)) \; : \; |f(x) - f(y)| \leq d(x, y) \, \forall x, y \right\}, \tag{7}$$

i.e., we maximize the difference between the expectations of f with respect to the two probabilities, among all functions f that are 1-Lipschitz with respect to the distance d.

Benamou-Brenier formula

Benamou and Brenier [10] noticed that, in many settings, e.g. for $\mathcal{X} = \mathcal{Y} \subseteq \mathbb{R}^d$ and the Euclidean distance, one can one can naturally interpolate a coupling π between σ and ρ via a continuous curve $(\mu_t)_{t \in [0,1]}$ that evolves according to a continuity equation $\partial_t \mu_t = \text{div}(b_t \mu_t)$, where $(b_t)_{t \in [0,1]}$ is a suitable velocity field, and then equivalently compute the quadratic Wasserstein distance by minimizing a "kinetic energy" functional:

$$W_2^2(\sigma, \rho) = \min_{(\mu_t, b_t)_{t \in [0,1]}} \int_0^1 \int_{\mathbb{R}^d} |b_t|^2 d\mu_t dt. \tag{8}$$

Similar expressions hold as well for different powers, e.g. for W_1 one minimizes a length functional. However, the case $p = 2$ is particulary rich in structure, since it can be a starting point to develop Otto's calculus rigorously.

One should notice however that typically such Benamou-Brenier formulas hold in a continuous setting, e.g., on manifolds or length metric spaces, as the support of μ_t will not be confined to the original set $\mathcal{X}$. Their extension to discrete spaces poses some issues, which were first addressed by Maas [11] and later served as a basis for an analogue construction in quantum settings.

Comparison with other distances

Clearly, the Wasserstein distance is only one (family) among many other distances between probability distributions, such as the total variation distance, the Hellinger distance or the Jensen-Shannon distance (modelled after the relative entropy). When compared with these examples, the Wasserstein distance has the possible advantage of exploiting the underlying geometry on the set $\mathcal{X}$ induced by the distance d. This indeed is a key feature that lead to successful applications in a variety of fields, such as functional inequalities, PDEs (as gradient flows), and geometry (synthetic Ricci curvature bounds) [1, 4]. More recent applications include statistics and machine learning, where it quantifies differences between empirical distributions, aids in data analysis, and serves as a discriminator in generative models [12]. Additionally, it enables geometric interpolation between probabilities and plays a role in proving concentration of measure and other functional inequalities [2, 13].

Despite its usefulness, the Wasserstein distance has drawbacks. In high-dimensional settings, the curse of dimensionality can affect the comparison of empirical distributions, although this issue is not unique to the Wasserstein distance. Computationally, solving the optimization problem for the Wasserstein distance can be expensive. However, approaches such as adding strictly convex terms to the cost [14] or relaxing the Lipschitz condition [12] have been proposed to address these challenges and improve practicality in various scenarios.

3 Quantum systems

Before we describe the theories of quantum optimal transport, we briefly recall some general concepts and notation for quantum systems. Quantum mechanics provides a mathematical framework to describe the behavior of particles at the atomic and subatomic levels.

Table 1 Classical notions and their quantum counterparts

Classical	Quantum
$\mathcal{X}$ (finite set)	$\mathcal{H}$ (finite dimensional)
$x \in \mathcal{X}$	$|\psi\rangle \in \mathcal{H}$
subset $S \subseteq \mathcal{X}$	subspace $V < H$
[0.5ex] $f : \mathcal{X} \to \mathbb{C}$	$A : \mathcal{H} \to \mathcal{H}$ linear; $A \in \mathcal{L}(\mathcal{H})$
f^*	$A^* : \mathcal{H} \to \mathcal{H}$, i.e., the adjoint of A
$|f|^2$	$A^* A$
$f : \mathcal{X} \to \mathbb{R}$	observable $A = A^*$; $A \in \mathcal{O}(H)$
$f : \mathcal{X} \to [0, \infty)$	$A \in \mathcal{O}(\mathcal{H})$, $\sigma(A) \subseteq [0, \infty)$; $A \geq 0$
[0.5ex] $\sum_{x \in \mathcal{X}} f(x)$	$\mathrm{tr}[A]$
probability $(\rho(x))_{x \in \mathcal{X}}$	state $\rho \geq 0$, $\mathrm{tr}[\rho] = 1$; $\rho \in \mathcal{S}(\mathcal{H})$
Dirac delta $x \mapsto 1_{x_0 = x}$	pure state $\rho = |\psi\rangle \langle\psi|$
Cartesian product $\mathcal{X} \times \mathcal{Y}$	Tensor product $\mathcal{H} \otimes \mathcal{K}$
Partial sum $\sum_x f(x, y)$	Partial trace $\mathrm{tr}_{\mathcal{H}}[A]$
Shannon entropy	Von Neumann entropy
$S(\rho) = -\sum_x \rho(x) \log \rho(x)$	$S(\rho) = -\mathrm{tr}[\rho \log \rho]$
Relative entropy	Quantum relative entropy
$D(\rho||\sigma) = \sum_x \rho(x) \ln(\rho(x)/\sigma(x))$	$S(\rho||\sigma) = \mathrm{tr}[\rho(\log \rho - \log \sigma)]$
Markov kernel $(N(x, y))_{x \in \mathcal{X}, y \in \mathcal{Y}}$	Quantum channel $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K})$

A key aspect of this theory is the replacement of commutative objects, such as functions and probabilities, with non-commutative ones represented by operators on complex Hilbert spaces.

To keep the exposition simple, we focus on finite-dimensional systems as an analogy to the case of finite sets and discrete measures, although some formulations of quantum optimal transport are natural on infinite-dimensional spaces, corresponding to continuous variable systems. We recommend any comprehensive monographs such as [15–19] for detailed explanations.

Observables and states

Every quantum system is (postulated) to be associated with a Hilbert space, denoted as $\mathcal{H}$, equipped with a scalar product $\langle\cdot|\cdot\rangle$, conventionally anti-linear in the left argument. For simplicity, we only describe the theory when $\mathcal{H}$ is finite dimensional, and use the same symbol to refer to both the quantum system and its associated Hilbert space. Thus, up to the choice of an orthonormal basis, one can for practical purposes identify $\mathcal{H}$ with some $\mathbb{C}^n$. Following Dirac's notation, we write $|\psi\rangle \in \mathcal{H}$ and $\langle\psi| \in \mathcal{H}^*$ for the corresponding linear functional. One should think of $\mathcal{H}$ as the quantum counterpart of a set, where classical objects of probability and measure theory have their natural quantum counterpart, see Table 1.

The key concepts in quantum mechanics are two (and dual to each other):

1. Quantum observables, that correspond to classical functions or random variables and are represented by self-adjoint operators $A : \mathcal{H} \to \mathcal{H}$, where $\mathcal{H}$ is the associated Hilbert space.

2. Quantum states (also known as density operators), which correspond to classical probability distributions and are described by self-adjoint operators $\rho : \mathcal{H} \rightarrow \mathcal{H}$ that are positive in the sense of quadratic forms and have unit trace $\text{tr}[\rho] = 1$.

We denote the set of linear operators from $\mathcal{H}$ to itself as $\mathcal{L}(\mathcal{H})$, the set of observables as $\mathcal{O}(\mathcal{H})$, and the set of density operators as $\mathcal{S}(\mathcal{H})$, so that $\mathcal{S}(\mathcal{H}) \subseteq \mathcal{O}(\mathcal{H}) \subseteq \mathcal{L}(\mathcal{H})$. Given a state $\rho \in \mathcal{S}(\mathcal{H})$ and an observable $A \in \mathcal{O}(\mathcal{H})$, the expected value of A is defined as $\langle A \rangle_\rho = \text{tr}[A\rho]$.

Given a state $\rho \in \mathcal{S}(\mathcal{H})$, the spectral theorem yields eigenvalues $p_i \in [0, 1]$ (counted with multiplicity) such that $\sum_i p_i = \text{tr}[\rho] = 1$. A state is pure if $\rho = |\psi\rangle \langle\psi| \in \mathcal{S}(\mathcal{H})$, i.e., its spectrum is only $\{0, 1\}$. One may think of pure states as the quantum counterparts of Dirac distributions at a point $x_0 \in \mathcal{X}$. A first difference between classical and quantum theories is that pure states may nevertheless show uncertainty in the outcome of the measurement of some observable A (that is the case when $|\psi\rangle$ is not an eigenvector of A): this is physically interpreted by referring to the pure state as being in a quantum superposition – although mathematically it simply means that $|\psi\rangle$ is a non-trivial linear combination of the eigenvectors of A corresponding to different eigenvalues.

The simplest example of a non-trivial quantum system is the qubit space $\mathcal{H} = \mathbb{C}^2$, with basis $|0\rangle = (1, 0)$, $|1\rangle = (0, 1) \in \mathbb{C}^2$. This provides the quantum analogue of a two-point space $\mathcal{X} = \{0, 1\}$. Observables and states are then naturally represented by 2×2 complex Hermitian matrices. States can be put is natural correspondence with points in the unit ball in $\mathbb{R}^3$ (called Bloch ball) with pure states on its boundary (the Bloch sphere).

Composite systems

The quantum analogue of strings of bits of general length n is obtained via the theory of composite systems. A composite quantum system $\mathcal{H} \otimes \mathcal{K}$ is formed by taking the tensor product of two systems $\mathcal{H}$ and $\mathcal{K}$, with a natural definition of scalar product. Given two linear operators $A \in \mathcal{L}(\mathcal{H})$, $B \in \mathcal{L}(\mathcal{K})$, one naturally induces an operator $A \otimes B$ on the composite system. Conversely, the partial trace operator $\text{tr}_\mathcal{K}$ (over $\mathcal{K}$) naturally maps operators on $\mathcal{H} \otimes \mathcal{K}$ into operators on $\mathcal{H}$, and similarly $\text{tr}_\mathcal{H}$. When applied to observables or states, they provide the non-commutative counterparts of partial integration, i.e., taking marginals, on a product space. In particular, for $\Pi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$, one obtains the so-called reduced density operators $\text{tr}_\mathcal{K}[\Pi] \in \mathcal{S}(\mathcal{H})$ and $\text{tr}_\mathcal{H}[\Pi] \in \mathcal{S}(\mathcal{K})$.

A second difference between classical and quantum theories is the fact that pure states do not necessarily have pure reduced density operators: these states are called entangled and realize non-classical correlations between the two quantum systems. Entangled states are mathematically simple to construct, but they provide a key resource for possible quantum advantage with respect to classical theories.

Quantum channels

The partial trace operators serve as fundamental examples of quantum channels (also called quantum operations), which are the quantum counterparts of classical Markov operators obtained through integration with respect to a probability kernel. A quantum channel from a quantum system $\mathcal{H}$ to a system $\mathcal{K}$ can be defined as a linear, completely positive, and trace-preserving operator $\Phi : \mathcal{L}(\mathcal{H}) \rightarrow \mathcal{L}(\mathcal{K})$. This means in particular that Φ maps states (on $\mathcal{H}$) into states (on $\mathcal{K}$): this represents mathematically the result of a physical interaction of the

system $\mathcal{H}$ with a larger system, leading to a final state on a system $\mathcal{K}$ (which may be equal to $\mathcal{H}$). Not all linear maps that preserve states are quantum channels, but only those represented by a collection of so-called Kraus operators $(B_i)_i$, where $B_i : \mathcal{H} \to \mathcal{K}$ is linear:

$$\Phi(A) = \sum_i B_i A B_i^* \quad \text{for every} A \in \mathcal{L}(\mathcal{H}), \tag{9}$$

and

$$\sum_i B_i^* B_i = \mathbb{1}_{\mathcal{H}}, \tag{10}$$

where $*$ denotes the adjoint and $\mathbb{1}_{\mathcal{H}}$ the identity operator.

An example of a linear transformation of quantum states that is not a quantum channel is the transpose operation $\rho \mapsto \rho^\tau$ (i.e., by seeing ρ as a complex matrix). Indeed, although it maps states into states, one can prove that the partial transpose operation, $M \otimes N \mapsto M \otimes N^\tau$ on any (non-trivial) joint system may fail to preserve states. However, if τ were an actual quantum channel, then such property should hold as well on any joint system.

4 Quantum optimal transport

The needs of quantum computing and communication led to various quantum analogues of distances between classical probability distributions. These include the trace distance, analogue of the total variation distance, the quantum fidelity, which is analogue of the Hellinger distance, and also the quantum relative entropy. Similarly to their classical counterparts, they can be defined for general systems and can be computed or at least approximated with relatively little effort, taking into account the system's dimension. Furthermore, they are not specific to any particular "geometry" in the underlying space, as they are invariant with respect to any change of basis on $\mathcal{H}$. More generally, they are are monotone with respect to the action of any quantum channel Φ from $\mathcal{H}$ into $\mathcal{K}$. Similar properties for the Wasserstein distance are not true nor to be expected, and actually can be used to single out geometric properties of the underlying space (e.g. a function F contracts the Wasserstein distance if and only if it is 1-Lipschitz, but also curvature properties of the space can be revealed by the contraction along the heat semigroup). This motivates the study of distances that are adapted to specific settings.

Taxonomy

In recent years, various proposals for quantum optimal transport problems and induced Wasserstein distances between quantum states have emerged, with diverse applications. In chronological order, the earliest formulation can be traced back to Connes and Lott in 1992 [20], who defined the spectral distance in non-commutative geometry. Another early approach was presented by Zyczkowski and Slomczynski in 1997 [21], where they computed the Wasserstein distance between Husimi probability distributions associated with states in Bosonic systems. In the context of free probability, Biane and Voiculescu proposed an analogue of the Wasserstein metric in [22]. Since 2012, Maas and Carlen [23–25] have developed a distance that formulates a quantum analogue of the classical Benamou-Brenier formula, which provides a continuous-time formulation of the optimal transport problem. In 2013, Agredo [26] proposed a Wasserstein distance that extends any given distance on a set of basis vectors. In 2016, Golse, Mouhot, and Paul introduced a quantum Kantorovich problem

Table 2 Classifying quantum optimal transport

Monge-Kantorovich
- distance between Husimi functions [21]
- transport via couplings [27]
- transport via channels [38]
Dual problem
- spectral distance [20]
- distance on a basis [26]
- Wasserstein distance of order 1 [34]
Benamou-Brenier
- quantum Benamou-Brenier [23]

using quantum couplings, with applications to semiclassical limits of many-body quantum systems [27–30]. In 2021, De Palma and the author proposed a problem based on quantum channels [31], which has found extensions [32] and recent applications in rate-distortion theory [33]. A further different proposal by De Palma, Marvian, Lloyd and the author was done in [34], generalizing somehow the dual point of view on n-qubit systems, and led to applications in quantum state learning [35] and concentration inequalities [36] and the study of limitations of variational quantum algorithms [37].

We can roughly classify all the proposals for quantum optimal transport according to the point of view they most emphasize, in the equivalent formulations of classical optimal transport, i.e., the Monge-Kantorovich problem, the dual formulation, or the Benamou-Brenier formula, see Table 2. However, we point out that all these formulations define convex problems hence they necessarily admit dual versions. In the following subsections, we discuss in more detail the above proposals.

Kantorovich problem

In [21] a metric for measuring distances between quantum states was defined, by computing the Wasserstein distance between the Husimi distributions (or Q-functions) of two given quantum states σ, ρ. The Husimi function is a probability distribution commonly used in quantum optics to represent the phase space distribution of a state (e.g. of light), and can be interpeted as the outcome of a specific quantum operation, which maps quantum states into classical states. Therefore, slightly simplifying the approach from [21] into our setting where Hilbert spaces are finite-dimensional, the proposal goes as follows: first, measure the quantum states and record the possible outcomes with the associated (classical) probabilities; then, solve an optimal transport problem between such classical probabilities. Despite this simplicity, the authors argued that the resulting distance exhibits properties that may be relevant for studying the semiclassical limit of quantum mechanics, and they illustrate their case by computing the distance in various examples, e.g. coherent states, squeezed states and between Fock states.

The second proposal that fits into this framework was first formulated in [27], with applications in the study of mean-field and classical limits of quantum evolutions. The main problem they address is to quantify two effective limits for quantum systems: the mean-field framework, which describes systems with many interacting particles (formally, a number $n \to \infty$ of particles), and the classical limit, where roughly speaking one lets the Planck's constant $\hbar \to 0$. Searching for bounds in terms of a suitable quantum optimal transport is

motivated by well-known fact that Wasserstein distance quantifies mean-field limits of classical interacting particle systems, at least if the interaction potential is sufficiently smooth and coercive. Golse, Mouhot and Paul in [27] introduced the following quantum analogue of the left hand side of (4):

$$W_{GMP}(\sigma, \rho)^2 = \min_{\Pi \in \mathcal{C}_{GMP}(\sigma, \rho)} \text{tr}\,[C\Pi], \tag{11}$$

where σ, ρ are given source and target quantum states in $\mathcal{H} = L^2(\mathbb{R}^d)$, the cost

$$C = (Q \otimes \mathbb{1}_{\mathcal{H}} - \mathbb{1}_{\mathcal{H}} \otimes Q)^2 + (P \otimes \mathbb{1}_{\mathcal{H}} - \mathbb{1}_{\mathcal{H}} \otimes P)^2 \tag{12}$$

is a sum of squared increments of position Q and momentum P observables and Π belongs to the set of quantum couplings $\mathcal{C}_{GMP}(\sigma, \rho)$. These are defined as density operators on the joint system $\mathcal{H} \otimes \mathcal{H}$, with reduced density operators respectively given by σ and ρ. In this case, the Hilbert space $\mathcal{H}$ is infinite-dimensional and the observables are unbounded self-adjoint operators – in order to obey the canonical commutation relation $[Q, P] = i\hbar\mathbb{1}_{\mathcal{H}}$: this is motivated by the application to particle systems, but one can also consider simpler variants on finite dimensional systems or use bounded observables. If compared with the first proposal, i.e., [21], we see here that the "transport" is performed at the quantum level, i.e., the coupling directly involves the two quantum states. This allows for a better integration with the bounds (of Grönwall type) they obtained studying the transport cost along the evolution dynamics. However, differently from the distance between the Husimi functions, we notice that $W_{GMP}(\sigma, \sigma)$ may be strictly positive, hence it is not an actual distance. For a detailed presentation, we refer to the notes in the upcoming monograph [3].

The third proposal, from [38], is based instead on the quantum analogue of the right hand side of (4), using quantum channels instead of couplings. In the classical setting, it amounts to replace couplings $\pi(x, y)$ with plans $\pi(y|x) = \pi(x, y)/\sigma(x)$, an operation however that has no quantum counterpart. Indeed, due to entangled states, there is no analogue (in general) of conditional distributions for quantum states. However, one can directly formulate a minimization problem over quantum channels Φ that map σ into ρ, i.e., $\Phi(\sigma) = \rho$, which may be seen as the counterpart of the classical $T_\sharp(\sigma) = \rho$. Denoting such set with $\mathcal{C}_{DT}(\sigma, \rho)$, one can setup a correspondence between such states and couplings, however, one does not recover exactly the two marginals σ and ρ, but instead obtains σ^τ (the transpose) and ρ. Thus, $\mathcal{C}_{GMP}(\sigma, \rho)$ and $\mathcal{C}_{DT}(\sigma, \rho)$ differ by a partial transpose operation which we already remarked above is not a quantum channel: this ultimately yields a different notion of distance, even if the two definitions appear at the beginning very similar. In [3], it is noticed that, when the cost observable $C = \sum_i (R_i \otimes \mathbb{1} - \mathbb{1} \otimes R_i)^2$ is a sum of squares, as in (12) but for a general set of observables $(R_i)_i$, by developing the square one obtains a direct formulation in terms of channels:

$$W_{DT}^2(\sigma, \rho) = \min_{\Phi(\sigma)=\rho} \sum_i \left(\text{tr}[R_i^2\sigma] + \text{tr}[R_i^2\rho] - 2\,\text{tr}[R_i\,\sqrt{\sigma}\,\Phi^\dagger(R_i)\,\sqrt{\sigma}] \right), \tag{13}$$

It turns out that W_{DT} shares many properties with W_{GMP}, such as the upper and lower bounds that are employed in the main results from [27] and the fact that $W_{DT}(\sigma, \sigma)$ can be strictly positive – hence it is not an actual distance. Furthermore one can show that the identity channel Φ is always optimal when computing the distance from a state to itself, as well as establish a "modified" triangle inequality

$$W_{DT}(\sigma, \rho) \le W_{DT}(\sigma, \tau) + W_{DT}(\tau, \tau) + W_{DT}(\tau, \rho). \tag{14}$$

We remark however that main conceptual difference between W_{GMP} and W_{DT} is the fact that the optimal coupling can be interpreted in the latter case as a physical operation. It would be interesting to understand whether employing W_{DT} instead of W_{GMP} in the problem of classical limit of many body quantum systems may provide further relevant information. Finally, we mention that W_{DT} has recently found application in the study of quantum rate-distortion theory [33], i.e., in quantifying fundamental bounds for lossy transmission rates of quantum information.

Dual formulation

It is straightforward to check that the quantum optimal transport problems defined in the previous subsection are convex optimization problems, and therefore they admit a dual formulation – at least for the finite dimensional quantum systems: in the infinite dimensional framework, duality is developed in [30] for W_{GMP}. In this section, we focus however on quantum optimal transport problems that are proposed from the very beginning in what classically is the "dual" formulation.

In their seminal contribution, Lott and Connes [20] focused on the metric properties of non-commutative geometry. They proposed a new notion of non-commutative metric space, by introducing a triple $(\mathcal{A}, \mathcal{H}, D)$, consisting of a Hilbert space $\mathcal{H}$, an involutive algebra $\mathcal{A}$ of operators on $\mathcal{H}$, and a selfadjoint "Dirac" operator D on $\mathcal{H}$. The key observation is that, in smooth commutative settings, where $\mathcal{A}$ reduce to usual complex-valued functions, the Lipschitz norm of a function $f \in \mathcal{A}$ is obtained as the norm of the commutator $[D, f]$ – the commutator $[D, \cdot]$ acting as a derivation. Recalling the classical duality (7), this leads to the definition of the spectral distance between states (that are in this setting positive normalized linear functionals acting on $\mathcal{A}$)

$$W_{LC}(\sigma, \rho) := \sup_{\|[D,A]\| \leq 1} \langle A \rangle_\sigma - \langle A \rangle_\rho. \tag{15}$$

They demonstrate that this framework captures various examples of spaces, including Riemannian manifolds, finite spaces, spaces with non-integer Hausdorff dimension, group rings of discrete subgroups of Lie groups, configuration spaces in supersymmetric quantum field theory, and "quantum" tori. They develop a differential calculus on non-commutative spaces that reproduces the differential forms calculus on Riemannian manifolds, using operator theoretic tools instead of traditional differential and integral calculus. The connection with optimal transport and in particular the Wasserstein distance of order 1 was explored in subsequent works by other authors, see e.g. [39, 40].

As a second proposal of quantum Wasserstein distance naturally formulated in dual terms, we mention Agredo's work [26]. Motivated by the problem of measuring deviations from equilibrium in quantum Markov semigroups, i.e., one-parameter families of quantum channels, he defines a distance $W_A(\sigma, \rho)$ over the states $\sigma, \rho \in \mathcal{S}(\mathcal{H})$ starting from any chosen orthonormal basis $(e_i)_{i \in I}$ of $\mathcal{H}$ and a (usual) distance function $d : I \times I \to [0, \infty)$ over the index set of the basis. The definition resembles again (4) and (15), but the set of 1-Lipschitz observables is given by those $A \in \mathcal{O}(\mathcal{H})$ such that

$$\left\| [|e_i\rangle\langle e_j| + |e_j\rangle\langle e_i|, A] \right\| \leq d(i, j) \quad \text{for every} i, j \in I. \tag{16}$$

He shows that the distance between states that are diagonal with respect to the chosen basis, hence can be identified with classical probability distributions over I, coincides with the classical Wasserstein distance of order 1 with respect to the chosen distance d. This property

may be useful when diagonal quantum states are used to codify classical probabilities, e.g. when transmitting information, and does not hold in general for other distances – see e.g. [28] for W_{GMP}. The key result in [26] is a characterization of a quantum version of the detailed balance condition for a quantum Markov semigroup in terms of an "entropy rate" defined in terms of the resulting Wasserstein distance obtained by choosing an orthonormal basis that diagonalizes the invariant state of the semigroup.

The third proposal that we include here is the quantum Wasserstein distance of order 1 for systems of n qubits first introduced in [34]. In this case, the system $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ is a composition of n single qubit systems (hence $\dim(\mathcal{H}) = 2^n$), and the aim is to define a distance between quantum states that looks like the classical Wasserstein distance with respect to the Hamming distance between strings (also called Ornstein's $\bar{d}$ distance in the stochastic processes literature). The distance can be naturally defined as a supremum as in (15), for a suitable notion of 1-Lipschitz observables $A \in \mathcal{O}(\mathcal{H})$, which in this case reads as the condition

$$2 \max_{i=1,...,n} \min_{A^{(i)}} \left\| A - \mathbb{1}_i \otimes A^{(i)} \right\| \leq 1, \tag{17}$$

where $A^{(i)}$ is any observable over the system where the i-th qubit has been removed (and $\mathbb{1}_i$ denotes the identity over the single qubit system at i). We refer also to a dedicated chapter in [3] for a more detailed exposition. Here, we notice that the resulting distance $W_1(\sigma, \rho)$ is enjoys several desirable properties. For example, it can be upper and lower bounded by the trace distance (but not uniformly with respect to n), it recovers the classical Wasserstein-Hamming distance for diagonal states in the computational basis, and the von Neumann entropy is continuous with an explicit modulus of continuity. Even more relevant for application is the fact that the distance can be used as a tool to establish concentration inequalities for Lipschitz observables [36] or as a cost in training quantum machine learning models [35]. As a further relevant research direction, we mention that in [41] the re-normalized limit as $n \to \infty$ was studied, with possible applications to quantum dynamical systems.

Benamou-Brenier

We end this section by briefly discussing the approach put forward by Maas and Carlen in [23, 24] (see also the dedicated chapter in [3]). Their objective is to introduce a metric on quantum states that may allow for similar computations as in Otto's calculus on probability distributions, possibly leading to novel functional inequalities, in particular modified log-Sobolev inequalities and contraction rates for quantum Markov semigroup (hence in a setting similar to Agredo's [26]).

Otto's calculus yields, for many ergodic Markov diffusion semigroups in the commutative setting, that one can interpret the semigroup as the gradient flow of the relative entropy with respect to the (unique) invariant distribution. A similar situation may hold for quantum Markov semigroups, where the quantum relative entropy (see Table 1) always decreases along the semigroup. Maas and Carlen therefore search for a Wasserstein-like metric such that the semigroup can be recovered as gradient flow, in a similar way. The key point is that, once such a metric is defined, besides investigating its properties (in particular its geodesics), if the quantum relative entropy happens to be convex along the geodesics, this would imply log-Sobolev inequalities and contraction rates for the semigroup.

Without entering too much in technical details, they search for a quantum Benamou-Brenier formula analogue to (8) and realize that one needs a suitable notion of continuity equation and a Riemannian like metric on the tangent space to the quantum states (in order

to define the energy as the integral of the metric). Not all the metrics however play the same role, since the key identity used in Otto's calculus they need to replicate reads

$$\Delta\rho = \operatorname{div}(\rho\nabla\log\rho).$$

This is a trivial consequence of the chain rule in the Euclidean setting (or on manifolds), but already not obvious in discrete settings, much less in the quantum case. Their metric eventually is quite explicitly defined, although the computations in actual cases may become a bit cumbersome, but see e.g. [23, section 6] for examples.

When compared with the previous approaches, it seems reasonable to conjecture that the metric built from the Benamou-Brenier formula could appear as the length distance with respect to some Kantorovich like problem. To the author's knowledge, no connection has been discovered so far. This may be possibly due the fact that the continuity equation defined by Carlen and Maas does not seem in general to describe the physical evolution of a quantum state, hence it cannot be lifted to a quantum channel. By comparison, the classical analogue of this property is often known as the superposition principle and one can argue that it holds in extreme generality [42].

5 Conclusion

We briefly presented old and new approaches to optimal transport problems for quantum systems. Several alternatives have been explored, and possibly other ones will be introduced, since this research field has become quite active: we mention also the recent works [32, 43–46]. Such diversity is a valuable resource, as different distances may prove better suited for particular applications, whether in quantum state tomography, computing or machine learning [35, 37, 47–50], or the analysis of quantum dynamical systems [51–53].

Some particularly promising directions for future research include the design of efficient classical and quantum algorithms for computing such distances, as well as a deeper investigation into the geometric properties induced by the optimal transport structure. Analyzing the structural characteristics of the optimizers themselves, and elucidating the connections between the various proposed frameworks, are also likely to lead to important new developments.

Overall, the field continues to evolve and novel ideas will be built upon and applied, enhancing our understanding and control of quantum systems and their dynamics.

Declarations

References

1. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)
2. Peyré, G., Cuturi, M.: Computational optimal transport: With applications to data science. Found. Trends Mach. Learn. **11**(5–6), 355–607 (2019)
3. Maas, J., Rademacher, S., Titkos, T., Virosztek, D. (eds.): Optimal Transport on Quantum Structures. Bolyai Society Mathematical Studies, Berlin (2024)
4. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: in Metric Spaces and in the Space of Probability Measures. Springer, Berlin (2005)
5. Santambrogio, F.: Optimal transport for applied mathematicians. Birkhäuser **55**(58–63), 94 (2015)
6. Figalli, A., Glaudo, F.: An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows. EMS Textbooks in Mathematics. EMS Press, Berlin (2021)
7. Monge, G.: Mémoire sur la théorie des déblais et des remblais. Mem. Math. Phys. Acad. Royale Sci., 666–704 (1781)
8. Vershik, A.M.: Long history of the monge-kantorovich transportation problem:(marking the centennial of lv kantorovich's birth!). Mathematical Intelligencer **35**(4) (2013)
9. Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Comm. Partial Differential Equ. **26**(1–2), 101–174 (2001)
10. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. Numerische Mathematik **84**(3), 375–393 (2000)
11. Maas, J.: Gradient flows of the entropy for finite markov chains. J. Funct. Anal. **261**(8), 2250–2292 (2011)
12. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223 (2017). PMLR
13. Gozlan, N., Léonard, C.: Transport inequalities a survey. Markov Processes Related Fields **16**, 635–736 (2010)
14. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems **26** (2013)
15. Alicki, R., Fannes, M.: Quantum Dynamical Systems, p. 278. Oxford University Press, Oxford (2001)
16. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2010)
17. Bengtsson, I., Życzkowski, K.: Geometry of Quantum States: an Introduction to Quantum Entanglement. Cambridge University Press, Cambridge (2017)
18. Holevo, A.S.: Quantum Systems, Channels, Information: a Mathematical Introduction. Walter de Gruyter GmbH & Co KG, (2019)
19. Moretti, V.: Fundamental Mathematical Structures of Quantum Theory. Springer, Berlin (2019)
20. Connes, A., Lott, J.: The metric aspect of noncommutative geometry. In: New Symmetry Principles in Quantum Field Theory, pp. 53–93. Springer, (1992)
21. Zyczkowski, K., Slomczynski, W.: The monge distance between quantum states. J. Phys. A **31**(45), 9095 (1998)
22. Biane, P., Voiculescu, D.: A free probability analogue of the wasserstein metric on the trace-state space. Geometric & Funct. Anal. GAFA **11**(6), 1125–1138 (2001)
23. Carlen, E.A., Maas, J.: An analog of the 2-wasserstein metric in non-commutative probability under which the fermionic fokker-planck equation is gradient flow for the entropy. Commun. Math. Phys. **331**(3), 887–926 (2014)
24. Carlen, E.A., Maas, J.: Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. J. Funct. Anal. **273**(5), 1810–1869 (2017)
25. Carlen, E.A., Maas, J.: Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. J. Stat. Phys. **178**(2), 319–378 (2020)

26. Agredo, J.: A wasserstein-type distance to measure deviation from equilibrium of quantum markov semigroups. Open Syst. Inform. Dynam. **20**(02), 1350009 (2013)
27. Golse, F., Mouhot, C., Paul, T.: On the mean field and classical limits of quantum mechanics. Commun. Math. Phys. **343**(1), 165–205 (2016)
28. Caglioti, E., Golse, F., Paul, T.: Quantum optimal transport is cheaper. J. Stat. Phys. **181**, 149–162 (2020)
29. Golse, F., Paul, T.: Quantum and semiquantum pseudometrics and applications. Journal of Functional Analysis (2022)
30. Caglioti, E., Golse, F., Paul, T.: Towards optimal transport for quantum densities. Annali della Scuola Normale Superiore di Pisa, Classe di Scienze **XXIV**(4) (2023)
31. Gosson, M.A.: Quantum harmonic analysis. In: Quantum Harmonic Analysis. De Gruyter, (2021)
32. Duvenhage, R.: Quadratic wasserstein metrics for von neumann algebras via transport plans. J. Operator Theory **88**(2), 289–308 (2022)
33. Garmaroudi, H.M., Pradhan, S.S., Chen, J.: Rate-limited quantum-to-classical optimal transport: A lossy source coding perspective. In: 2023 IEEE International Symposium on Information Theory (ISIT), pp. 1925–1930 (2023). IEEE
34. De Palma, G., Marvian, M., Trevisan, D., Lloyd, S.: The quantum wasserstein distance of order 1. IEEE Trans. Inform. Theory **67**(10), 6627–6643 (2021)
35. Kiani, B.T., De Palma, G., Marvian, M., Liu, Z.-W., Lloyd, S.: Learning quantum data with the quantum earth mover's distance. Quantum Sci. Technol. **7**(4), 045002 (2022)
36. De Palma, G., Rouzé, C.: Quantum concentration inequalities. Ann. Henri Poincaré **23**(9), 3391–3429 (2022)
37. De Palma, G., Marvian, M., Rouzé, C., Franca, D.S.: Limitations of variational quantum algorithms: a quantum optimal transport approach. PRX Quantum **4**(1), 010309 (2023)
38. De Palma, G., Trevisan, D.: Quantum optimal transport with quantum channels. Ann. Henri Poincaré **22**(10), 3199–3234 (2021)
39. Rieffel, M.A.: Metrics on state spaces. Docum. Math. **4**, 559–600 (1999)
40. D'andrea, F., Martinetti, P.: A view on optimal transport from noncommutative geometry. SIGMA. Symmetry, Integrability and Geometry: Methods and Applications 6, 057 (2010)
41. De Palma, G., Trevisan, D.: The wasserstein distance of order 1 for quantum spin systems on infinite lattices. Annales Henri Poincaré **24**, 4237–4282 (2023)
42. Stepanov, E., Trevisan, D.: Three superposition principles: currents, continuity equations and curves of measures. J. Funct. Anal. **272**(3), 1044–1103 (2017)
43. Tóth, G., Pitrik, J.: Quantum wasserstein distance based on an optimization over separable states. Quantum **7**, 1143 (2023)
44. Feliciangeli, D., Gerolin, A., Portinale, L.: A non-commutative entropic optimal transport approach to quantum composite systems at positive temperature. J. Funct. Anal. **285**(4), 109963 (2023)
45. Bistroń, R., Eckstein, M., Życzkowski, K.: Monotonicity of a quantum 2-wasserstein distance. J. Phys. A **56**(9), 095301 (2023)
46. Duvenhage, R., Mapaya, M.: Quantum wasserstein distance of order 1 between channels. Infinite Dimensional Analysis, Quantum Probability and Related Topics (2023)
47. Chakrabarti, S., Yiming, H., Li, T., Feizi, S., Wu, X.: Quantum wasserstein generative adversarial networks. Advances in Neural Information Processing Systems **32** (2019)
48. Rouzé, C., França, D.S.: Learning quantum many-body systems from a few copies. Quantum **8**, 1319 (2024)
49. Onorati, E., Rouzé, C., França, D.S., Watson, J.D.: Efficient learning of ground & thermal states within phases of matter. arXiv:2301.12946 (2023)
50. Hirche, C., Rouzé, C., França, D.S.: Quantum differential privacy: An information theory perspective. IEEE Trans. Inform. Theory **69**, 5771–5787 (2023)
51. Capel, A., Rouzé, C., Franca, D.S.: The modified logarithmic sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions. arXiv preprint arXiv:2009.11817 (2020)
52. Gao, L., Rouzé, C.: Ricci curvature of quantum channels on non-commutative transportation metric spaces. arXiv:2108.10609 (2021)
53. Gao, L., Rouzé, C.: Complete entropic inequalities for quantum markov chains. Archive for Rational Mechanics and Analysis **245**(1), 183–238 (2022)