QUANTUM OPTIMAL TRANSPORT: QUANTUM CHANNELS AND QUBITS

GIACOMO DE PALMA AND DARIO TREVISAN

ABSTRACT. These notes are based on the lectures given by the second author at the School on Optimal Transport on Quantum Structures at Erdős Center in September 2022. The focus of the exposition is on two recently introduced approaches on quantum optimal transport: one based on quantum channels as generalized transport plans, the other based on the notion of Hamming-Wasserstein distance of order 1 on multiple-qubit systems. The material is presented in an elementary manner with a focus on the finite-dimensional setting.

1. INTRODUCTION

Quantum Optimal Transport is a rapidly growing field at the intersection of quantum mechanics and optimal transport theory. While optimal transport theory searches for the most efficient way to transport resources or information from one location to another, in the quantum setting such a problem becomes more challenging due to the non-commutativity of the involved quantities, i.e., states and observables. In recent years, there has been a surge of interest in quantum optimal transport, from both theoretical and computational perspectives, making it an exciting and promising area of research.

These lecture notes closely follow the material presented by the second author in a series of lectures on quantum optimal transport given at the Erdős Center, Budapest, in September 2022. The main focus is on two recent different proposals for defining quantum optimal transport and their properties. The first proposal, introduced in [22], is based on the use of quantum channels to transport one quantum state onto another. The second proposal, following [20], is well-suited for composite systems, in particular multiple-qubit systems, and is based on a natural notion of neighbouring states. Both approaches provide distinct and complementary perspectives on the problem of quantum optimal transport, and we believe they provide significant examples and applications in the field. The notes aim to a detailed and accessible introduction to these proposals, in particular focusing on the mathematical foundations, and highlight some key concepts and some elementary applications, e.g. to concentration inequalities.

The exposition is structured as follow: in Section 2, we review some basic notions on the classical optimal transport problem and the related Wasserstein distances. Section 3 collects some elementary facts on quantum systems, states, observables and quantum channels. Section 4 reviews some literature on quantum optimal transport, providing a rough classification of

Date: November 23, 2023.
the main approaches to the subject. Section 5 is devoted to the formulation of quantum optimal transport using quantum channels, while Section 6 focuses on the quantum Wasserstein distance on qubit systems.

Although the field is rapidly evolving, we hope that the readers may gain from these notes a basic mathematical understanding of some of the latest developments in quantum optimal transport and their potential impact on the field of quantum information science.

Acknowledgments. We thank the organizers of the School and Workshop on Optimal Transport on Quantum Structures at Erdös Center, J. Maas, S. Rademacher, T. Titkos, D. Virosztek, as well as the other lecturers, A. Figalli, E.A. Carlen and F. Golse, as well as the speakers and participants, for many stimulating conversations. We also acknowledge three anonymous referees for many suggestions and improvements in the manuscript.

GDP has been supported by the HPC Italian National Centre for HPC, Big Data and Quantum Computing - Proposal code CN00000013, CUP J33C22001170001 and by the Italian Extended Partnership PE01 - FAIR Future Artificial Intelligence Research - Proposal code PE00000013, CUP J33C22002830006. GDP is a member of the “Gruppo Nazionale per la Fisica Matematica (GNFM)” of the “Istituto Nazionale di Alta Matematica “Francesco Severi” (INdAM)”.

DT has been supported by the HPC Italian National Centre for HPC, Big Data and Quantum Computing - Proposal code CN1 CN00000013, CUP I53C22000690001, and by the INdAM-GNAMPA project 2023 “Teoremi Limite per Dinamiche di Discesa Gradiente Stocastica: Convergenza e Generalizzazione”.

2. Classical Optimal Transport

In this section, we review some notation and basic facts about the classical optimal transport theory. We provide no proofs and focus on the elementary setting of finite sets and hence discrete measures: the general theory is exposed in several excellent monographs, e.g. [48, 3, 46, 28, 43].

2.1. Monge’s problem. Historically, G. Monge’s seminal work on optimal transport theory, published in 1781, laid the foundations for a mathematical framework to study the optimal ways of transporting goods or mass from one location to another. He introduced the idea of finding an optimal transportation map as a way to minimize the total cost, or energy, required to move mass between two locations, and it provided a first mathematical framework to study this problem.

Although Monge considered only absolutely continuous distributions of mass (i.e., densities with respect to 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 positions of the masses (without loss of generality, one usually assumes $\mathcal{X} = \mathcal{Y}$),
2. a source distribution of mass $\sigma = (\sigma(x))_{x \in \mathcal{X}}$ (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 $y \in \mathcal{Y}$)
and a cost function for transporting a unit of mass from position $x \in \mathcal{X}$ to position $y \in \mathcal{Y}$,

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

(2.1)

Monge’s optimal transport problem searches for a transport map $T : \mathcal{X} \to \mathcal{Y}$ that moves $\sigma$ into $\rho$ with minimal total transport cost, defined as

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

(2.2)

The condition that $T$ moves $\sigma$ into $\rho$ can be stated as

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

(2.3)

Summation upon $y \in \mathcal{Y}$, yields immediately that the total masses of the two distributions must be equal, hence, due to homogeneity of (2.2), one usually considers only probability distributions.

2.2. The Kantorovich problem. Despite being a natural formulation of the problem, simple examples show that transport maps $T$ satisfying (2.3) may not exist for general probability distributions $\sigma, \rho$. To overcome this issue, L. Kantorovich extended Monge’s original approach by considering the possibility of splitting mass into fractions and directing it to different target locations. He introduced the concept of a coupling as a probability distribution over the product space and formulated the problem of finding an optimal coupling as a linear programming problem.

Precisely, a coupling $\pi$ between $\sigma$ and $\rho$ is defined as a probability distribution on the product set $\mathcal{X} \times \mathcal{Y}$, such that

$$\sum_{x \in \mathcal{X}} \pi(x, y) = \rho(y), \quad \sum_{y \in \mathcal{Y}} \pi(x, y) = \sigma(x), \quad \text{for every } x \in \mathcal{X}, y \in \mathcal{Y},$$

(2.4)

i.e., the marginal distributions are respectively $\sigma$ and $\rho$. Let us denote with $C(\sigma, \rho)$ the set of couplings between $\sigma$ and $\rho$, which is always non empty, since it contains the product distribution $\pi(x, y) = \sigma(x)\rho(y)$. The transport cost associated to a coupling $\pi$ is defined as the expectation

$$\langle c \rangle_{\pi} = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} c(x, y)\pi(x, y).$$

(2.5)

Kantorovich formulation extends Monge’s problem, since any transport map $T : \mathcal{X} \to \mathcal{Y}$ that moves $\sigma$ into $\rho$ in the sense of (2.3), is naturally associated to the coupling $\pi(x, y) = \sigma(x)1_{\{y = T(x)\}}$, and $\langle c \rangle_{\pi}$ equals (2.2).

Moreover, minimization over all couplings, keeping $\sigma$ and $\rho$ fixed, is a linear programming problem, i.e. on optimization problem of a linear cost functional with linear constraints in the variables, that can be solved e.g. via the simplex algorithm.

2.3. The Wasserstein distance. When $\mathcal{X} = \mathcal{Y}$ and the cost $c(x, y) = d(x, y)$ is a distance, the optimal transport cost

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

(2.6)
induces a distance between probability distributions over \( \mathcal{X} \), usually called Wasserstein distance, but also known as the Earth Mover’s distance, as it measures the minimum amount of work needed to transform one probability distribution into another, where the amount of work is proportional to the distance travelled by each unit of mass during transportation. More precisely, \( W_1 \) is called Wasserstein distance of order 1, since for every \( p \geq 1 \), one can define the Wasserstein distance of order \( p \) as

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

which also induces a distance. For \( p \in (0, 1) \), one defines instead

\[
W_p(\sigma, \rho) = \min_{\pi \in C(\sigma, \rho)} \sum_{x, y \in \mathcal{X}} d(x, y)^p \pi(x, y),
\]

which is a Wasserstein distance of order 1, with respect to the distance \((x, y) \mapsto d(x, y)^p\).

Kantorovich also introduced the fundamental concept of duality in linear programming problems. It is a general tool that establishes a relationship between two different linear programming problems, usually referred to as the primal and the dual one. Roughly speaking, 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, with respect to a distance \( d \), its expression is particularly simple:

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

i.e., we maximize the difference between the expectations \( \langle f \rangle_\sigma - \langle f \rangle_\rho \) among all functions \( f \) that are 1-Lipschitz with respect to the distance \( d \).

2.4. Comparison with other distances. Of course, there are plenty of other distances between probability distributions, such as

a) the Total Variation distance

\[
\| \rho - \sigma \|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\sigma(x) - \rho(x)|
\]

b) the Hellinger distance

\[
H(\sigma, \rho) = \sqrt{\frac{1}{2} \sum_{x \in \mathcal{X}} (\sqrt{\sigma(x)} - \sqrt{\rho(x)})^2}
\]

c) the relative entropy or Kullback-Leibler divergence

\[
D_{KL}(\sigma || \rho) = \sum_{x \in \mathcal{X}} \sigma(x) \ln \left( \frac{\sigma(x)}{\rho(x)} \right),
\]

provided that \( \sigma(x) = 0 \) whenever \( \rho(x) = 0 \), otherwise we set it to \(+\infty\) (although this is not in general an actual distance, it is commonly employed to compare two distributions).
Comparing the definitions above with (2.6), it is clear that a possible advantage of the Wasserstein distance is to make use of the underlying geometry on the set $\mathcal{X}$ induced by the distance $d$. This is indeed the case, and the Wasserstein distance has found numerous applications in many fields. If we limit ourselves to statistics and machine learning, one of its main applications is in quantifying the difference between empirical distributions of sampled points, which has immediate implications in data analysis. In particular, the Wasserstein distance has been proposed as a discriminator in generative models [4], where it measures the distance between the generated and real distributions. Additionally, the Wasserstein distance can be used as a tool for geometric interpolation between probabilities, allowing for natural transformation of one distribution into another [43]. The Wasserstein distance is also a theoretical tool that can be used to prove concentration of measure and other functional inequalities [34], which has important implications in probability theory and statistics. The versatility and usefulness of the Wasserstein distance have made it a central concept in modern mathematics and its applications continue to be explored in various research areas.

While the Wasserstein distance has many useful applications, it also has some downsides compared to other metrics such as total variation or relative entropy. One of the main drawbacks becomes apparent in high-dimensional settings where the curse of dimensionality can become an issue when comparing empirical distributions of data – but to be fair, this issue is common also for other distances. In addition to this, the computation of the Wasserstein distance involves solving an optimization problem, which can be computationally expensive. However, several solutions have been proposed to address these issues. For example, strictly convex terms can be added to the cost [18], to improve the convergence of iterative algorithms, or in the dual formulation, the notion of Lipschitz functions can be relaxed, for instance, by parametrizing over a class of neural networks [4]. These solutions can help overcome the computational challenges associated with the Wasserstein distance and make it more practical to use in a variety of settings.

**Exercise 2.1.** Show that the Total variation distance is the Wasserstein distance with respect to the trivial distance $d(x, y) = 1_{\{x \neq y\}}$. Deduce that

$$\lim_{p \to +\infty} W_p(\sigma, \rho) = \|\rho - \sigma\|_{TV}. \quad (2.13)$$

3. **Quantum systems**

Quantum mechanics is a mathematical framework that describes the behaviour of particles at the atomic and subatomic level. One of the key features of the theory is that it replaces commutative objects, such as functions and probabilities, with non-commutative ones, precisely given as operators on a complex Hilbert space. To keep the exposition simple and avoid topological considerations, in this notes we focus only on the case of finite dimensional systems, which should be considered as the analogue of the case of finite sets and discrete measures from the previous section. Again, we give no proofs of the basic facts recalled in this section, and we recommend to
the interested reader any of the excellent monographs [2, 40, 6, 36, 39] for a
detailed exposition.

3.1. Systems, states and observables. Any quantum system has an
associated Hilbert space $\mathcal{H}$ with a scalar product $\langle \cdot | \cdot \rangle$, which is conventionally
anti-linear in the left argument. With a small abuse of notation, in these
notes we will identify the quantum system with its Hilbert space and denote
them with the same symbol. Following Dirac’s notation, we write $|\psi\rangle \in \mathcal{H}$
and $\langle \psi| \in \mathcal{H}^*$ for the linear functional
$$\mathcal{H} \ni |\varphi\rangle \mapsto \langle \psi|\varphi\rangle.$$ (3.1)

Many classical objects of probability and measure theory have their natural
quantum counterpart. The correspondence is summarized in Table 1 below.
The most relevant ones are:

1. quantum (real-valued) observables, which correspond to classical
functions (or random variables) and are given by linear self-adjoint
operators $A : \mathcal{H} \to \mathcal{H}$, i.e., $\langle \psi|A\varphi\rangle = \langle A\psi|\varphi\rangle$ for every $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$.

2. quantum states, which correspond to classical probability distribu-
tions, given by operators $\rho : \mathcal{H} \to \mathcal{H}$ that are self-adjoint, positive
in the sense of quadratic forms, i.e., $\langle \psi|\rho\psi\rangle \geq 0$ for every $|\psi\rangle \in \mathcal{H}$,
and with unit trace $\text{tr}[\rho] = 1$. Such operators are also called density
operators.

We write $\mathcal{L}(\mathcal{H})$ for the set of linear operators from $\mathcal{H}$ into itself, $\mathcal{O}(\mathcal{H})$ for
the set of observables, and $\mathcal{S}(\mathcal{H})$ for the set of density operators. Notice
that both $\mathcal{S}(\mathcal{H}) \subseteq \mathcal{O}(\mathcal{H}) \subseteq \mathcal{L}(\mathcal{H})$. We write $1_{\mathcal{H}} \in \mathcal{O}(\mathcal{H})$ for the identity
operator (often simply $1$ when the space $\mathcal{H}$ is understood) and $1_V$ for the
orthogonal projector on a subspace $V < \mathcal{H}$. Given a state $\rho \in \mathcal{S}(\mathcal{H})$ and an
observable $A \in \mathcal{O}(\mathcal{H})$, its expected value is defined as
$$\langle A\rangle_\rho = \text{tr}[A\rho].$$ (3.2)

For a chosen orthonormal basis $(|\psi_i\rangle)_i \subseteq \mathcal{H}$, any operator $A \in \mathcal{L}(\mathcal{H})$ can be
represented as a square complex matrix $(\langle \psi_i|A\psi_j\rangle)_{i,j}$, i.e., in Dirac notation
$$A = \sum_{i,j} |\psi_i\rangle \langle \psi_i|A\psi_j\rangle \langle \psi_j|.$$ (3.3)

In particular, we have the representation
$$\text{tr}[A] = \sum_i \langle \psi_i|A\psi_i\rangle,$$ (3.4)

(which however does not depend on the chosen basis) and the matrix is
Hermitian if and only if $A$ is self-adjoint. By the spectral theorem, one can
choose a basis of eigenvectors of $A$, so that
$$A = \sum_i \lambda_i |\psi_i\rangle \langle \psi_i|,$$ (3.5)

where $\lambda_i \in \sigma(A)$ are the eigenvalues of $A$. In particular $\text{tr}[A]$ is the sum
(with multiplicities) of the eigenvalues of $A$.

When $A = \rho \in \mathcal{S}(\mathcal{H})$ is a density operator, then $\lambda_i = p_i \in [0,1]$ and one
obtains a classical probability density by counting the eigenvalues with their
multiplicity, since $\text{tr}[\rho] = 1$. A state is called 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 delta probabilities at \(x_0 \in \mathcal{X}\), i.e., \(x \mapsto 1_{\{x=x_0\}}\).

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

Example 3.1. The simplest example of a non-trivial quantum system is given by a two-dimensional space \(\mathcal{H} = \mathbb{C}^2\), where one denotes the standard basis as \(|0\rangle = (1, 0), |1\rangle = (0, 1) \in \mathbb{C}^2\). This should be seen as the quantum analogue of a two-point space \(\mathcal{X} = \{0, 1\}\) (or of a single bit sequence), hence \(\mathbb{C}^2\) is also referred as a single-qubit system. To define the quantum analogue of \(n\)-bits strings \(\mathcal{X} = \{0, 1\}^n\), we discuss first the notion of composite quantum systems.

3.2. Composite systems and partial trace. A composite quantum system \(\mathcal{H} \otimes \mathcal{K}\) is defined as the tensor product between two (finite dimensional) systems \(\mathcal{H}, \mathcal{K}\). One often writes \(|\psi, \varphi\rangle = |\psi\rangle \otimes |\varphi\rangle \in \mathcal{H} \otimes \mathcal{K}\) for the elementary tensor products, whose linear span yields the entire space. The composite system \(\mathcal{H} \otimes \mathcal{K}\) is endowed with the following scalar product:

\[
\langle \psi, \varphi|\psi', \varphi'\rangle_{\mathcal{H} \otimes \mathcal{K}} = \langle \psi|\psi'\rangle_{\mathcal{H}} \langle \varphi|\varphi'\rangle_{\mathcal{K}},
\]

defined on elementary tensor products and naturally extended by linearity. Given orthonormal bases \(\{|\psi_i\rangle\}_{i \subseteq \mathcal{H}}, \{|\varphi_j\rangle\}_{j \subseteq \mathcal{K}}\), the elements \(|\psi_i, \varphi_j\rangle\) yield an orthonormal basis of \(\mathcal{H} \otimes \mathcal{K}\), which in particular has dimension \(\dim(\mathcal{H} \otimes \mathcal{K}) = (\dim \mathcal{H})(\dim \mathcal{K})\). Writing \(A \in \mathcal{L}(\mathcal{H} \otimes \mathcal{K})\) in matrix form

\[
A = \sum_{i,j,k,\ell} A_{ij,kl} |\psi_i, \varphi_j\rangle \langle \psi_k, \varphi_\ell|,
\]

(3.7)
one defines the partial trace of $A$ over $\mathcal{H}$ as the operator on $\mathcal{K}$ given by

$$\text{tr}_\mathcal{H}[A] = \sum_{j,\ell} \text{tr}_\mathcal{H}[A]_{j\ell} |\varphi_j\rangle \langle \varphi_\ell|,$$

where

$$\text{tr}_\mathcal{H}[A]_{j\ell} = \sum_i A_{ij,il}.$$  

(3.8)

Similarly, the partial trace of $A$ over $\mathcal{K}$ is the operator on $\mathcal{H}$ given by

$$\text{tr}_\mathcal{K}[A] = \sum_{i,k} \text{tr}_\mathcal{H}[A]_{ik} |\psi_i\rangle \langle \psi_k|,$$

where

$$\text{tr}_\mathcal{K}[A]_{ik} = \sum_j A_{ij,kj}.$$  

(3.10)

It is simple to prove that both $\text{tr}_\mathcal{H}[A]$ and $\text{tr}_\mathcal{K}[A]$ do not depend on the chosen bases. When $A = \rho \in S(\mathcal{H} \otimes \mathcal{K})$ is a state, the states obtained as partial traces $\rho_\mathcal{K} = \text{tr}_\mathcal{H}[\rho]$, $\rho_\mathcal{H} = \text{tr}_\mathcal{K}[\rho]$ play the roles of marginal probabilities and are called reduced density operators.

For composite systems $\bigotimes_{i=1}^n \mathcal{H}_i$ joining $n$ quantum systems $(\mathcal{H}_i)_{i=1}^n$, we define similarly the partial trace over the system $\mathcal{H}_i$, and write

$$\text{tr}_i[A] = \text{tr}_{\mathcal{H}_i}[A] \in \mathcal{L}(\bigotimes_{j \neq i} \mathcal{H}_j).$$

(3.12)

Given $I \subseteq \{1, \ldots, n\}$, and $\rho \in S(\bigotimes_{i=1}^n \mathcal{H}_i)$ we also write

$$\rho_I = \text{tr}_{\{1, \ldots, n\}\setminus I}[\rho]$$

(3.13)

for the reduced density operator on the composite sub-system $\bigotimes_{i \in I} \mathcal{H}_i$. In particular, when $\mathcal{H}_i = \mathbb{C}^2$, one obtains the system $(\mathbb{C}^2)^\otimes n$, the quantum analogue of $n$-bits sequences, which plays a fundamental role in quantum computing. By taking tensor products of the standard basis elements $\{|0\rangle, |1\rangle\}$, one obtains the so-called computational basis

$$\{|x\rangle\}_{x \in \{0,1\}^n} \subseteq (\mathbb{C}^2)^\otimes n.$$  

(3.14)

Usually, one drastically simplifies the notation writing e.g.

$$|000\rangle = |0\rangle \otimes |0\rangle \otimes |0\rangle, \quad |001\rangle = |0\rangle \otimes |0\rangle \otimes |1\rangle,$$

and similarly for all binary sequences. Although there are many bases available, the computational basis has a distinctive role. For example, classical probabilities over the set of binary sequences are always seen in correspondence with diagonal states in the computational basis.

3.3. **Quantum channels.** The partial trace operators $\text{tr}_\mathcal{H}[,\cdot]$, $\text{tr}_\mathcal{K}[,\cdot]$ are linear and positive, i.e. map positive operators (on $\mathcal{H} \otimes \mathcal{K}$) into positive operators (respectively, on $\mathcal{K}$ and $\mathcal{H}$). In fact, they provide a fundamental example of quantum channels, which are the quantum analogues of classical Markov operators, i.e. the operators induced by integration with respect to probability kernels.
A quantum channel from a system \( \mathcal{H} \) into a system \( \mathcal{K} \) can be abstractly defined as linear, completely positive, and trace preserving operator \( \Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K}) \), where the latter means that

\[
\text{tr}[\Phi(A)] = \text{tr}[A] \quad \text{for every } A \in \mathcal{L}(\mathcal{H}).
\]

(3.16)

In particular, \( \Phi \) maps states on \( \mathcal{H} \) into states on the systems \( \mathcal{K} \). We do not enter into the details of the theory, in particular concerning complete positivity, which is a strengthening on the concept of positive operator. We only recall here the fundamental structure result, stating that quantum channels from \( \mathcal{H} \) into \( \mathcal{K} \) are precisely those linear operators \( \Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{K}) \) that can be represented via a suitable (finite) family of Kraus operators \((B_i)_i\), where each \( B_i : \mathcal{H} \to \mathcal{K} \) is linear, in the following way:

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

(3.17)

and moreover

\[
\sum_i B_i^* B_i = 1_{\mathcal{H}}.
\]

(3.18)

Moreover, the number of Kraus operators in the representation is always bounded from above by \((\dim \mathcal{H})(\dim \mathcal{K})\). In particular, both the trace \(\text{tr}[\cdot] : \mathcal{L}(\mathcal{H}) \to \mathbb{C} = \mathcal{L}(\mathbb{C})\) and the partial trace operators enjoy such a Kraus representation. Another example is the conjugation via of a unitary map \( U : \mathcal{H} \to \mathcal{K} \), i.e., \( U^* U = 1_{\mathcal{H}} \), letting \( \Phi(A) = UAU^* \). Using the representation in terms of Kraus operators, one sees immediately that the set of all quantum channels from \( \mathcal{H} \) into \( \mathcal{K} \) define a convex compact subset of all the linear operators from \( \mathcal{L}(\mathcal{H}) \) to \( \mathcal{L}(\mathcal{K}) \). Finally, given a quantum channel \( \Phi \) from \( \mathcal{H} \) into \( \mathcal{K} \), its adjoint (with respect to the Hilbert-Schmidt scalar product both on \( \mathcal{L}(\mathcal{H}) \) and \( \mathcal{L}(\mathcal{K}) \)) is usually denoted by \( \Phi^\dagger \) and represented by the adjoint of the Kraus operators associated to \( \Phi \):

\[
\Phi^\dagger(A) = \sum_i B_i^* A B_i, \quad \text{for every } A \in \mathcal{L}(\mathcal{K}).
\]

(3.19)

Notice that, by (3.18), \( \Phi^\dagger(1_{\mathcal{K}}) = 1_{\mathcal{H}} \), i.e., \( \Phi^\dagger \) is unital.

Not all the linear operators that map states into states are quantum channels, i.e., enjoy a representation in terms of Kraus operators. A natural example is provided by the transpose map. Given \( A \in \mathcal{L}(\mathcal{H}) \), its transpose \( A^T \in \mathcal{L}(\mathcal{H}^*) \) is defined as

\[
\mathcal{H}^* \ni \langle \psi \rangle \mapsto A^T \langle \psi \rangle = \langle \psi | A, \quad \text{for every } \langle \psi \rangle \in \mathcal{H},
\]

(3.20)

where \( \langle \psi | A | \varphi \rangle = \langle \psi | A \varphi \rangle \) for every \( | \varphi \rangle \in \mathcal{H} \). Given any orthonormal basis \( (| \psi_i \rangle)_i \subseteq \mathcal{H} \), the matrix representing \( A^T \) with respect to the dual basis \( (\langle \psi_i |)_i \subseteq \mathcal{H}^* \) is indeed the transpose (but not conjugate) of the matrix representing \( A \). Since the spectrum of \( A \) equals the spectrum of its transpose, we see immediately that the transpose maps \( \mathcal{S}(\mathcal{H}) \) into \( \mathcal{S}(\mathcal{H}^*) \). However, as soon as \( \dim(\mathcal{H}) \geq 2 \) it is possible to prove that it is not a quantum channel.

**Exercise 3.1.** On a single-qubit quantum system \( \mathcal{H} = \mathbb{C}^2 \), for \( p \in [0, 1] \), define

\[
\rho_p = (1 - p) |0\rangle \langle 0| + p |1\rangle \langle 1|,
\]

(3.21)
and
\[ |\psi_p\rangle = \sqrt{1 - p} |0\rangle + \sqrt{p} |1\rangle. \] (3.22)
Prove that \( \rho_p, |\psi_p\rangle \langle \psi_p| \in \mathcal{S}(\mathbb{C}^2) \). Are they equal?

Exercise 3.2. On a two-qubit quantum system \( \mathbb{C}^2 \otimes \mathbb{C}^2 \), consider
\[ |\Phi^+\rangle := \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \] (3.23)
and the associated Bell state \( \pi = |\Phi^+\rangle \langle \Phi^+| \in \mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^2) \) (which is a pure state). Compute the reduced density operators on the two single-qubit subsystems.

4. An overview of Quantum Optimal Transport

The development of quantum computing and quantum information theory has led to the exploration of various mathematical concepts, including quantum analogues of classical distances on probabilities. Given states \( \sigma, \rho \in \mathcal{S}(\mathcal{H}) \) on a (finite dimensional) quantum system \( \mathcal{H} \),

1. the quantum analogue of the total variation distance (2.10) is called the trace distance, and defined as
\[ D_{\text{tr}}(\sigma, \rho) = \frac{1}{2} \text{tr}[[\rho - \sigma]], \] (4.1)
where \( [\rho - \sigma] \) is defined via functional calculus, so that \( \text{tr}[[\rho - \sigma]] \) is the sum (with multiplicities) of the moduli of the eigenvalues of \( \rho - \sigma \in \mathcal{O}(\mathcal{H}) \),

2. the quantum analogue of the Hellinger distance (2.11) is related to the quantum fidelity, defined as
\[ F(\sigma, \rho) = \text{tr}\left[ \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right]^2, \] (4.2)
while the actual distance corresponding to the Hellinger distance is the Bures metric \( \sqrt{(1 - F(\sigma, \rho))/2} \),

3. the quantum analogue of the relative entropy (2.12) is the (Umegaki) quantum relative entropy, defined as
\[ S(\rho \parallel \sigma) = \text{tr}[\rho (\log \rho - \log \sigma)], \] (4.3)
whenever the kernel of \( \sigma \) is contained in the kernel of \( \rho \) (otherwise one sets the relative entropy to be \(+\infty\), as in the classical case).

Like their classical counterparts, these distances can be defined on quite general systems and can be computed or approximated with a relatively small effort, taking into account the dimension of the system. Moreover, they are not adapted to any specific geometry in the underlying space. This can be made precise by noticing that they are unitarily invariant, i.e., for every unitary \( U : \mathcal{H} \to \mathcal{H} \),
\[ D_{\text{tr}}(U\sigma U^* ,U\rho U^*) = D_{\text{tr}}(\sigma, \rho), \] (4.4)
and similarly for the quantum fidelity and the relative entropy. More generally, all these quantities are monotone with respect to the action of any quantum channel \( \Phi \) from \( \mathcal{H} \) into \( \mathcal{K} \), i.e.,
\[ D_{\text{tr}}(\Phi(\sigma), \Phi(\rho)) \leq D_{\text{tr}}(\sigma, \rho), \] (4.5)
and similarly for the relative entropy
\[ S(\Phi(\rho)\|\Phi(\sigma)) \leq S(\rho\|\sigma) \] (4.6)
and the quantum fidelity (with a reverse inequality). Such inequalities are the quantum analogues of their classical counterparts.

Several proposals for quantum optimal transport problems and induced distances between quantum states have been put forward in recent years, with various applications. Although these lecture notes are focused on two recent proposals by the authors and their collaborators [22, 20], in the literature, many alternative approaches are available, and the (possible) links between them are still not completely explored. The earliest formulation dates back to Connes and Lott in 1992 [17], who defined the spectral distance in non-commutative geometry. Another early approach was put forth by Zyczkowski and Slomczynski in 1997 [49], who computed the Wasserstein distance between the Husimi (classical) probability distributions associated to states in Bosonic systems. In the setting of free probability, Biane and Voiculescu proposed an analogue of the Wasserstein metric in [7]. Since 2012, Maas and Carlen [13, 15, 14] have devised a distance formulating a quantum analogue of the classical Benamou-Brenier formula, which gives a continuous-time formulation of the optimal transport problem (for more information, see also the lecture notes by E. Carlen in this volume). In 2013, Agredo [1] 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 using quantum couplings, with applications to semi-classical limits [32, 9, 33, 10] (for more information, see also the lecture notes by F. Golse in this volume).

We can roughly classify all these definitions according to the point of view on classical optimal transport they mostly emphasize, i.e., the (primal) Monge-Kantorovich problem, the dual problem, or the Benamou-Brenier formulation, see Table 2.

**Table 2: Classifying quantum optimal transport theories.**

| Monge-Kantorovich               |
|---------------------------------|
| - distance between Husimi functions [49] |
| - transport via couplings [32]   |
| - transport via channels [22]    |

| Dual problem                    |
|---------------------------------|
| - spectral distance [17]        |
| - distance on a basis [1]       |
| - Wasserstein distance of order 1 [20] |

| Benamou-Brenier                |
|---------------------------------|
| - quantum Benamou-Brenier [13]  |
5. Quantum optimal transport via quantum channels

In this section, we focus on the proposal [22], where a notion of optimal transport between quantum states is formulated using quantum channels.

5.1. Quantum transport plans. Following the analogy with the classical theory, we write \(X, Y\) for two finite-dimensional quantum systems (possibly \(X = Y\)) and let \(\sigma \in S(X)\), \(\rho \in S(Y)\) be density operators. Recalling the notion of coupling in the sense of Kantorovich, it seems quite natural to define a quantum coupling as a state \(\pi \in S(X \otimes Y)\) such that

\[\text{tr}_Y[\pi] = \sigma, \quad \text{tr}_X[\pi] = \rho.\]  

(5.1)

If a cost function is replaced by an operator \(C \in O(X \otimes Y)\) and the average cost of a coupling is \(\langle C \rangle_\pi = \text{tr}[C\pi]\), this strategy yields a straightforward notion of quantum optimal transport cost, by optimizing the average cost among all the couplings. Indeed, this is the point of view developed in [32] with applications e.g. to semi-classical limits of interacting systems.

However, a different approach can also be devised. Indeed, in the classical theory, i.e., when \(X\) and \(Y\) are sets, Kantorovich couplings \(\pi(x, y)\) can be interpreted as generalized maps (transport plans) by conditioning, e.g. with respect to the first variable, and defining

\[\pi(y|x) = \frac{\pi(x, y)}{\sigma(x)},\]  

(5.2)

provided that \(\sigma(x) > 0\). This defines a Markov kernel pushing \(\sigma\) into \(\rho\).

The quantum analogues of Markov kernels are quantum channels \(\Phi\), so we begin with the following definition.

**Definition 5.1.** Given \(\sigma \in S(X)\), \(\rho \in S(Y)\) a quantum transport plan \(\Phi\) from \(\sigma\) to \(\rho\) is a quantum channel \(\Phi\) from \(X\) to \(Y\) such that \(\Phi(\sigma) = \rho\).

However, we also need to define a notion of transport cost. The issue is that we apparently need a suitable state in a composite system where both states \(\sigma\) and \(\Phi(\sigma) = \rho\) are encoded. The strategy is to rely first upon the so-called purification of a quantum state to build two “copies” of \(\sigma\) and then act with \(\Phi\) only on one such copy.

Let us prove the following well-known result (valid in fact also for infinite dimensional systems).

**Proposition 5.2** (purification of a state). Given any \(\sigma \in S(H)\), there exists an auxiliary quantum system \(K\) and a pure state \(|\Psi\rangle \langle \Psi|\) \(\in S(H \otimes K)\) such that

\[\text{tr}_K[|\Psi\rangle \langle \Psi|] = \sigma.\]  

(5.3)

**Proof.** Let \(K = H^*\) be the dual of \(H\), and consider the canonical isomorphism

\[|\psi\rangle \otimes \langle \varphi| \mapsto |\psi\rangle \langle \varphi|\]  

(5.4)

between \(H \otimes H^*\) and \(L(H)\). We can thus find \(|\Psi\rangle \in H \otimes K\) corresponding to the operator \(\sqrt{\sigma} \in L(H)\) (defined via spectral calculus). We claim that such \(|\Psi\rangle\) satisfies the thesis. Indeed, by choosing an orthonormal basis \((|\psi_i\rangle)\) \(\subseteq H\) of eigenvectors of \(\sigma\) with corresponding eigenvalues \((p_i)\), we have

\[\sqrt{\sigma} = \sum_i \sqrt{p_i} |\psi_i\rangle \langle \psi_i|,\]  

(5.5)
hence

$$\ket{\Psi} = \sum_i \sqrt{p_i} |\psi_i\rangle \otimes \langle \psi_i|.$$  

(5.6)

Thus,

$$\bra{\Psi} \ket{\Psi} = \sum_{i,j} \sqrt{p_i p_j} (|\psi_i\rangle \otimes \langle \psi_i|)(\langle \psi_j| \otimes |\psi_j\rangle),$$  

(5.7)

and taking the partial trace over $K$ yields

$$\text{tr}_K[\bra{\Psi} \ket{\Psi}] = \sum_i p_i |\psi_i\rangle \langle \psi_i| = \sigma.$$  

(5.8)

□

Of course, neither $\ket{\Psi}$ nor $K$ is by any means unique. We thus refer to the construction in the proof above as the canonical purification of $\rho$, so that $K = H^*$ and one can check easily that

$$\text{tr}_{H^*}[\bra{\Psi} \ket{\Psi}] = \sum_i p_i \langle \psi_i| |\psi_i\rangle \in \mathcal{L}(H^*) = \sigma^T$$  

(5.9)

is the transpose operator of $\sigma$.

**Example 5.3.** On a single-qubit system $H = \mathbb{C}^2$, the state

$$\sigma = \frac{1}{2} \mathbb{1}_{\mathbb{C}^2} = \frac{1}{2} |0\rangle \langle 0| + \frac{1}{2} |1\rangle \langle 1|$$  

(5.10)

admits as a purification the Bell state

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}} |00\rangle + \frac{1}{\sqrt{2}} |11\rangle \in (\mathbb{C}^2)^\otimes 2.$$  

(5.11)

Notice that the above is the canonical purification, up to identifying $\mathbb{C}^2$ with its dual.

Let us also notice that an analogous concept to that of purification cannot exist in classical probability, since pure states correspond to Dirac deltas, hence their marginals will be Dirac deltas too. Indeed, the purifications will yield (in general) so-called entangled states, which is a peculiar quantum property that has no classical counterpart.

Back to the optimal transport problem, given the states $\sigma \in \mathcal{S}(\mathcal{X})$ and $\rho \in \mathcal{S}(\mathcal{Y})$, we associate to any quantum transport plan $\Phi$ from $\sigma$ to $\rho$, the density operator

$$\pi_{\Phi} = (\Phi \otimes \mathbb{1}_{\mathcal{L}(\mathcal{X}^*)}) (\ket{\Psi} \bra{\Psi}) \in \mathcal{S}(\mathcal{Y} \otimes \mathcal{X}^*),$$  

(5.12)

where $\ket{\Psi} \in \mathcal{X} \otimes \mathcal{X}^*$ denotes the canonical purification of $\sigma$. The linear operator $\Phi \otimes \mathbb{1}_{\mathcal{L}(\mathcal{X}^*)}$ acts on elementary tensor products $A \otimes B \in \mathcal{L}(\mathcal{X} \otimes \mathcal{X}^*)$ as follows:

$$\Phi \otimes \mathbb{1}_{\mathcal{L}(\mathcal{X}^*)} (A \otimes B) = \Phi(A) \otimes B$$  

(5.13)

and is then extended by linearity to the whole $\mathcal{L}(\mathcal{X} \otimes \mathcal{X}^*) = \mathcal{L}(\mathcal{X}) \otimes \mathcal{L}(\mathcal{X}^*)$. Using that $\Phi$ admits a representation in terms of Kraus operators, one sees immediately that also $\Phi \otimes \mathbb{1}_{\mathcal{L}(\mathcal{X}^*)}$ admits a similar representation, hence it is also a quantum channel and in particular when applied to $|\Phi\rangle \langle \Phi|$, it yields indeed a quantum state. Since the channel acts only on the system $\mathcal{X}$, it is not difficult to prove that reduced density operators of $\pi_{\Phi}$ are

$$\text{tr}_Y[\pi_{\Phi}] = \sigma^T, \quad \text{tr}_{\mathcal{X}^*}[\pi_{\Phi}] = \Phi(\sigma) = \rho.$$  

(5.14)
From this correspondence, we are thus led to a notion of quantum coupling different than (5.1).

**Definition 5.4.** Given $\sigma \in S(X)$, $\rho \in S(Y)$, a quantum coupling between $\sigma$ and $\rho$ is any state $\pi \in S(Y \otimes X^*)$ such that

$$\text{tr}_Y[\pi] = \sigma^T \quad \text{tr}_{X^*}[\pi] = \rho.$$  \hspace{0.5cm} (5.15)

The set of such quantum couplings is denoted with $C(\sigma, \rho)$.

Some examples:
- As in the classical case, the set of quantum couplings is always non-empty, since it contains the product coupling $\pi = \rho \otimes \sigma^T$, induced by the trivial quantum channel

$$\Phi(A) = \text{tr}[A] \rho,$$  \hspace{0.5cm} (5.16)

which maps any state (in particular $\sigma$) into $\rho$.
- If $\sigma = |\psi\rangle \langle \psi|$ is pure, then the product coupling is the only coupling, hence $C(\sigma, \rho) = \{\rho \otimes \sigma^T\}$.
- If $\rho = \sigma$ (hence $X = Y$) the coupling induced by the identity channel $\Phi(A) = A$, which corresponds to the canonical purification of $\sigma$, belongs to $C(\sigma, \sigma)$.

**Remark 5.5.** Although the definition of transport plans is not symmetric, since we require that $\Phi(\sigma) = \rho$, symmetry is restored at the level of couplings, since $C(\sigma, \rho)$ is in natural correspondence with $C(\rho, \sigma)$, via a swap-transpose map, acting on elementary tensor products as

$$\alpha \otimes \beta^T \mapsto \beta \otimes \alpha^T \quad \forall \alpha \in L(Y), \beta \in L(X),$$  \hspace{0.5cm} (5.17)

and extended by linearity to $L(Y) \otimes L(X^*)$.

The set of quantum couplings is in correspondence with quantum plans, as the following construction shows (see [22] for details). Given $\Pi \in C(\sigma, \rho) \subseteq S(Y \otimes X^*)$, by the spectral theorem, write

$$\Pi = \sum_i p_i |\Psi_i\rangle \langle \Psi_i|$$  \hspace{0.5cm} (5.18)

with $(|\Psi_i\rangle)_i \subseteq Y \otimes X^*$ is an orthonormal basis and $\sum_i p_i = 1$. Using the canonical isomorphism between $Y \otimes X^*$ and the space of linear operators from $X$ to $Y$,

$$|\psi\rangle \otimes \langle \varphi| \mapsto |\psi\rangle \langle \varphi|,$$  \hspace{0.5cm} (5.19)

we have an orthonormal basis $(F_i)_i$ corresponding to $(|\Psi_i\rangle)_i$. The identity (5.18) yields that

$$\sum_i p_i F_i^* F_i = (\text{tr}_Y \Pi)^T = \sigma, \quad \text{and} \quad \sum_i p_i F_i F_i^* = \text{tr}_{X^*} \Pi = \rho.$$  \hspace{0.5cm} (5.20)

We thus define the linear operator $\Phi : L(X) \to L(Y)$,

$$\Phi(A) = \sum_i \sqrt{p_i} F_i \sigma^{-\frac{1}{2}} A \sigma^{-\frac{1}{2}} F_i^* \sqrt{p_i},$$  \hspace{0.5cm} (5.21)
where for simplicity we have assumed that $\sigma$ is invertible (otherwise one needs to use a pseudo-inverse). Defining the Kraus operators $B_i = \sqrt{p_i} F_i \sigma^{-\frac{1}{2}}$, using (5.20) one obtains
\[
\sum_i B_i^* B_i = \sum_i \sigma^{-\frac{1}{2}} F_i^* \sqrt{p_i} \sqrt{p_i} F_i \sigma^{-\frac{1}{2}} = \mathbb{1}_X,
\]
and
\[
\Phi(\sigma) = \sum_i \sqrt{p_i} F_i \sigma^{-\frac{1}{2}} \sigma \sigma^{-\frac{1}{2}} F_i^* \sqrt{p_i} = \sum_i p_i F_i F_i^* = \rho,
\]
hence $\Phi$ is a quantum channel such that $\Phi(\sigma) = \rho$.

5.2. Quantum optimal transport cost. We are now in a position to follow the same path as in [32], with this alternative notion of coupling. Given states $\sigma \in S(X)$, $\rho \in S(Y)$ and a cost operator $C \in \mathcal{O}(Y \otimes X^*)$, we search for the coupling $\pi \in \mathcal{C}(\sigma, \rho)$ which minimizes the average cost $\langle C \rangle_\pi = \text{tr}[C \pi]$. Since the set of couplings is closed and convex and the average cost is linear, an optimal coupling $\pi$ always exists (operationally, it is a semidefinite programming problem).

In the remainder of this section, we focus on the particular case of quadratic cost, i.e. we model the cost after the squared Euclidean distance in $\mathbb{R}^d$. Precisely, we let $X = Y$, we fix a set of $d$ observables $(R_i)_{i=1}^d \subseteq \mathcal{O}(X)$ and define the cost operator
\[
C = \sum_{i=1}^d (R_i \otimes \mathbb{1}_{X^*} - \mathbb{1}_X \otimes R_i^T)^2.
\]
We then write
\[
D(\sigma, \rho)^2 = \min_{\pi \in \mathcal{C}(\sigma, \rho)} \text{tr}[C \pi],
\]
which plays the role of a squared Wasserstein distance of order 2 in this quantum setting, where $(R_i)_{i=1}^d$ are the “directions” that are used to measure the transport cost. Although this is not necessarily a distance, we are going to see that it enjoys several natural properties.

First, let us notice that the structure of the cost operator allows for rewriting $\langle C \rangle_\pi$ for any coupling $\pi$ in terms of the associated transport plan $\Phi$, i.e. such that
\[
\pi = (\Phi \otimes \mathbb{1}_{L(X^*)}) (|\Psi \rangle \langle \Psi|),
\]
where $|\Psi \rangle \in X \otimes X^*$ is the canonical purification of $\sigma$. Writing $\Phi^\dagger$ for the adjoint of the channel $\Phi$, we have
\[
\text{tr}[\pi C] = \text{tr} \left[ |\Psi \rangle \langle \Psi| \left( \Phi^\dagger \otimes \mathbb{1}_{L(X^*)} \right) (C) \right] = \langle \Psi | \left( \Phi^\dagger \otimes \mathbb{1}_{L(X^*)} \right) (C) | \Psi \rangle.
\]
By expanding each square in the the definition of $C$, we have
\[
(R_i \otimes \mathbb{1}_{X^*} - \mathbb{1}_X \otimes R_i^T)^2 = R_i^2 \otimes \mathbb{1}_{X^*} + \mathbb{1}_X \otimes (R_i^2)^T - 2R_i \otimes R_i^T,
\]
so that
\[
(\Phi^\dagger \otimes \mathbb{1}_{L(X^*)}) \left( (R_i \otimes \mathbb{1}_{X^*} - \mathbb{1}_X \otimes R_i^T)^2 \right) = \Phi^\dagger (R_i^2) \otimes \mathbb{1}_{X^*} + \mathbb{1}_X \otimes (R_i^2)^T - 2\Phi^\dagger (R_i) \otimes R_i^T.
\]
Using that \( \Psi \) is the canonical purification of \( \sigma \), we have that
\[
\langle \Psi \Phi^\dagger(R_i^2) \otimes \mathbb{1}_X \mid \Psi \rangle = \text{tr}[\Phi^\dagger(R_i^2)\sigma] = \text{tr}[R_i^2\Phi(\sigma)] = \text{tr}[R_i^2\rho].
\] (5.30)
Similarly,
\[
\langle \Psi \mid \mathbb{1}_X \otimes (R_i^2)^T \mid \Psi \rangle = \text{tr}[(R_i^2)^T\sigma^T] = \text{tr}[R_i^2\sigma].
\] (5.31)
These terms do not depend on the specific transport plan (or the coupling). Of course, the third term instead does depend on \( \Phi \). Recalling that \( \Psi \in X \otimes X^c \) corresponds to \( \sqrt{\sigma} \in \mathcal{L}(X) \) in the construction of canonical purification
Proposition 5.2, we have the identity
\[
\langle \Psi | \Phi^\dagger(R_i) \otimes R_i^T | \Psi \rangle = \text{tr}[\sqrt{\sigma} \Phi^\dagger(R_i) \sqrt{\sigma} R_i].
\] (5.32)
Summing all these contributions, we have the equivalent expression that uses only the transport plan \( \Phi \):
\[
\langle C \rangle_\pi = \sum_{i=1}^d \left( \text{tr}[R_i^2\sigma] + \text{tr}[R_i^2\rho] - 2 \text{tr}[R_i \sqrt{\sigma} \Phi^\dagger(R_i) \sqrt{\sigma}] \right),
\] (5.33)
which shows that minimizing \( \langle C \rangle_\pi \) is equivalent to maximizing the correlation-like quantity
\[
\sum_{i=1}^d \text{tr}[R_i \sqrt{\sigma} \Phi^\dagger(R_i) \sqrt{\sigma}].
\] (5.34)

5.3. Properties. Let us obtain some properties of \( D(\sigma, \rho) \). First, notice that by Remark 5.5 and the structure of the cost \( C \), it follows easily that \( D(\sigma, \rho) = D(\rho, \sigma) \).

Notice that, if \( \sigma = \rho \) and \( \Phi = \mathbb{1}_\mathcal{L}(X) \) is the identity channel, then the average cost becomes
\[
2 \sum_{i=1}^d \left( \text{tr}[R_i^2\sigma] - \text{tr}[R_i \sqrt{\sigma} R_i \sqrt{\sigma}] \right).
\] (5.35)
We can prove that this quantity is indeed \( D^2(\sigma, \sigma) \), i.e. the identity channel is always an optimal plan, as one would expect. Indeed, for general \( \sigma, \rho \in \mathcal{S}(X) \), we establish the inequality
\[
D^2(\sigma, \rho) \geq \sum_{i=1}^d \left( \text{tr}[R_i^2\sigma] - \text{tr}[R_i \sqrt{\sigma} R_i \sqrt{\sigma}] + \text{tr}[R_i^2\rho] - \text{tr}[R_i \sqrt{\rho} R_i \sqrt{\rho}] \right),
\] (5.36)
which for \( \sigma = \rho \) becomes an equality (since in the minimization that defines the left-hand side one can take the identity channel). To prove (5.36), by (5.33) it is sufficient to consider any transport plan \( \Phi \) from \( \sigma \) to \( \rho \) and argue that, for every \( i \),
\[
2 \text{tr}[R_i \sqrt{\sigma} \Phi^\dagger(R_i) \sqrt{\sigma}] \leq \text{tr}[R_i \sqrt{\sigma} R_i \sqrt{\sigma}] + \text{tr}[R_i \sqrt{\rho} R_i \sqrt{\rho}].
\] (5.37)
Using the inequality
\[
2 \text{tr}[AB] \leq \text{tr}[A^2] + \text{tr}[B^2], \quad \text{for } A, B \in \mathcal{O}(X),
\] (5.38)
with \( A = \sigma^{1/4} R_i \sigma^{1/4}, B = \sigma^{1/4} \Phi^\dagger(R_i) \sigma^{1/4} \), the thesis reduces to the validity of
\[
\text{tr}[\Phi^\dagger(R_i) \sqrt{\sigma} \Phi^\dagger(R_i) \sqrt{\sigma}] \leq \text{tr}[R_i \sqrt{\rho} R_i \sqrt{\rho}].
\] (5.39)
Since \( \rho = \Phi(\sigma) \), this is a direct application of the monotonicity version of Lieb’s concavity theorem (see [12, Theorem 1.6] with \( t = 1/2 \), but also E. Carlen’s lecture notes in this volume).

Inequality (5.36) with \( \sigma = \rho \) yields that \( D(\sigma, \sigma)^2 \) can be in fact strictly positive – simple examples can be found, e.g. in the single-qubit case [31]. Furthermore, (5.36) implies that for any two quantum states \( \rho, \sigma \in \mathcal{S}(\mathcal{H}) \) we have

\[
D(\rho, \sigma)^2 \geq \frac{1}{2} D(\rho, \rho)^2 + \frac{1}{2} D(\sigma, \sigma)^2.
\]  

(5.40)

The second interesting property that we mention, is a modified triangle inequality:

\[
D(\sigma, \rho) \leq D(\sigma, \tau) + D(\tau, \rho) \quad \forall \rho, \sigma, \tau \in \mathcal{S}(\mathcal{X}).
\]  

(5.41)

The proof of the triangle inequality for the classical Wasserstein distance uses a gluing procedure between couplings, which however uses conditioning. In terms of transport plans, however, it is clear that starting from any \( \Phi_{\sigma \rightarrow \tau} \) (from \( \sigma \) to \( \tau \)) and \( \Phi_{\tau \rightarrow \rho} \) (from \( \tau \) to \( \rho \)), the composition \( \Phi_{\tau \rightarrow \rho} \circ \Phi_{\sigma \rightarrow \tau} \) yields a plan from \( \sigma \) to \( \rho \), which should provide an upper bound to \( D(\sigma, \rho) \). The actual proof however is more involved, and we refer to [22, Theorem 2] for details, explaining in particular the appearance of the extra term \( D(\tau, \tau) \) which again may be strictly positive.

**Remark 5.6.** In view of (5.40), we conjecture that the quantity

\[
\sqrt{D(\rho, \sigma)^2 - \frac{1}{2} D(\rho, \rho)^2 - \frac{1}{2} D(\sigma, \sigma)^2}
\]  

(5.42)

may indeed be an actual distance (up to some non-degeneracy assumptions on the \( R_i \)'s to ensure that the quantity is null if and only if \( \rho = \sigma \)).

6. **The quantum Wasserstein distance of order 1 for qubits**

In this section, following [20], we discuss a notion of quantum Wasserstein distance of order 1, generalizing the classical one induced by the Hamming distance. To keep exposition simple, we limit ourselves to \( n \)-qubit systems, i.e. \( (\mathbb{C}^2)^\otimes n \), which play a fundamental role in quantum computing and quantum information theory.

6.1. **The Hamming-Wasserstein distance.** Before we describe the quantum Wasserstein distance of order 1, let us briefly review the classical case.

The set \( \mathcal{X} = \{0, 1\}^n \) of binary strings of length \( n \) (also called the discrete \( n \)-cube) can be endowed with the Hamming distance

\[
|x - y|_H = \sum_{i=1}^{n} |x_i - y_i| = \sum_{i=1}^{n} 1_{\{x_i \neq y_i\}},
\]  

(6.1)

for \( x = (x_i)_{i=1}^{n}, y = (y_i)_{i=1}^{n} \in \mathcal{X} \). Such a distance induces by (2.6) the Wasserstein distance of order 1 between probability distributions \( \sigma, \rho \) over strings

\[
W_1(\sigma, \rho) = \min_{\pi \in \mathcal{C}(\sigma, \rho)} \sum_{x, y \in \{0, 1\}^n} |x - y|_H \pi(x, y),
\]  

(6.2)

which roughly speaking measures the optimal number of characters in the string that on average one has to change to transform \( \sigma \) into \( \rho \). For example,
if $\sigma$ is a uniform distribution over the possible $2^n$ binary strings, and $\rho$ is a Dirac delta at the string consisting of all zeros, then the only possible coupling leads to

$$W_1(\sigma, \rho) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \sum_{i=1}^n |x_i| = \frac{n}{2}. \quad (6.3)$$

The dual formulation (2.9) reads in this case

$$W_1(\sigma, \rho) = \max \left\{ \sum_{x \in \{0,1\}^n} f(x)(\rho(x) - \sigma(x)) \right\}, \quad (6.4)$$

where maximization runs among all functions $f : \{0,1\}^n \to \mathbb{R}$ that are 1-Lipschitz with respect to the Hamming distance, i.e.,

$$|f(x) - f(y)| \leq |x - y|_H = \sum_{i=1}^n 1_{\{x_i \neq y_i\}} \text{ for every } x, y \in \{0,1\}^n. \quad (6.5)$$

Arguing either from the primal or the dual formulation, it is not difficult to conclude that, if $\sigma$ and $\rho$ have the same $(n - 1)$-marginal distribution, e.g.

$$\sigma(0, x) + \sigma(1, x) = \rho(0, x) + \rho(1, x) \quad \text{for every } x \in \{0,1\}^{n-1}, \quad (6.6)$$

then

$$W_1(\sigma, \rho) \leq 1. \quad (6.7)$$

This is because one can build a coupling $\pi \in C(\sigma, \rho)$ which keeps unchanged all letters but one. More generally, if for some $1 \leq k \leq n$, $\sigma$ and $\rho$ have the same $(n - k)$-marginal distributions, then

$$W_1(\sigma, \rho) \leq k. \quad (6.8)$$

Finally, the following inequalities hold, for probability distributions $\sigma$, $\rho$ on $\{0,1\}^n$:

$$\|\sigma - \rho\|_{TV} \leq W_1(\sigma, \rho) \leq n\|\sigma - \rho\|_{TV}. \quad (6.9)$$

This can be seen by combining Exercise 2.1 with the trivial inequality

$$1_{\{x \neq y\}} \leq |x - y|_H \leq n1_{\{x \neq y\}}. \quad (6.10)$$

6.2. Construction of the distance. Back to the quantum case, i.e., $\mathcal{X} = (\mathbb{C}^2)^{\otimes n}$, our aim is to define a suitable notion of quantum Wasserstein distance of order 1, recovering the classical distance for classical states (i.e., density operators that are diagonal with respect to the computational basis) and enjoying useful properties, such as (6.9). This eventually leads, by duality, to a notion of quantum Lipschitz observables with interesting features.

The definition of the quantum Wasserstein distance of order 1 between two states $\sigma, \rho \in \mathcal{S}((\mathbb{C}^2)^{\otimes n})$ relies upon the fact that, also in the classical case, it is induced by a norm, and indeed we write $\|\rho - \sigma\|_{W_1}$ in the quantum case to emphasize this fact. We proceed in multiple steps.

(1) First, mimicking the classical case, we postulate that a state $\rho$ is at distance $\leq 1$ from a state $\sigma$ if there exists $i = 1, \ldots, n$ such that (recalling the notation (3.12))

$$\text{tr}_i[\sigma] = \text{tr}_i[\rho], \quad (6.11)$$
i.e. the reduced density operators coincide after discarding the $i$-th qubit of the system.

(2) Then, we define the unit ball $\mathcal{B}_n$ (centred at 0) as the convex envelope of differences of states at distance less than one, i.e.,

$$
\mathcal{B}_n = \left\{ \sum_{i=1}^{n} p_i \left( \rho^{(i)} - \sigma^{(i)} \right) : p_i \geq 0, \; \sum_{i=1}^{n} p_i = 1, \; \rho^{(i)}, \sigma^{(i)} \in \mathcal{S}(\mathbb{C}^2 \otimes^n) \right\} \tag{6.12}
$$

and induce a norm via the Minkowski functional:

$$
\|X\|_{W_1} = \min \left( t \geq 0 : X \in t \mathcal{B}_n \right). \tag{6.13}
$$

(3) Finally, we define the distance between states $\sigma$, $\rho$, as $\|\rho - \sigma\|_{W_1}$.

Following this construction we are led to the following quantity, which operationally is a semi-definite programming problem.

**Definition 6.1** (quantum Wasserstein distance of order 1 on $n$ qubits). Let $\sigma, \rho \in \mathcal{S}(\mathbb{C}^2 \otimes^n)$. Then, the quantum Wasserstein distance of order 1 between $\sigma$ and $\rho$ is defined as the quantity

$$
\|\sigma - \rho\|_{W_1} = \min \left\{ \sum_{i=1}^{n} c_i : \sigma - \rho = \sum_{i=1}^{n} c_i (\sigma^{(i)} - \rho^{(i)}), c_i \geq 0, \right\}
$$

$$
\sigma^{(i)}, \rho^{(i)} \in \mathcal{S}(\mathbb{C}^2 \otimes^n) \text{ are such that } \text{tr}_i[\rho^{(i)}] = \text{tr}_i[\sigma^{(i)}] \tag{6.14}
$$

where the minimum runs among all possible $c_i$, $\sigma^{(i)}$ and $\rho^{(i)}$s.

6.3. **Basic properties.** From the very definition, we see that the quantum Wasserstein distance of order 1 is invariant with respect to permutations of the qubits, and unitary operations $U$ acting on a single qubit

$$
\|\sigma - \rho\|_{W_1} = \|U \sigma U^* - U \rho U^*\|_{W_1}. \tag{6.15}
$$

The quantum Wasserstein distance of order 1 can be compared with the trace distance, via the following analogue of (6.9):

$$
D_{\text{tr}}(\sigma, \rho) \leq \|\sigma - \rho\|_{W_1} \leq n D_{\text{tr}}(\sigma, \rho). \tag{6.16}
$$

To see why this holds, consider any representation as in (6.14)

$$
\sigma - \rho = \sum_{i=1}^{n} c_i (\sigma^{(i)} - \rho^{(i)}), \tag{6.17}
$$

take the 1-Schatten norm on both sides, and use the triangle inequality:

$$
\text{tr}[\|\sigma - \rho\|] \leq \sum_{i=1}^{n} c_i \text{ tr}[\|\sigma^{(i)} - \rho^{(i)}\|] \leq 2 \sum_{i=1}^{n} c_i. \tag{6.18}
$$

Minimization upon the representations gives the first inequality in (6.16).

To prove the second inequality, assume first that $D_{\text{tr}}(\sigma, \rho) = 1$, and let

$$
\rho^{(i)} = \rho_{\{1, \ldots, i\}} \otimes \sigma_{\{i+1, \ldots, n\}}, \; \sigma^{(i)} = \rho_{\{1, \ldots, i-1\}} \otimes \sigma_{\{i, \ldots, n\}}. \tag{6.19}
$$
Then,
\[
\rho - \sigma = \sum_{i=1}^{n} \left( \rho^{(i)} - \sigma^{(i)} \right),
\]
(6.20)
hence
\[
\|\rho - \sigma\|_{W_1} \leq n = nD_{tr}(\sigma, \rho).
\]
(6.21)

In the general case, we use the spectral theorem on the operator \(\rho - \sigma\) to write it as
\[
\rho - \sigma = (\rho' - \sigma')D_{tr}(\sigma, \rho),
\]
(6.22)
where \(D_{tr}(\sigma', \rho') = 1\). Since the quantum Wasserstein distance of order 1 is induced by a norm, we have
\[
\|\rho - \sigma\|_{W_1} = D_{tr}(\sigma, \rho)\|\rho' - \sigma'\|_{W_1} \leq nD_{tr}(\sigma, \rho).
\]
(6.23)

The quantum Wasserstein distance of order 1 is well-adapted to the tensor product structure of \((\mathbb{C}^2)^{\otimes n}\) as the next proposition shows (see [20] for a proof).

**Proposition 6.2** (tensorization). Given states \(\sigma, \rho \in \mathcal{S}(\mathbb{C}^2)^{\otimes n}\), it holds, for any \(I \subseteq \{1, \ldots, n\}\),
\[
\|\sigma - \rho\|_{W_1} \geq \|\sigma_I - \rho_I\|_{W_1} + \|\sigma_{\{1, \ldots, n\}\setminus I} - \rho_{\{1, \ldots, n\}\setminus I}\|_{W_1}
\]
(6.24)
with equality if
\[
\sigma = \sigma_I \otimes \sigma_{\{1, \ldots, n\}\setminus I} \quad \text{and} \quad \rho = \rho_I \otimes \rho_{\{1, \ldots, n\}\setminus I}.
\]
(6.25)

For diagonal states with respect to the computational basis, i.e.,
\[
\sigma = \sum_{x \in \{0, 1\}^n} p(x)|x\rangle\langle x|, \quad \rho = \sum_{x \in \{0, 1\}^n} q(x)|x\rangle\langle x|,
\]
(6.26)
the quantum Wasserstein distance of order 1 recovers precisely the Wasserstein distance of order 1 with respect to the Hamming distance, i.e.,
\[
\|\rho - \sigma\|_{W_1} = W_1(p, q).
\]
(6.27)

Indeed, by the tensorization property for product states (Proposition 6.2), we have that for every \(x, y \in \{0, 1\}^n\),
\[
\|\langle x| \rangle - \langle y| \rangle\|_{W_1} = \sum_{i=1}^{n} 1_{\{x_i \neq y_i\}} = |x - y|_H.
\]
(6.28)

Given any classical coupling \(\pi \in \mathcal{C}(p, q)\), between \(p\) and \(q\), the triangle inequality yields
\[
\|\rho - \sigma\|_{W_1} = \left\| \sum_{x,y} \pi(x, y) (|x\rangle\langle x| - |y\rangle\langle y|) \right\|_{W_1} \leq \sum_{x,y} \pi(x, y)|x - y|_H,
\]
(6.29)
yielding that the quantum Wasserstein distance of order 1 is cheaper. To argue that it is not strictly cheaper, consider any representation as in (6.14),
\[
\sigma - \rho = \sum_{i=1}^{n} c_i(\sigma^{(i)} - \rho^{(i)}),
\]
(6.30)
and discard the off-diagonal terms of each $\rho^{(i)}$ and $\sigma^{(i)}$, thus defining classical probability distributions $q^{(i)}$, $p^{(i)}$ over $\{0, 1\}^n$ such that

$$p - q = \sum_{i=1}^n c_i (p^{(i)} - q^{(i)}). \quad (6.31)$$

After discarding the $i$-th bit, the $n-1$ marginals of $p^{(i)}$ and $q^{(i)}$ coincide (because $\operatorname{tr}_i[\sigma^{(i)}] = \operatorname{tr}_i[\rho^{(i)}]$), hence

$$W_1(p^{(i)}, q^{(i)}) \leq 1. \quad (6.32)$$

Since also the classical Wasserstein distance of order 1 is induced by a norm, by triangle inequality we conclude that

$$W_1(p, q) \leq \sum_{i=1}^n c_i W_1(p^{(i)}, q^{(i)}) \leq \sum_{i=1}^n c_i, \quad (6.33)$$

hence equality holds in (6.27).

### 6.4. Distance and channels.

Consider a quantum channel $\Phi = \Phi^{(k)} \otimes 1_{\mathcal{L}(C^2 \otimes (n-k))}$ on a system of $n$ qubits but acting only on a subset of $k$ qubits (without loss of generality, the first $k$ qubits). Then, we argue that

$$\|\rho - \Phi(\rho)\|_{W_1} \leq 2k. \quad (6.34)$$

Indeed, letting $\sigma = \Phi(\rho)$, we have that

$$\operatorname{tr}_{1...k}[^{}\rho\!] = \operatorname{tr}_{1...k}[\sigma]. \quad (6.35)$$

We claim that the condition (6.35) implies

$$\|\rho - \sigma\|_{W_1} \leq 2k. \quad (6.36)$$

Indeed, let us define, for $i = 1, \ldots, k$,

$$\sigma^{\geq i} = 2^{-i} 1_{1...i} \otimes \operatorname{tr}_{1...i}[^{}\sigma\!], \quad \rho^{\geq i} = 2^{-i} 1_{1...i} \otimes \operatorname{tr}_{1...i}[\rho], \quad (6.37)$$

where $1_{1...i}$ denotes the identity on the composite system of the first $i$ qubits. Then,

$$\operatorname{tr}_i[\sigma^{\geq i-1}] \otimes 1_{i}/2 = \sigma^{\geq i}, \quad \operatorname{tr}_i[\rho^{\geq i-1}] \otimes 1_{i}/2 = \rho^{\geq i} \quad (6.38)$$

and for $i = k$, using (6.35),

$$\rho^{\geq k} = \sigma^{\geq k}. \quad (6.39)$$

Summing telescopically, we obtain

$$\rho - \sigma = \sum_{i=1}^k \left((\rho^{\geq i-1} - \rho^{\geq i}) - (\sigma^{\geq i-1} - \sigma^{\geq i})\right)$$

$$= 2 \sum_{i=1}^k \left(\frac{1}{2}(\sigma^{\geq i} + \rho^{\geq i-1}) - \frac{1}{2}(\rho^{\geq i} + \sigma^{\geq i-1})\right). \quad (6.40)$$

Since

$$\operatorname{tr}_i[\sigma^{\geq i} + \rho^{\geq i-1}] = \operatorname{tr}_i[\rho^{\geq i} + \sigma^{\geq i-1}], \quad (6.41)$$

we have

$$\left\|\frac{1}{2}(\sigma^{\geq i} + \rho^{\geq i-1}) - \frac{1}{2}(\rho^{\geq i} + \sigma^{\geq i-1})\right\|_{W_1} \leq 1 \quad (6.42)$$

hence (6.36) by triangle inequality.
It is also important for applications to understand how much the quantum Wasserstein distance of order 1 expands under the action of a quantum channel $\Phi$ acting on systems of qubits (in fact, it does not even need to be the same number of qubits). We introduce the $W_1$-contraction coefficient of $\Phi$ as the quantity

$$\|\Phi\|_{W_1 \rightarrow W_1} = \max_{\rho \neq \sigma} \frac{\|\Phi(\rho) - \Phi(\sigma)\|_{W_1}}{\|\rho - \sigma\|_{W_1}}.$$  

(6.43)

Simple examples show that it is not true that $\|\Phi\|_{W_1 \rightarrow W_1} \leq 1$ in general – indeed, in the classical case the analogous quantity is related to Ollivier’s coarse Ricci curvature [41]. For many families of channels this can be efficiently computed or estimated: see [20] for examples.

6.5. Lipschitz observables. We next search for a dual formulation

$$\|\rho - \sigma\|_{W_1} = \max_A \{\text{tr}[A(\rho - \sigma)] : \|A\|_L \leq 1\},$$  

(6.44)

where maximization is among observables $A \in \mathcal{O}((\mathbb{C}^2)^\otimes n)$ and $\|\cdot\|_L$ denotes a suitable notion of quantum Lipschitz constant of $A$. Recalling that the Wasserstein distance of order 1 is induced by a norm, it follows that $\|\cdot\|_L$ must be the dual norm:

$$\|A\|_L = \max_{\rho, \sigma} \{\text{tr}[A(\rho - \sigma)] : \|\rho - \sigma\|_{W_1} \leq 1\}. \quad (6.45)$$

If we take this as a definition, then duality (6.44) is straightforward. However, this may seem too abstract. Interestingly, we have instead the following “operational” definition of the quantum Lipschitz constant:

$$\|A\|_L = 2 \max_{i=1, \ldots, n} \min_{A^{(i)}} \left\|A - \mathbb{1}_i \otimes A^{(i)}\right\|_{\infty}, \quad (6.46)$$

where $A^{(i)}$ runs among the observables that do not act on the $i$-th qubit system, and $\|\cdot\|_{\infty}$ denotes the operator norm.

To prove the equivalence between (6.45) and (6.46), write temporarily

$$\|A\|_{L} = 2 \max_{i=1, \ldots, n} \min_{A^{(i)}} \left\|A - \mathbb{1}_i \otimes A^{(i)}\right\|_{\infty}. \quad (6.47)$$

We prove first that $\|A\|_L \leq \|A\|'_L$. Given states $\rho^{(i)}$, $\sigma^{(i)}$ such that $\text{tr}_i[\rho^{(i)}] = \text{tr}_i[\sigma^{(i)}]$, then

$$\text{tr}[\mathbb{1}_i \otimes A^{(i)}(\rho^{(i)} - \sigma^{(i)})] = \text{tr}[A^{(i)} \text{tr}_i[\rho^{(i)} - \sigma^{(i)}]] = 0. \quad (6.48)$$

It follows that, given states $\sigma$, $\rho$ and any representation as in (6.14), i.e.,

$$\rho - \sigma = \sum_{i=1}^n c_i (\rho^{(i)} - \sigma^{(i)}), \quad (6.49)$$

we have the identity

$$\text{tr}[A(\rho - \sigma)] = \sum_{i=1}^n c_i \text{tr}[A(\rho^{(i)} - \sigma^{(i)})] = \sum_{i=1}^n c_i \text{tr}[(A - \mathbb{1}_i \otimes A^{(i)})(\rho^{(i)} - \sigma^{(i)})]. \quad (6.50)$$
Using the inequality \( \text{tr}[BC] \leq \|B\|_\infty \|C\|_1 \), we bound from above

\[
\text{tr}[A(\rho - \sigma)] \leq \sum_{i=1}^{n} c_i \|A\|'_L. \tag{6.51}
\]

Assuming \( \|\rho - \sigma\|_{W_1} \leq 1 \), we obtain the inequality \( \|A\|_L \leq \|A\|'_L \). To prove the converse we use the general fact that, for an observable \( A \in \mathcal{O}(\mathcal{H}) \),

\[
2 \min_{c \in \mathbb{R}} \|A - c\mathbb{1}\|_\infty = \max_{\rho, \sigma \in \mathcal{S}(\mathcal{H})} \text{tr}[A(\rho - \sigma)], \tag{6.52}
\]

whose proof (via the spectral theorem) is left as an exercise to the reader.

For \( i = 1, \ldots, n \) we write

\[
\min_{A^{(i)}} \|A - \mathbb{1}_i \otimes A^{(i)}\|_\infty = \min_{A^{(i)}} \min_{c \in \mathbb{R}} \|A - \mathbb{1}_i \otimes A^{(i)} - c\mathbb{1}\|_\infty
\]

\[
= \frac{1}{2} \min_{A^{(i)}} \max_{\rho, \sigma} \text{tr}[(A - \mathbb{1}_i \otimes A^{(i)})(\rho - \sigma)] \tag{6.53}
\]

\[
= \frac{1}{2} \max_{\rho, \sigma} \min_{A^{(i)}} \text{tr}[(A - \mathbb{1}_i \otimes A^{(i)})(\rho - \sigma)],
\]

where in the last step we exchanged minimization and maximization by Fenchel duality theorem. Next, we reduce maximization to \( \rho, \sigma \) such that \( \text{tr}_i[\rho] = \text{tr}_i[\sigma] \), otherwise one can choose a sequence \( A^{(i)}_n \) so that

\[
\text{tr}[\mathbb{1}_i \otimes A^{(i)}_n(\rho - \sigma)] = \text{tr}[A^{(i)}_n \text{tr}_i[\rho - \sigma]] \to -\infty. \tag{6.54}
\]

Therefore, \( \text{tr}[\mathbb{1}_i \otimes A^{(i)}(\rho - \sigma)] = 0 \), and

\[
\min_{A^{(i)}} \|A - \mathbb{1}_i \otimes A^{(i)}\|_\infty \leq \max_{\rho, \sigma} \text{tr}[A(\rho - \sigma)] \leq \|A\|_L \max_{\rho, \sigma} \|\rho - \sigma\|_{W_1}. \tag{6.55}
\]

Finally, we use that \( \|\rho - \sigma\|_{W_1} \leq 1 \), since \( \text{tr}_i[\rho] = \text{tr}_i[\sigma] \).

**Remark 6.3.** It always holds \( \|A\|_L \leq 2 \min_{c \in \mathbb{R}} \|A - c\mathbb{1}\|_\infty \leq 2\|A\|_\infty \). Moreover, if

\[
A = \sum_{I \subseteq \{1, \ldots, n\}} A_I \tag{6.56}
\]

is a sum of local operators, i.e., \( A_I \) acts only on qubits in the subset \( I \), then

\[
\|A\|_L \leq 2 \max_{i=1, \ldots, n} \left\| \sum_{I \ni i} A_I \right\|_{\infty}, \tag{6.57}
\]

simply by taking for \( i = 1, \ldots, n \),

\[
\mathbb{1}_i \otimes A^{(i)} = \sum_{I \notni i} A_I. \tag{6.58}
\]

### 6.6. Gaussian concentration inequalities

It is not difficult to argue that, for a Lipschitz observable \( A \), all the eigenvalues \( \lambda \) must belong to the interval

\[
\text{tr}[A]/2^n - n\|A\|_L \leq \lambda \leq \text{tr}[A]/2^n + n\|A\|_L. \tag{6.59}
\]

The above bound however is not very useful, and in fact most eigenvalues belong to a much smaller interval, of length \( \approx \sqrt{n}\|A\|_L \). This is the content of the following result.
Proposition 6.4 (concentration inequality). Given $A \in \mathcal{O}((\mathbb{C}^2)^\otimes n)$, for every $\delta > 0$, it holds
\[
\dim(A \geq (\text{tr}[A]/2^n + \delta \sqrt{n} \|A\|_L/2) \mathbb{1}) \leq 2^n \exp(-\delta^2/2).
\]

(6.60)

Since $\dim((\mathbb{C}^2)^\otimes n) = 2^n$, it yields that the relative distribution of eigenvalues is (roughly) concentrated as a Gaussian law, with mean $\text{tr}[A]/2^n$ and standard deviation $\sqrt{n} \|A\|_L/2$.

A classical tool to establish concentration results are transport-entropy inequalities of the form
\[
W_1(\sigma, \rho) \leq \sqrt{\frac{n}{2} D_{KL}(\rho || \sigma)}.
\]

(6.61)

for probabilities $\sigma, \rho$, and $D_{KL}(\rho || \sigma)$ is the relative entropy (2.12). Relevant cases in the discrete setting, due to K. Marton [38], include the case where $q$ is a product distribution or more generally a Markov chain, under mild assumptions. Following this route, we establish a quantum analogue of (6.61) for product states on qubits. For an extension to non-product states, see [21].

Proposition 6.5 (Quantum Marton’s inequality). For any $\rho, \sigma \in S((\mathbb{C}^2)^\otimes n)$, with
\[
\sigma = \sigma_1 \otimes \ldots \otimes \sigma_n
\]
product state, the following inequality holds:
\[
\|\rho - \sigma\|_W \leq \sqrt{\frac{n}{2} S(\rho || \sigma)}.
\]

(6.62)

To prove (6.62), we argue first in the case $n = 1$, so that the quantum Wasserstein distance of order 1 coincides with the trace distance, and the inequality becomes
\[
D_{tr}(\sigma, \rho) \leq \sqrt{\frac{1}{2} S(\rho || \sigma)}.
\]

(6.63)

This inequality is well-known in the literature as the quantum analogue of the classical Pinsker’s inequality. Its proof is quite straightforward from the classical Pinsker’s inequality: given states $\rho, \sigma$, using spectral calculus on $\rho - \sigma$, we introduce the orthogonal projectors $\Pi_+ = \mathbb{1}_{\{\rho - \sigma \geq 0\}}$, $\Pi_- = \mathbb{1}_{\{\rho - \sigma < 0\}}$ and probabilities on a two-point space $\{-, +\}$
\[
\begin{align*}
    r_+ &= \text{tr}[\Pi_+ \rho], \\
    s_+ &= \text{tr}[\Pi_+ \sigma].
\end{align*}
\]

So that
\[
2D_{tr}(\sigma, \rho) = |r_+ - s_+| + |r_- - s_-| = \|r - s\|_1 \leq \sqrt{2S(r || s)},
\]

(6.64)

having used the classical Pinsker’s inequality. Since $\{\Pi_+, \Pi_-\}$ gives a measurement (often called Helstrom measurement), the monotonicity of the quantum relative entropy (4.6) yields that
\[
S(r || s) \leq S(\rho || \sigma),
\]

hence (6.63). For the general case, write
\[
\rho - \sigma = \sum_{i=1}^n (\rho_{1 \ldots i} \otimes \sigma_{i+1 \ldots n} - \rho_{1 \ldots i-1} \otimes \sigma_{i \ldots n}),
\]

(6.65)
and apply Pinsker’s inequality, for every \( i = 1, \ldots, n \):

\[
\| \rho_{1\ldots i} \otimes \sigma_{i+1\ldots n} - \rho_{1\ldots i-1} \otimes \sigma_{i\ldots n} \|_1 \leq \sqrt{2S(\rho_{1\ldots i} \otimes \sigma_{i+1\ldots n}\|\rho_{1\ldots i-1} \otimes \sigma_{i\ldots n})}.
\]

Summing upon \( i \) and using concavity of the square root,

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

Finally, we conclude using the identity

\[
S(\rho_{1\ldots i} \otimes \sigma_{i+1\ldots n}\|\rho_{1\ldots i-1} \otimes \sigma_{i\ldots n}) = S(\rho_{1\ldots i}\|\rho_{1\ldots i-1} \otimes \sigma_{i})
\]

\[
= -S(\rho_{1\ldots i}) + S(\rho_{1\ldots i-1}) - \text{tr}[\rho_i \ln \sigma_i]
\]

(6.68)

and a telescopic summation.

Starting from (6.62), we now deduce the Gaussian concentration for quantum Lipschitz observables. Let \( A \in O((\mathbb{C}^2)^{\otimes n}) \) and for simplicity assume that \( \text{tr}[A] = 0 \) and \( \|A\|_L \leq 1 \). We prove that

\[
\text{tr}[e^{tA}] \leq 2^n \exp(nt^2/8), \quad \text{for every } t \in \mathbb{R},
\]

(6.69)

so that concentration follows by spectral calculus and Markov inequality. For simplicity, assume that \( t > 0 \) (otherwise argue with \( -A, -t \) instead of \( A \) and \( t \)). We choose in (6.62) the state \( \sigma = \mathbb{1}/2^n \) which is a product state (also called the maximally mixed state). Duality for the quantum Wasserstein distance of order 1 and Marton’s inequality yield, for any state \( \rho \),

\[
t \text{tr}[A \rho] = t \text{tr}[A(\rho - \sigma)] \leq t\|\rho - \sigma\|_{W_1} \leq t \sqrt{\frac{n}{2} S(\rho\|\sigma)} \leq \frac{nt^2}{8} + S(\rho\|\sigma).
\]

(6.70)

We then choose \( \rho = e^{tA}/\text{tr}[e^{tA}] \) (a Gibbs state) so that \( \ln \rho = tA - \ln \text{tr}[e^{tA}] \) and

\[
S(\rho\|\sigma) = \text{tr}[\rho \ln \rho] + \ln 2^n = t \text{tr}[A \rho] - \ln \text{tr}[e^{tA}] + \ln 2^n.
\]

(6.71)

Rearranging the terms we conclude that

\[
\ln \text{tr}[e^{tA}] \leq \frac{nt^2}{8} + \ln 2^n,
\]

(6.72)

that is equivalent to (6.60).

### 6.7. Continuity of entropy

As a second application of the quantum Wasserstein distance of order 1, we discuss a modulus of continuity for the quantum entropy. In classical information theory, Shannon’s entropy

\[
S(\rho) = -\sum_{x \in X} \rho(x) \ln \rho(x)
\]

(6.73)

quantifies the amount the information contained in a classical probability distribution \( (\rho(x))_{x \in X} \). Usually the logarithm is in base 2 (entropy is measured in bits of information), but for simplicity we use the natural basis here. The notation \( S(X) = S(\rho) \) is also quite common, where \( X \) is a random variable with values in \( X \) and law \( \rho \). For example, the entropy of a
random variable \( X \) with uniform distribution over \( d \) values is \( \ln d \), hence if \( d = 2^n \), we have that \( S(X) = n \ln 2 \).

For a state \( \rho \in S(\mathcal{H}) \) on a quantum system \( \mathcal{H} \), the quantum analogue of Shannon’s entropy is defined as

\[
S(\rho) = -\text{tr}[\rho \ln \rho],
\]

and was first introduced by von Neumann. Interestingly, this definition was historically prior to Shannon’s work – but this should not be surprising since entropy had already been considered in thermodynamics and statistical physics. As its classical counterpart, the quantum entropy plays a fundamental role in information theory.

Clearly, the entropy is continuous as a function of its argument (as long as we consider finite sets or finite-dimensional quantum systems), however a precise modulus of continuity would be quite useful in applications. Polyan- 

ski and Wu \([44]\) proved the following explicit bound, for probability distributions \( \sigma, \rho \) on the discrete cube \( X = \{0, 1\}^n \):

\[
|S(\sigma) - S(\rho)| \leq n h_2 \left( \frac{W_1(\sigma, \rho)}{n} \right),
\]

where \( h_2(x) = -(1-x) \ln(1-x) - x \ln x \) the so-called binary entropy function, i.e., Shannon’s entropy of a Bernoulli distribution with parameter \( x \in [0, 1] \). To grasp the relevance of this result, assume that \( \sigma \) is uniform over the \( 2^n \) values, so that \( S(\sigma) \approx n \) and \( \rho \) is close to \( \sigma \), i.e., \( W_1(\sigma, \rho) \) is much smaller than \( n \). Then, \( h_2(W_1(\sigma, \rho)/n) \) is also small, hence \( S(\rho) \) is also of order \( n \).

In the quantum setting, given states \( \rho, \sigma \in S(\mathcal{H}) \) on a quantum system \( \mathcal{H} \), Fannes \([26]\) and Audenaert \([5]\) proved that

\[
|S(\rho) - S(\sigma)| \leq h_2 \left( D_\text{tr}(\sigma, \rho) \right) + D_\text{tr}(\sigma, \rho) \ln(\dim(\mathcal{H}) - 1). \tag{6.76}
\]

When compared to (6.75), we see that if \( \dim(\mathcal{H}) = 2^n \) as in the case of \( n \)-qubit systems, the inequality becomes much less effective, because of the second term in the right hand side grows linearly as \( n \) grows. Moreover, recalling (6.16), it would be better to replace the trace distance with the quantum Wasserstein distance of order 1, divided by \( n \).

This is precisely what was obtained in [20] and later improved in [23]: for \( \rho, \sigma \in S((\mathbb{C}^2)^\otimes n) \), it holds

\[
|S(\rho) - S(\sigma)| \leq n h_2 \left( \frac{\|\rho - \sigma\|_{W_1}}{n} \right) + \|\rho - \sigma\|_{W_1} \ln(3). \tag{6.77}
\]

Dividing both sides by \( n \), one obtains an inequality that can be even extended to the \( n = \infty \) case, as investigated in [23].

7. Conclusion

We presented two recent approaches to the theory of optimal transport for quantum systems. As described in Section 4, these are not the only ones, and several alternatives have been explored, and possibly other ones will be introduced, since this research field has become quite active, see [47, 27, 8, 24, 25] for even more recent and interesting variants. This variety is indeed a resource for the field, since a particular distance may be better suited for the application one has in mind, from quantum computing, machine
learning and quantum state tomography [16, 45, 42, 37, 19, 35], to the study of dynamics and convergence to equilibria of systems [11, 30, 29]. From a purely mathematical perspective, we are sure that future investigations on the links between these formulations will provide new insights on the geometrical properties of quantum systems.

References

[1] J. Agredo. “A Wasserstein-type distance to measure deviation from equilibrium of quantum Markov semigroups”. In: Open Systems & Information Dynamics 20.02 (2013), p. 1350009 (cit. on p. 11).
[2] R. Alicki and M. Fannes. Quantum dynamical systems. Oxford University Press, Oxford, 2001, pp. xiv+278 (cit. on p. 6).
[3] L. Ambrosio, N. Gigli, and G. Savaré. Gradient flows: in metric spaces and in the space of probability measures. Springer Science & Business Media, 2005 (cit. on p. 2).
[4] M. Arjovsky, S. Chintala, and L. Bottou. “Wasserstein generative adversarial networks”. In: International conference on machine learning. PMLR. 2017, pp. 214–223 (cit. on p. 5).
[5] K. Audenaert. “A sharp continuity estimate for the von Neumann entropy”. In: Journal of Physics A: Mathematical and Theoretical 40.28 (2007), p. 8127 (cit. on p. 26).
[6] I. Bengtsson and K. Życzkowski. Geometry of quantum states: an introduction to quantum entanglement. Cambridge university press, 2017 (cit. on p. 6).
[7] P. Biane and D. Voiculescu. “A free probability analogue of the Wasserstein metric on the trace-state space”. In: Geometric & Functional Analysis GAFA 11.6 (2001), pp. 1125–1138 (cit. on p. 11).
[8] R. Bistron, M. Eckstein, and K. Życzkowski. “Monotonicity of a quantum 2-Wasserstein distance”. In: Journal of Physics A: Mathematical and Theoretical 56.9 (2023), p. 095301 (cit. on p. 26).
[9] E. Caglioti, F. Golse, and T. Paul. “Quantum optimal transport is cheaper”. In: Journal of Statistical Physics 181 (2020), pp. 149–162 (cit. on p. 11).
[10] E. Caglioti, F. Golse, and T. Paul. “Towards Optimal Transport for Quantum Densities”. In: Annali della Scuola Normale Superiore di Pisa, Classe di Scienze (2022) (cit. on p. 11).
[11] A. Capel, C. Rouzé, and D. Franca. “The modified logarithmic Sobolev inequality for quantum spin systems: classical and commuting nearest neighbour interactions”. In: arXiv preprint arXiv:2009.11817 (2020) (cit. on p. 27).
[12] E. Carlen. “On some convexity and monotonicity inequalities of Elliott Lieb”. In: The physics and mathematics of elliott lieb. EMS Press. 2022, pp. 143–209 (cit. on p. 17).
[13] E. Carlen and J. 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”. In: Communications in mathematical physics 331.3 (2014), pp. 887–926 (cit. on p. 11).
[14] E. Carlen and J. Maas. “Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems”. In: Journal of Statistical Physics 178.2 (2020), pp. 319–378 (cit. on p. 11).
[15] E. A. Carlen and J. Maas. “Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance”. In: Journal of Functional Analysis 273.5 (2017), pp. 1810–1869 (cit. on p. 11).
[16] S. Chakrabarti, H. Yiming, T. Li, S. Feizi, and X. Wu. “Quantum wasserstein generative adversarial networks”. In: Advances in Neural Information Processing Systems 32 (2019) (cit. on p. 27).
[17] A. Connes and J. Lott. “The metric aspect of noncommutative geometry”. In: New symmetry principles in quantum field theory. Springer, 1992, pp. 53–93 (cit. on p. 11).
[18] M. Cuturi. “Sinkhorn distances: Lightspeed computation of optimal transport”. In: Advances in neural information processing systems 26 (2013) (cit. on p. 5).
[19] G. De Palma, M. Marvian, C. Rouzé, and D. Franca. “Limitations of variational quantum algorithms: a quantum optimal transport approach”. In: PRX Quantum 4.1 (2023), p. 010309 (cit. on p. 27).
[20] G. De Palma, M. Marvian, D. Trevisan, and S. Lloyd. “The quantum Wasserstein distance of order 1”. In: IEEE Transactions on Information Theory 67.10 (2021), pp. 6627–6643 (cit. on pp. 1, 11, 17, 20, 22, 26).
[21] G. De Palma and C. Rouzé. “Quantum concentration inequalities”. In: Annales Henri Poincaré 23.9 (2022), pp. 3391–3429 (cit. on p. 24).
[22] G. De Palma and D. Trevisan. “Quantum optimal transport with quantum channels”. In: Annales Henri Poincaré 22.10 (2021), pp. 3199–3234 (cit. on pp. 1, 11, 12, 14, 17).
[23] G. De Palma and D. Trevisan. “The Wasserstein Distance of Order 1 for Quantum Spin Systems on Infinite Lattices”. In: Annales Henri Poincaré. Springer. 2023, pp. 1–46 (cit. on p. 26).
[24] R. Duvenhage. “Quadratic Wasserstein metrics for von Neumann algebras via transport plans”. In: Journal of Operator Theory 88.2 (2022), pp. 289–308 (cit. on p. 26).
[25] R. Duvenhage and M. Mapaya. “Quantum Wasserstein distance of order 1 between channels”. In: Infinite Dimensional Analysis, Quantum Probability and Related Topics (2023) (cit. on p. 26).
[26] M. Fannes. “A continuity property of the entropy density for spin lattice systems”. In: Communications in Mathematical Physics 31 (1973), pp. 291–294 (cit. on p. 26).
[27] D. Feliciangeli, A. Gerolin, and L. Portinale. “A non-commutative entropic optimal transport approach to quantum composite systems at positive temperature”. In: Journal of Functional Analysis 285.4 (2023), p. 109963 (cit. on p. 26).
[28] A. Figalli and F. Glaudo. An invitation to optimal transport, Wasserstein distances, and gradient flows. EMS Textbooks in Mathematics. EMS Press, Berlin, 2021, pp. vi+136 (cit. on p. 2).
[29] L. Gao and C. Rouzé. “Complete entropic inequalities for quantum Markov chains”. In: *Archive for Rational Mechanics and Analysis* 245.1 (2022), pp. 183–238 (cit. on p. 27).

[30] L. Gao and C. Rouzé. “Ricci curvature of quantum channels on non-commutative transportation metric spaces”. In: *arXiv:2108.10609* (2021) (cit. on p. 27).

[31] G. Gehér, J. Pitrik, T. Titkos, and D. Virosztek. “Quantum Wasserstein isometries on the qubit state space”. In: *Journal of Mathematical Analysis and Applications* 522.2 (2023), p. 126955 (cit. on p. 17).

[32] F. Golse, C. Mouhot, and T. Paul. “On the mean field and classical limits of quantum mechanics”. In: *Communications in Mathematical Physics* 343.1 (2016), pp. 165–205 (cit. on pp. 11, 12, 15).

[33] F. Golse and T. Paul. “Quantum and Semiquantum Pseudometrics and applications”. In: *Journal of Functional Analysis* (2022) (cit. on p. 11).

[34] N. Gozlan and C. Léonard. “Transport inequalities. A survey”. In: *Markov Processes And Related Fields* 16 (2010), pp. 635–736 (cit. on p. 5).

[35] C. Hirche, C. Rouzé, and D. S. França. “Quantum differential privacy: An information theory perspective”. In: *IEEE Transactions on Information Theory* (2023) (cit. on p. 27).

[36] A. S. Holevo. *Quantum systems, channels, information: a mathematical introduction*. Walter de Gruyter GmbH & Co KG, 2019 (cit. on p. 6).

[37] B. Kiani, G. De Palma, M. Marvian, Z.-W. Liu, and S. Lloyd. “Learning quantum data with the quantum earth mover’s distance”. In: *Quantum Science and Technology* 7.4 (2022), p. 045002 (cit. on p. 27).

[38] K. Marton. “Bounding d-distance by informational divergence: a method to prove measure concentration”. In: *The Annals of Probability* 24.2 (1996), pp. 857–866 (cit. on p. 24).

[39] V. Moretti. *Fundamental Mathematical Structures of Quantum Theory*. Springer, 2019 (cit. on p. 6).

[40] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010 (cit. on p. 6).

[41] Y. Ollivier. “Ricci curvature of metric spaces”. In: *Comptes Rendus Mathematique* 345.11 (2007), pp. 643–646 (cit. on p. 22).

[42] E. Onorati, C. Rouzé, D. S. França, and J. D. Watson. “Efficient learning of ground & thermal states within phases of matter”. In: *arXiv:2301.12946* (2023) (cit. on p. 27).

[43] G. Peyré, M. Cuturi, et al. “Computational optimal transport: With applications to data science”. In: *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pp. 355–607 (cit. on pp. 2, 5).

[44] Y. Polyanskiy and Y. Wu. “Wasserstein continuity of entropy and outer bounds for interference channels”. In: *IEEE Transactions on Information Theory* 62.7 (2016), pp. 3992–4002 (cit. on p. 26).

[45] C. Rouzé and D. Franca. “Learning quantum many-body systems from a few copies”. In: *arXiv preprint arXiv:2107.03333* (2021) (cit. on p. 27).
[46] F. Santambrogio. “Optimal transport for applied mathematicians”. In: *Birkäuser, NY* 55.58-63 (2015), p. 94 (cit. on p. 2).

[47] G. Toth and J. Pitrik. “Quantum Wasserstein distance based on an optimization over separable states”. In: *Bulletin of the American Physical Society* (2023) (cit. on p. 26).

[48] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer, 2009 (cit. on p. 2).

[49] K. Zyczkowski and W. Slomczynski. “The Monge distance between quantum states”. In: *Journal of Physics A: Mathematical and General* 31.45 (1998), p. 9095 (cit. on p. 11).

G. DP.: Dipartimento di Matematica, Università di Bologna, 40126, Bologna, Italy
Email address: giacomo.depalma@unibo.it

D.T.: Dipartimento di Matematica, Università di Pisa, 56125 Pisa, Italy
Email address: dario.trevisan@unipi.it