



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Quantum Optimal Transport with Quantum Channels

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Quantum Optimal Transport with Quantum Channels / De Palma, G; Trevisan, D. - In: ANNALES HENRI POINCARÉ. - ISSN 1424-0637. - ELETTRONICO. - 22:10(2021), pp. 3199-3234. [10.1007/s00023-021-01042-3]

*Availability:*

This version is available at: <https://hdl.handle.net/11585/845014> since: 2022-01-12

*Published:*

DOI: <http://doi.org/10.1007/s00023-021-01042-3>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**De Palma, G., & Trevisan, D. (2021). Quantum optimal transport with quantum channels. *Annales Henri Poincare*, 22(10), 3199-3234**

The final published version is available online at: <https://dx.doi.org/10.1007/s00023-021-01042-3>

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# Quantum optimal transport with quantum channels

Giacomo De Palma and Dario Trevisan

**Abstract.** We propose a new generalization to quantum states of the Wasserstein distance, which is a fundamental distance between probability distributions given by the minimization of a transport cost. Our proposal is the first where the transport plans between quantum states are in natural correspondence with quantum channels, such that the transport can be interpreted as a physical operation on the system. Our main result is the proof of a modified triangle inequality for our transport distance. We also prove that the distance between a quantum state and itself is intimately connected with the Wigner-Yanase metric on the manifold of quantum states. We then specialize to quantum Gaussian systems, which provide the mathematical model for the electromagnetic radiation in the quantum regime. We prove that the noiseless quantum Gaussian attenuators and amplifiers are the optimal transport plans between thermal quantum Gaussian states, and that our distance recovers the classical Wasserstein distance in the semiclassical limit.

## Contents

1. Introduction	2
1.1. Our contribution	3
2. Quantum transport plans and quantum couplings	5
2.1. Preliminaries	5
2.2. Quantum transport	6
3. Quantum transport cost and quantum Wasserstein distance	8
4. Connection with the Wigner-Yanase metric	13
5. The Wasserstein distance for quantum Gaussian systems	14
5.1. Introduction to quantum Gaussian systems	14
5.2. The Wasserstein distance	17
5.3. Semiclassical limit	20
5.4. Quantum Stam inequality	23
6. Conclusions and perspectives	24
Acknowledgements	25

Appendix A. Continuity of the quantum Wasserstein distance	25
A.1. Compactness of quantum plans	25
A.2. Energy	28
A.3. Convergence of the quantum Wasserstein distance	30
Appendix B.	32
References	32

## 1. Introduction

Optimal transport theory [1–3] is the study of the optimal transportation of resources and has now become a fundamental part of functional analysis with continuously growing applications. Indeed, optimal transport theory provides novel tools to tackle fundamental problems such as:

- the study partial differential equations, by interpreting many evolution equations as gradient flows with respect to transport-induced metrics [4];
- geometric analysis, with quantitative isoperimetric inequalities [5] and synthetic notions of Ricci curvature bounds [6, 7];
- stochastic analysis in infinite dimensions [8];
- random combinatorial optimization problems [9];
- statistics and machine learning [10].

In its original formulation [11], the optimal transport problem looks for the cheapest way to transport a source mass distribution onto a target one. In mathematical terms, source and target are modelled via probability measures  $\mu, \nu$  in  $\mathbb{R}^n$ , and the cost of transporting a unit of mass from a position  $x$  to a position  $y$  is a given function  $c(y, x)$ , whose most common choice is of the form  $c(y, x) = |y - x|^p$  for some  $p > 0$ . The assignment from the source  $\mu$  to the target  $\nu$  can be modelled as a “transport map”  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that for every open set  $A \subseteq \mathbb{R}^n$

$$\int_A d\nu(x) = \int_{f^{-1}(A)} d\mu(x), \quad (1)$$

where  $f^{-1}(A)$  is the preimage of  $A$ . The overall transportation cost associated to  $f$  is then

$$C(f) = \int_{\mathbb{R}^n} c(f(x), x) d\mu(x), \quad (2)$$

and any minimizer  $f^*$  of such a cost is called an optimal transport map. The existence of a transport map  $f$  satisfying (1) is in general not guaranteed. For example,  $f$  does not exist if  $\mu$  and  $\nu$  have support on finite sets with different number of points. This crucial issue was solved by relaxing the problem [12], introducing the so-called “transport plans” or couplings, *i.e.*, probability measures  $\pi$  on the product  $\mathbb{R}^n \times \mathbb{R}^n$  such that their first and second marginal laws are respectively  $\mu$  and  $\nu$ . We denote with  $\mathcal{C}(\mu, \nu)$  the set of such couplings. Any transport map  $f$  induces the coupling  $\pi(y, x) = \delta_{y=f(x)} \mu(x)$ , but other coupling always exist, *e.g.*, the product measure  $\pi = \nu \otimes \mu$ .

From the disintegration theorem [13, Volume II, Section 10.6], we can associate to each coupling  $\pi$  a stochastic map  $\phi$  that assigns to each  $x$  in the support of  $\mu$  the probability measure  $\phi(\cdot|x)$  on  $\mathbb{R}^n$  such that for any measurable function  $\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \psi(y, x) d\phi(y|x) d\mu(x) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \psi(y, x) d\pi(y, x), \quad (3)$$

*i.e.*,  $d\phi(y|x)$  is the conditional probability distribution of  $y$  given  $x$  induced by  $\pi$ . The stochastic map  $\phi$  sends a probability measure  $\rho$  defined on the support of  $\mu$  to the probability measure on  $\mathbb{R}^n$

$$\phi(\rho) = \int_{\mathbb{R}^n} \phi(\cdot|x) d\rho(x). \quad (4)$$

We notice that from (3),  $\phi(\mu) = \nu$ .

The cost associated to the coupling  $\pi$  is

$$C(\pi) = \int_{\mathbb{R}^n \times \mathbb{R}^n} c(y, x) d\pi(y, x), \quad (5)$$

and the optimal transport cost is given by

$$W(\mu, \nu)^2 = \inf_{\pi \in \mathcal{C}(\mu, \nu)} C(\pi). \quad (6)$$

The optimal transport problem is now relaxed to a linear optimization problem that under mild regularity assumptions always admits a solution. In special cases, one can a posteriori prove that an optimal coupling is in fact a transport map, the most notable case being that of  $c(y, x) = |y-x|^2$ , when  $\mu$  is absolutely continuous [14]. In fact, with the same cost, the square root of the associated optimal transport cost provides a distance on the space of probability measures, commonly denoted  $W(\mu, \nu)$  and called Wasserstein distance, which induces a Riemannian metric on the manifold of probability measures on  $\mathbb{R}^n$  and whose geometric properties play an essential role in many of the applications mentioned above [4–9].

### 1.1. Our contribution

There have been two recent proposals to generalize the Wasserstein distance to the quantum setting. The first proposal by Carlen, Maas, Datta and Rouzé [15–21] is built on the equivalent dynamical definition of the Wasserstein distance provided by Benamou and Brenier [22], which assigns a length to each path of probability measures that connects the source with the target. The key property of this proposal is that the resulting quantum distance is induced by a Riemannian metric on the manifold of quantum states, and the quantum generalization of the heat semigroup is the gradient flow of the von Neumann entropy with respect to this metric. This quantum generalization of the Wasserstein distance has been shown to be intimately linked to both entropy and Fisher information [19], and has led to determine the rate of convergence of the quantum Ornstein-Uhlenbeck semigroup [16, 23]. The second proposal by Golse, Mouhot, Paul and Caglioti [24–29] arose in the context of the study of the semiclassical limit of quantum mechanics and is built on the definition of the Wasserstein distance through couplings (5). This distance

was the key element to prove that the mean-field limit of quantum mechanics is uniform in the semiclassical limit [24], and has been employed as a cost function to train the quantum counterpart of deep generative adversarial networks [30, 31]. In the following, we will refer to this distance as the GMPC distance.

A fundamental property of the classical transport plans is that they are in one-to-one correspondence with stochastic maps, and this correspondence provides the operational interpretation of the Wasserstein distance as the distance associated to a physical operation that is performed on the system of interest. The mathematical model for the physical operations that can be performed on a quantum system are the quantum channels, which are the completely-positive and trace-preserving linear maps on the set of trace-class operators on the Hilbert space of the quantum system, and are the quantum counterpart of the stochastic maps [32–35]. We propose a new quantum generalization of the Wasserstein distance that builds on the GMPC distance and has the key property that the associated set of quantum transport plans is in a one-to-one correspondence with the set of quantum channels. Our proposal is the first that has this property, which allows for the operational interpretation of quantum transport plans as physical operations performed on the quantum system.

Our main result is that our quantum generalization of the Wasserstein distance satisfies a modified triangle inequality (Theorem 2), whose validity for the GMPC distance is not known. Our distance shares with the GMPC distance the peculiar property of being nonzero even for coinciding quantum states, and we prove that our distance between a quantum state and itself is intimately connected to the Wigner-Yanase metric on the set of quantum states [36, 37]. We then focus on quantum Gaussian systems, which provide the mathematical model for the electromagnetic radiation in the quantum regime. Quantum Gaussian systems play a central role in quantum information, since photons traveling through optical fibers provide the main platform for quantum key distribution and one of the most promising platforms for quantum computation [38–40]. We prove that the optimal transport plans between thermal quantum Gaussian states are noiseless quantum Gaussian attenuators or amplifiers, which model the attenuation of electromagnetic signals traveling through optical fibers and their optimal amplification, respectively (Theorem 3). We also show that the distance between generic states is convex with respect to mixing with a beamsplitter (Theorem 4) and subadditive with respect to the addition of classical noise (Theorem 5). Moreover, we prove that our distance between a state of a quantum Gaussian system and itself satisfies a Stam inequality (Theorem 8). Finally, we prove that our distance recovers the classical Wasserstein distance in the semiclassical limit. Specifically, our distance is lower bounded by the Wasserstein distance between the Husimi Q representations of the quantum states (Theorem 7), and if the quantum states are semiclassical, it is also upper bounded by

the Wasserstein distance between their Glauber-Sudarshan P representations (Theorem 6).

The paper is structured as follows. In section 2, we present our definition of quantum transport plans and show that they are in a one-to-one correspondence with quantum channels. In section 3, we define our quantum Wasserstein distance and prove that it satisfies a modified triangle inequality, and in section 4 we show the connection with the Wigner-Yanase metric. In subsection 5.1 we introduce quantum Gaussian systems, in subsection 5.2 we determine the optimal transport plans between thermal quantum Gaussian states, and in subsection 5.3 we study the semiclassical limit of the quantum Wasserstein distance. We conclude in section 6. We prove in Appendix A some results that are needed to deal with Hilbert spaces with infinite dimension and unbounded operators, and Appendix B contains the proof of an auxiliary lemma.

## 2. Quantum transport plans and quantum couplings

### 2.1. Preliminaries

Let  $\mathcal{H}$  be a separable Hilbert space and let  $\mathcal{L}(\mathcal{H})$  be the set of linear operators on  $\mathcal{H}$ . Let  $\mathcal{B}(\mathcal{H})$  be the set of bounded linear operators on  $\mathcal{H}$ , let

$$\mathcal{T}_1(\mathcal{H}) = \left\{ X \in \mathcal{B}(\mathcal{H}) : \mathrm{Tr}_{\mathcal{H}} \sqrt{X^\dagger X} < \infty \right\} \quad (7)$$

be the set of trace-class operators on  $\mathcal{H}$ , and let

$$\mathcal{S}(\mathcal{H}) = \{ \rho \in \mathcal{T}_1(\mathcal{H}) : \rho \geq 0, \mathrm{Tr}_{\mathcal{H}} \rho = 1 \} \quad (8)$$

be the set of the quantum states of  $\mathcal{H}$ , which are the quantum counterpart of the classical probability measures. The quantum counterparts of the stochastic maps are the quantum channels, which are the completely-positive and trace-preserving linear maps on  $\mathcal{T}_1(\mathcal{H})$ . We recall that a linear map  $\Phi$  is positive if it preserves the set of nonnegative operators in  $\mathcal{T}_1(\mathcal{H})$  and completely positive if the linear map  $\Phi \otimes \mathbb{I}_n$  acting on  $\mathcal{T}_1(\mathcal{H} \otimes \mathbb{C}^n)$  is positive for any  $n \in \mathbb{N}$  [35]. Quantum channels preserve the set of quantum states even when they are applied only to a subsystem, and all the linear maps with this property are quantum channels.

Let

$$\mathcal{T}_2(\mathcal{H}) = \left\{ X \in \mathcal{B}(\mathcal{H}) : \|X\|_2^2 = \mathrm{Tr}_{\mathcal{H}} [X^\dagger X] < \infty \right\} \quad (9)$$

be the set of the Hilbert-Schmidt operators on  $\mathcal{H}$ .  $\mathcal{T}_2(\mathcal{H})$  is a Hilbert space if equipped with the hermitian product

$$\langle\langle X \| Y \rangle\rangle = \mathrm{Tr}_{\mathcal{H}} [X^\dagger Y], \quad X, Y \in \mathcal{T}_2(\mathcal{H}). \quad (10)$$

Let  $\mathcal{H}^*$  be the Hilbert space of continuous linear functionals on  $\mathcal{H}$ .

**Definition 1 (transpose).** For any  $X \in \mathcal{L}(\mathcal{H})$ , let  $X^T$  be the linear operator on  $\mathcal{H}^*$  given by

$$X^T \langle \phi | = \langle \phi | X, \quad \langle \phi | \in \mathcal{H}^*. \quad (11)$$

**Proposition 1** ([41,42]).  $\mathcal{H} \otimes \mathcal{H}^*$  is canonically isometric to  $\mathcal{T}_2(\mathcal{H})$  through the linear extension of the map

$$|\psi\rangle \otimes \langle\phi| \mapsto |\psi\rangle\langle\phi|, \quad |\psi\rangle \in \mathcal{H}, \langle\phi| \in \mathcal{H}^*. \quad (12)$$

We will make an extensive use of the following properties:

**Lemma 1** ([41,42]). For any  $X \in \mathcal{T}_2(\mathcal{H})$  and any  $A, B \in \mathcal{L}(\mathcal{H})$  such that  $AXB \in \mathcal{T}_2(\mathcal{H})$  we have

$$(A \otimes B^T \|X\|) = \|AXB\|. \quad (13)$$

**Lemma 2.** For any  $X \in \mathcal{T}_2(\mathcal{H})$  we have

$$\text{Tr}_{\mathcal{H}^*} \|X\| \langle\langle X \| = X X^\dagger, \quad \text{Tr}_{\mathcal{H}} \|X\| \langle\langle X \| = (X^\dagger X)^T. \quad (14)$$

A purification of a quantum state  $\rho \in \mathcal{S}(\mathcal{H})$  is a pure quantum state  $\gamma \in \mathcal{S}(\mathcal{H} \otimes \mathcal{K})$ , where  $\mathcal{K}$  is a separable Hilbert space, such that

$$\text{Tr}_{\mathcal{K}} \gamma = \rho. \quad (15)$$

The Hilbert space  $\mathcal{K}$  can always be chosen to be  $\mathcal{H}^*$ :

**Definition 2 (canonical purification [35]).** The canonical purification of a quantum state  $\rho \in \mathcal{S}(\mathcal{H})$  is  $\|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \| \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*)$ .

## 2.2. Quantum transport

We start defining our notion of quantum transport plan as the quantum counterpart of the classical stochastic map  $\phi$  defined in (3).

**Definition 3 (quantum transport plan).** For any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , the set  $\mathcal{M}(\rho, \sigma)$  of quantum transport plans from  $\rho$  to  $\sigma$  is the set of the quantum channels  $\Phi: \mathcal{T}_1(\text{supp } \rho) \rightarrow \mathcal{T}_1(\mathcal{H})$  such that  $\Phi(\rho) = \sigma$ .

GMPC associate to any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  the set of quantum couplings

$$\mathcal{C}_{\text{GMPC}}(\rho, \sigma) = \{\Pi \in \mathcal{S}(\mathcal{H}_2 \otimes \mathcal{H}_1) : \text{Tr}_{\mathcal{H}_2} \Pi = \rho, \text{Tr}_{\mathcal{H}_1} \Pi = \sigma\}, \quad (16)$$

where  $\mathcal{H}_{1,2}$  are two copies of  $\mathcal{H}$  [24]. With this definition, there is no straightforward way to associate a quantum transport plan to a quantum coupling.

We propose a new definition of quantum coupling that admits a one-to-one correspondence with quantum transport plans. First, we associate to any quantum transport plan  $\Phi \in \mathcal{M}(\rho, \sigma)$  the quantum state of  $\mathcal{H} \otimes \mathcal{H}^*$

$$\Pi_\Phi = (\Phi \otimes \mathbb{I}_{\mathcal{T}_1(\mathcal{H}^*)}) (\|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \|). \quad (17)$$

We get from Lemma 2

$$\text{Tr}_{\mathcal{H}^*} [\|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \|] = \rho, \quad \text{Tr}_{\mathcal{H}} [\|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \|] = \rho^T, \quad (18)$$

therefore

$$\text{Tr}_{\mathcal{H}} \Pi_\Phi = \rho^T, \quad \text{Tr}_{\mathcal{H}^*} \Pi_\Phi = \sigma. \quad (19)$$

**Definition 4 (quantum coupling).** In view of (19), we associate to any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  the set of quantum couplings

$$\mathcal{C}(\rho, \sigma) = \{\Pi \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*) : \text{Tr}_{\mathcal{H}} \Pi = \rho^T, \text{Tr}_{\mathcal{H}^*} \Pi = \sigma\}. \quad (20)$$



*Remark 1.* If the partial transpose on  $\mathcal{H}^*$  of  $\Pi \in \mathcal{C}(\rho, \sigma)$  is positive semi-definite, it belongs to  $\mathcal{C}_{\text{GMPC}}(\rho, \sigma)$ . Moreover, if the partial transpose on  $\mathcal{H}_1$  of  $\Pi' \in \mathcal{C}_{\text{GMPC}}(\rho, \sigma)$  is positive semi-definite, it belongs to  $\mathcal{C}(\rho, \sigma)$ . Therefore, the quantum couplings of  $\mathcal{C}(\rho, \sigma)$  with positive partial transpose are in a one-to-one correspondence with the quantum couplings of  $\mathcal{C}_{\text{GMPC}}(\rho, \sigma)$  with positive partial transpose. On the contrary, the couplings of  $\mathcal{C}(\rho, \sigma)$  with non-positive partial transpose do not have any counterpart in  $\mathcal{C}_{\text{GMPC}}(\rho, \sigma)$ , and vice-versa.

*Remark 2.* A definition of quantum coupling very similar to Definition 4 has been proposed in [43], where a coupling between  $\rho$  and  $\sigma$  is defined as a quantum state of  $\mathcal{H}^{\otimes 2}$  with marginals  $\rho$  and  $\sigma^T$ , where  $\sigma^T$  is the quantum state of  $\mathcal{H}$  whose density matrix is the transpose of the density matrix of  $\sigma$  in a given basis.

*Remark 3.* The quantum coupling associated to the trivial transport plan of  $\rho$  onto itself is the canonical purification of  $\rho$ :

$$\Pi_{\mathbb{I}} = \|\sqrt{\rho}\rangle\rangle \langle\langle\sqrt{\rho} \| . \quad (21)$$

Unless  $\rho$  is a pure state, the partial transpose of this coupling is not positive. Therefore, this coupling does not have any counterpart in  $\mathcal{C}_{\text{GMPC}}(\rho, \rho)$ .

The following Proposition 2 proves that our quantum couplings are in a one-to-one correspondence with the quantum transport plans.

**Proposition 2.** *For any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ , the map  $\Phi \mapsto \Pi_{\Phi}$  defined by (17) is a bijection between  $\mathcal{M}(\rho, \sigma)$  and  $\mathcal{C}(\rho, \sigma)$ .*

*Proof.* Let us prove that the map is surjective. Let  $\Pi \in \mathcal{C}(\rho, \sigma)$ , and let

$$\Pi = \sum_{n=0}^{\infty} p_n \|A_n\rangle\rangle \langle\langle A_n \| \quad (22)$$

be an orthonormal eigendecomposition of  $\Pi$  as a linear operator on the Hilbert space  $\mathcal{T}_2(\mathcal{H})$ , where the series converges in the trace norm. Since  $\Pi \in \mathcal{C}(\rho, \sigma)$ , we get from Lemma 2

$$\sum_{n=0}^{\infty} p_n A_n^\dagger A_n = \rho, \quad \sum_{n=0}^{\infty} p_n A_n A_n^\dagger = \sigma, \quad (23)$$

where both series converge in the trace norm. We define for any  $X \in \mathcal{T}_1(\text{supp } \rho)$

$$\Phi_{\Pi}(X) = \sum_{n=0}^{\infty} p_n A_n \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} A_n^\dagger \in \mathcal{T}_1(\mathcal{H}). \quad (24)$$

The series in (24) converges in the trace norm since for any  $N \in \mathbb{N}$

$$\begin{aligned} \sum_{n=0}^N p_n \|A_n \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} A_n^\dagger\|_1 &\leq \sum_{n=0}^N p_n \text{Tr} \left[ A_n \rho^{-\frac{1}{2}} |X| \rho^{-\frac{1}{2}} A_n^\dagger \right] \\ &= \text{Tr} \left[ |X| \sum_{n=0}^N p_n \rho^{-\frac{1}{2}} A_n^\dagger A_n \rho^{-\frac{1}{2}} \right] \leq \|X\|_1, \end{aligned} \quad (25)$$

where the last inequality follows from (23). The map  $\Phi_\Pi$  defined in (24) is linear and completely positive, and from (22) it is trace preserving and satisfies  $\Phi_\Pi(\rho) = \sigma$ , hence  $\Phi_\Pi \in \mathcal{M}(\rho, \sigma)$ . Moreover, if  $\Pi$  has degenerate spectrum,  $\Phi_\Pi$  does not depend on the choice of the eigenvectors of  $\Pi$  in (22). Indeed, for any  $X \in \mathcal{T}_1(\text{supp } \rho)$  such that  $\left\| \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} \right\|_\infty < \infty$  we have

$$\Phi_\Pi(X) = \text{Tr}_{\mathcal{H}^*} \left[ \left( \mathbb{I}_{\mathcal{H}} \otimes \left( \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} \right)^T \right) \Pi \right]. \quad (26)$$

Since  $\Phi_\Pi$  is completely positive and trace preserving, it is continuous with respect to the trace norm, hence  $\Phi_\Pi(X)$  does not depend on the choice of the eigenvectors for any  $X \in \mathcal{T}_1(\text{supp } \rho)$ . With the help of Lemma 1 we have

$$\begin{aligned} \Pi_{\Phi_\Pi} &= \sum_{n=0}^{\infty} p_n \left( A_n \rho^{-\frac{1}{2}} \otimes \mathbb{I}_{\mathcal{H}^*} \right) \|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \rangle\rangle \left( \rho^{-\frac{1}{2}} A_n^\dagger \otimes \mathbb{I}_{\mathcal{H}^*} \right) \\ &= \sum_{n=0}^{\infty} p_n \|A_n\| \langle\langle A_n \rangle\rangle = \Pi, \end{aligned} \quad (27)$$

hence the map  $\Phi \mapsto \Pi_\Phi$  is surjective.

Let us show that it is also injective. Let  $\Phi, \Phi' \in \mathcal{M}(\rho, \sigma)$  such that  $\Pi_\Phi = \Pi_{\Phi'} = \Pi$ . Recalling the definition (17) of  $\Pi_\Phi$ , for any  $X \in \mathcal{T}_1(\text{supp } \rho)$  such that  $\left\| \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} \right\|_\infty < \infty$  we have

$$\begin{aligned} \Phi_\Pi(X) &= \Phi \left( \text{Tr}_{\mathcal{H}^*} \left[ \left( \mathbb{I}_{\mathcal{H}} \otimes \left( \rho^{-\frac{1}{2}} X \rho^{-\frac{1}{2}} \right)^T \right) \|\sqrt{\rho}\| \langle\langle \sqrt{\rho} \rangle\rangle \right] \right) \\ &= \Phi \left( \text{Tr}_{\mathcal{H}^*} \left[ \left\| X \rho^{-\frac{1}{2}} \right\| \left\| \rho^{\frac{1}{2}} \right\| \right] \right) = \Phi(X), \end{aligned} \quad (28)$$

where the second equality follows from Lemma 1. Since  $\Phi_\Pi$  is continuous with respect to the trace norm, we get  $\Phi_\Pi(X) = \Phi(X)$  for any  $X \in \mathcal{T}_1(\text{supp } \rho)$ , hence  $\Phi_\Pi = \Phi$ . Analogously we get  $\Phi_\Pi = \Phi'$ , hence  $\Phi = \Phi'$ .  $\square$

### 3. Quantum transport cost and quantum Wasserstein distance

Given a set  $\{R_1, \dots, R_N\}$  of self-adjoint operators on  $\mathcal{H}$ , which we call *quadratures*, we propose the following operational definition of the transport cost associated to the transport plan  $\Phi$  applied to the quantum state  $\rho$ . We build  $N$  copies of the quantum state  $\Pi_\Phi$  defined in (17), and for each  $i = 1, \dots, N$  we measure  $R_i$  on the  $\mathcal{H}$  subsystem and  $R_i^T$  on the  $\mathcal{H}^*$  subsystem of the  $i$ -th copy, getting the outcomes  $r_i$  and  $r'_i$ , respectively. We define the transport cost as the expectation value of  $\sum_{i=1}^N (r_i - r'_i)^2$  over the above protocol. This cost has a simple expression in terms of the quantum transport plan:

**Definition 5 (quantum transport cost).** Let  $\Pi$  be a quantum state of  $\mathcal{H} \otimes \mathcal{H}^*$  whose eigenvectors all belong to the domain of each  $R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T$ .

We define the cost associated to  $\Pi$  as

$$C(\Pi) = \sum_{i=1}^N \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ (R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T) \Pi (R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T) \right]. \quad (29)$$

*Remark 4.* Let  $\Pi'$  be a quantum state of  $\mathcal{H}^{\otimes 2}$  whose eigenvectors all belong to the domain of each  $R_i \otimes \mathbb{I} - \mathbb{I} \otimes R_i$ . GMPC define the cost associated to  $\Pi'$  as

$$C_{\text{GMPC}}(\Pi') = \sum_{i=1}^N \text{Tr}_{\mathcal{H}^{\otimes 2}} \left[ (R_i \otimes \mathbb{I} - \mathbb{I} \otimes R_i) \Pi' (R_i \otimes \mathbb{I} - \mathbb{I} \otimes R_i) \right]. \quad (30)$$

Let  $T_{\mathcal{H}^*}$  denote the partial transposition on  $\mathcal{H}^*$ . Let  $\Pi \in \mathcal{C}(\rho, \sigma)$  be such that  $\Pi^{T_{\mathcal{H}^*}} \geq 0$ . Then,  $\Pi^{T_{\mathcal{H}^*}} \in \mathcal{C}_{\text{GMPC}}(\rho, \sigma)$ , and  $C(\Pi) = C_{\text{GMPC}}(\Pi^{T_{\mathcal{H}^*}})$ .

*Remark 5.* We define the cost operator

$$C = \sum_{i=1}^N (R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T)^2. \quad (31)$$

Whenever  $C\Pi \in \mathcal{T}_1(\mathcal{H} \otimes \mathcal{H}^*)$ , the cost of  $\Pi$  can also be expressed as

$$C(\Pi) = \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} [C\Pi]. \quad (32)$$

**Definition 6 (energy).** Let  $A$  be a self-adjoint operator on  $\mathcal{H}$  and let  $\rho \in \mathcal{S}(\mathcal{H})$  be a quantum state with eigendecomposition  $\rho = \sum_{i=0}^{\infty} p_i |\psi_i\rangle\langle\psi_i|$ . We define the energy of  $\rho$  with respect to  $A$  as

$$E_A(\rho) = \sum_{i=0}^{\infty} p_i \|A|\psi_i\rangle\|^2 \in [0, \infty] \quad (33)$$

if every  $|\psi_i\rangle$  belongs to the domain of  $A$ , and  $E_A(\rho) = \infty$  otherwise.

We define the energy of  $\rho$  with respect to the quadratures  $R_1, \dots, R_N$  as

$$E(\rho) = E_{R_1}(\rho) + \dots + E_{R_N}(\rho). \quad (34)$$

For the sake of a simpler notation, we will refer to (34) simply as the energy of  $\rho$ .

*Remark 6.* If  $\rho = |\psi\rangle\langle\psi|$  is pure, then  $E_A(\rho) = \|A|\psi\rangle\|^2$  if  $|\psi\rangle$  belongs to the domain of  $A$ ,  $E_A(\rho) = \infty$  otherwise.

*Remark 7.* If  $A \in \mathcal{B}(\mathcal{H})$ , then  $A\rho A \in \mathcal{T}_1(\mathcal{H})$  and we have

$$\text{Tr}_{\mathcal{H}} [A\rho A] = \sum_{i=0}^{\infty} p_i \|A|\psi_i\rangle\|^2 = E_A(\rho) < \infty. \quad (35)$$

**Lemma 3.** Let  $\rho$  be a quantum state with finite energy. Then,  $R_i \rho R_i \in \mathcal{T}_1(\mathcal{H})$  for any  $i = 1, \dots, N$ .

*Proof.* Let

$$\rho = \sum_{n=0}^{\infty} p_n |\psi_n\rangle\langle\psi_n| \quad (36)$$

be an eigendecomposition of  $\rho$ . We have

$$\mathrm{Tr}_{\mathcal{H}} [R_i \rho R_i] = \sum_{n=0}^{\infty} p_n \|R_i |\psi_n\rangle\|^2 < \infty, \quad (37)$$

and the claim follows.  $\square$

**Proposition 3.** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  have finite energy. Then, any plan  $\Pi \in \mathcal{C}(\rho, \sigma)$  has finite cost.*

*Proof.* See subsection A.3.  $\square$

**Definition 7 (swap transposition).** Let  $\Gamma$  be an operator on  $\mathcal{H} \otimes \mathcal{H}^*$ .  $\Gamma^T$  is an operator on  $\mathcal{H}^* \otimes \mathcal{H}$ . We define  $\Gamma^{ST}$  to be the operator on  $\mathcal{H} \otimes \mathcal{H}^*$  associated to  $\Gamma^T$  through the canonical identification between  $\mathcal{H}^* \otimes \mathcal{H}$  and  $\mathcal{H} \otimes \mathcal{H}^*$ .

The swap transposition provides a canonical identification between  $\mathcal{C}(\rho, \sigma)$  and  $\mathcal{C}(\sigma, \rho)$ :

**Proposition 4.** *For any  $\Pi \in \mathcal{C}(\rho, \sigma)$ , we have  $\Pi^{ST} \in \mathcal{C}(\sigma, \rho)$ , and the two couplings have the same cost.*

As in the classical case, we define the square Wasserstein distance as the minimum transport cost:

**Definition 8 (quantum Wasserstein distance).** We define for any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$

$$D(\rho, \sigma)^2 = \inf_{\Pi \in \mathcal{C}(\rho, \sigma)} C(\Pi). \quad (38)$$

*Remark 8.* From Proposition 4,  $D(\rho, \sigma) = D(\sigma, \rho)$ .

*Remark 9.* Definition 8 is completely analogous to the GMPC definition of distance

$$D_{\mathrm{GMPC}}(\rho, \sigma)^2 = \inf_{\Pi' \in \mathcal{C}_{\mathrm{GMPC}}(\rho, \sigma)} C_{\mathrm{GMPC}}(\Pi'). \quad (39)$$

As in the classical case, the quantum Wasserstein distance is additive with respect to the tensor product:

**Proposition 5 (additivity with respect to tensor product).** *Let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces with quadratures  $R_1^1, \dots, R_{N_1}^1$  and  $R_1^2, \dots, R_{N_2}^2$ , respectively, and let  $R_1^1 \otimes \mathbb{I}_{\mathcal{H}_2}, \dots, R_{N_1}^1 \otimes \mathbb{I}_{\mathcal{H}_2}, \mathbb{I}_{\mathcal{H}_1} \otimes R_1^2, \dots, \mathbb{I}_{\mathcal{H}_1} \otimes R_{N_2}^2$  be the quadrature operators of  $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ . Then, for any  $\rho_1, \sigma_1 \in \mathcal{S}(\mathcal{H}_1)$  and any  $\rho_2, \sigma_2 \in \mathcal{S}(\mathcal{H}_2)$  with finite energy,*

$$D(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)^2 = D(\rho_1, \sigma_1)^2 + D(\rho_2, \sigma_2)^2. \quad (40)$$

*Proof.* We have for any  $\Pi \in \mathcal{C}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$  with finite cost

$$\begin{aligned} C(\Pi) &= \sum_{i=1}^{N_1} E_{R_i^\dagger \otimes \mathbb{I}_{\mathcal{H}_2} \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T \otimes \mathbb{I}_{\mathcal{H}_2^*}}(\Pi) \\ &\quad + \sum_{i=1}^{N_2} E_{\mathbb{I}_{\mathcal{H}_1} \otimes R_i^2 \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes \mathbb{I}_{\mathcal{H}_1^*} \otimes R_i^{2T}}(\Pi) \\ &= C_1(\text{Tr}_{\mathcal{H}_2 \otimes \mathcal{H}_2^*} \Pi) + C_2(\text{Tr}_{\mathcal{H}_1 \otimes \mathcal{H}_1^*} \Pi) \geq D(\rho_1, \sigma_1)^2 + D(\rho_2, \sigma_2)^2, \end{aligned} \quad (41)$$

where we have used that  $\text{Tr}_{\mathcal{H}_2 \otimes \mathcal{H}_2^*} \Pi \in \mathcal{C}(\rho_1, \sigma_1)$  and  $\text{Tr}_{\mathcal{H}_1 \otimes \mathcal{H}_1^*} \Pi \in \mathcal{C}(\rho_2, \sigma_2)$ .

Conversely, for any  $\Pi_1 \in \mathcal{C}(\rho_1, \sigma_1)$  and  $\Pi_2 \in \mathcal{C}(\rho_2, \sigma_2)$ , we have that  $\Pi = \Pi_1 \otimes \Pi_2 \in \mathcal{C}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$ , hence

$$D(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)^2 \leq C(\Pi) = C_1(\Pi_1) + C_2(\Pi_2). \quad (42)$$

Taking the infimum of the right-hand side of (42) over  $\Pi_1 \in \mathcal{C}(\rho_1, \sigma_1)$  and  $\Pi_2 \in \mathcal{C}(\rho_2, \sigma_2)$  we get

$$D(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)^2 \leq D(\rho_1, \sigma_1)^2 + D(\rho_2, \sigma_2)^2, \quad (43)$$

and the claim follows.  $\square$

**Theorem 1.** For any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with finite energy and any  $\Pi \in \mathcal{C}(\rho, \sigma)$ ,

$$C(\Pi) \geq \frac{1}{2} (C(\|\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}\|) + C(\|\sqrt{\sigma}\rangle\rangle\langle\langle\sqrt{\sigma}\|)). \quad (44)$$

*Proof.* We consider the case  $N = 1$ , the extension to generic  $N$  being straightforward. Let

$$\Pi = \sum_{n=0}^{\infty} p_n \|X_n\rangle\rangle\langle\langle X_n\| \quad (45)$$

be an eigendecomposition of  $\Pi$  as a linear operator on the Hilbert space  $\mathcal{T}_2(\mathcal{H})$ . Since  $\rho$  and  $\sigma$  have finite energy, from Proposition 3  $\Pi$  has finite cost. We then have

$$\begin{aligned} C(\Pi) &= \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} [(R \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R^T) \Pi (R \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R^T)] \\ &= \sum_{n=0}^{\infty} p_n \|(R \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R^T) \|X_n\rangle\rangle\|^2 \stackrel{(a)}{=} \sum_{n=0}^{\infty} p_n \|[R, X_n]\|_2^2 \\ &\stackrel{(b)}{\geq} \sum_{n=0}^{\infty} p_n \text{Tr}_{\mathcal{H}} \left[ R (X_n^\dagger X_n + X_n X_n^\dagger) R - \sqrt{X_n^\dagger X_n} R \sqrt{X_n^\dagger X_n} R \right. \\ &\quad \left. - \sqrt{X_n X_n^\dagger} R \sqrt{X_n X_n^\dagger} R \right] \\ &\stackrel{(c)}{=} \text{Tr}_{\mathcal{H}} [R(\rho + \sigma)R] \\ &\quad - \sum_{n=0}^{\infty} p_n \text{Tr}_{\mathcal{H}} \left[ \sqrt{X_n^\dagger X_n} R \sqrt{X_n^\dagger X_n} R - \sqrt{X_n X_n^\dagger} R \sqrt{X_n X_n^\dagger} R \right], \end{aligned} \quad (46)$$

where (a) follows from Lemma 1, (b) follows from Lemma 7 of Appendix B and (c) follows since

$$\rho = (\text{Tr}_{\mathcal{H}} \Pi)^T = \sum_{n=0}^{\infty} p_n X_n^\dagger X_n, \quad \sigma = \text{Tr}_{\mathcal{H}^*} \Pi = \sum_{n=0}^{\infty} p_n X_n X_n^\dagger, \quad (47)$$

where both series converge in the trace norm. Lieb's concavity theorem [44] implies

$$\begin{aligned} \sum_{n=0}^{\infty} p_n \text{Tr}_{\mathcal{H}} \left[ \sqrt{X_n^\dagger X_n} R \sqrt{X_n^\dagger X_n} R \right] &\leq \text{Tr}_{\mathcal{H}} [\sqrt{\rho} R \sqrt{\rho} R], \\ \sum_{n=0}^{\infty} p_n \text{Tr}_{\mathcal{H}} \left[ \sqrt{X_n X_n^\dagger} R \sqrt{X_n X_n^\dagger} R \right] &\leq \text{Tr}_{\mathcal{H}} [\sqrt{\sigma} R \sqrt{\sigma} R], \end{aligned} \quad (48)$$

therefore

$$\begin{aligned} C(\Pi) &\geq \text{Tr}_{\mathcal{H}} [R(\rho + \sigma) R - \sqrt{\rho} R \sqrt{\rho} R - \sqrt{\sigma} R \sqrt{\sigma} R] \\ &= \frac{1}{2} \left( \| [R, \sqrt{\rho}] \|_2^2 + \| [R, \sqrt{\sigma}] \|_2^2 \right) \\ &= \frac{1}{2} \left( C(\|\sqrt{\rho}\|) \langle\langle \sqrt{\rho} \rangle\rangle + C(\|\sqrt{\sigma}\|) \langle\langle \sqrt{\sigma} \rangle\rangle \right), \end{aligned} \quad (49)$$

where we have used Lemma 1 again, and the claim follows.  $\square$

The fundamental consequence of Theorem 1 is that the identity channel is the optimal plan to transport a quantum state on itself.

**Corollary 1 (trivial transport).** *For any  $\rho \in \mathcal{S}(\mathcal{H})$  with finite energy, the optimal plan to transport  $\rho$  onto itself is the identity channel and*

$$D(\rho, \rho)^2 = C(\|\sqrt{\rho}\|) \langle\langle \sqrt{\rho} \rangle\rangle. \quad (50)$$

We can define the composition of quantum transport plans through the composition of the associated quantum channels. The possibility of composing quantum transport plans allows us to prove the following modified triangle inequality for the quantum Wasserstein distance.

**Theorem 2 (modified triangle inequality).** *For any  $\rho_A, \rho_B, \rho_C \in \mathcal{S}(\mathcal{H})$  with finite energy,*

$$D(\rho_A, \rho_C) \leq D(\rho_A, \rho_B) + D(\rho_B, \rho_B) + D(\rho_B, \rho_C). \quad (51)$$

*Proof.* It is sufficient to prove the thesis for quantum states with finite dimensional support and obtain the general case as a limit by Theorem 9 and Remark 11 of Appendix A.

For the sake of a simpler notation, we consider  $\rho_A, \rho_B$  and  $\rho_C$  as operators on the Hilbert spaces  $\mathcal{H}_A, \mathcal{H}_B$  and  $\mathcal{H}_C$ , respectively (each of which is canonically isomorphic to  $\mathcal{H}$ ). Let  $\Phi_{A \rightarrow B} \in \mathcal{M}(\rho_A, \rho_B)$  and  $\Phi_{B \rightarrow C} \in \mathcal{M}(\rho_B, \rho_C)$ , such that

$$\Phi_{A \rightarrow C} = \Phi_{B \rightarrow C} \circ \Phi_{A \rightarrow B} \in \mathcal{M}(\rho_A, \rho_C). \quad (52)$$

Let also

$$\begin{aligned}\Pi_{BA^*} &= \Pi_{\Phi_{A \rightarrow B}} \in \mathcal{C}(\rho_A, \rho_B), \\ \Pi_{CB^*} &= \Pi_{\Phi_{B \rightarrow C}} \in \mathcal{C}(\rho_B, \rho_C), \\ \Pi_{CA^*} &= \Pi_{\Phi_{A \rightarrow C}} \in \mathcal{C}(\rho_A, \rho_C).\end{aligned}\tag{53}$$

We have

$$\begin{aligned}\Pi_{CA^*} &= (\Phi_{B \rightarrow C} \otimes \mathbb{I}_{\mathcal{T}_1(\mathcal{H}_A^*)})(\Pi_{BA^*}) = \text{Tr}_{B^*} \left[ (\mathbb{I}_{\mathcal{H}_C} \otimes \Psi_{BA^*}^{TB}) (\Pi_{CB^*} \otimes \mathbb{I}_{\mathcal{H}_A^*}) \right] \\ &= \langle\langle \sqrt{\rho_B} \parallel (\Psi_{CB^*} \otimes \Psi_{BA^*}) \parallel \sqrt{\rho_B} \rangle\rangle,\end{aligned}\tag{54}$$

where  $TB$  denotes the partial transposition on  $\mathcal{H}_B$ ,  $\langle\langle \sqrt{\rho_B} \rangle\rangle \in \mathcal{H}_B \otimes \mathcal{H}_B^*$  and

$$\begin{aligned}\Psi_{CB^*} &= \left( \mathbb{I}_{\mathcal{H}_C} \otimes \rho_B^{-\frac{T}{2}} \right) \Pi_{CB^*} \left( \mathbb{I}_{\mathcal{H}_C} \otimes \rho_B^{-\frac{T}{2}} \right), \\ \Psi_{BA^*} &= \left( \rho_B^{-\frac{1}{2}} \otimes \mathbb{I}_{\mathcal{H}_A^*} \right) \Pi_{BA^*} \left( \rho_B^{-\frac{1}{2}} \otimes \mathbb{I}_{\mathcal{H}_A^*} \right),\end{aligned}\tag{55}$$

where  $\rho_B^{-1}$  is defined as the Moore–Penrose inverse [45] of  $\rho_B$ , which is identically zero on the orthogonal of the support of  $\rho_B$  and is equal to the inverse of  $\rho_B$  on its support. We notice that  $\rho_B^{-1}$  is bounded since  $\rho_B$  has finite rank.

We consider the case of one quadrature operator  $R$ , *i.e.*,  $N = 1$ , the extension to generic  $N$  being straightforward. Let  $R_A$ ,  $R_B$  and  $R_C$  be the operator  $R$  acting on  $\mathcal{H}_A$ ,  $\mathcal{H}_B$  and  $\mathcal{H}_C$ , respectively. We have from the triangle inequality for the Hilbert norm

$$\begin{aligned}D(\rho_A, \rho_C) &\leq \sqrt{\text{Tr}_{CA^*} \left[ (R_C - R_A^T) \Pi_{CA^*} (R_C - R_A^T) \right]} \\ &= \sqrt{\text{Tr}_{CA^*} \left[ \langle\langle \sqrt{\rho_B} \parallel (R_C - R_A^T) \Psi_{CB^*} \Psi_{BA^*} (R_C - R_A^T) \parallel \sqrt{\rho_B} \rangle\rangle \right]} \\ &= \left\| \sqrt{\Psi_{CB^*} \Psi_{BA^*}} (R_C - R_A^T) \parallel \sqrt{\rho_B} \right\| \\ &\leq \left\| \sqrt{\Psi_{CB^*} \Psi_{BA^*}} (R_C - R_B) \parallel \sqrt{\rho_B} \right\| \\ &\quad + \left\| \sqrt{\Psi_{CB^*} \Psi_{BA^*}} (R_B - R_B^T) \parallel \sqrt{\rho_B} \right\| \\ &\quad + \left\| \sqrt{\Psi_{CB^*} \Psi_{BA^*}} (R_B^T - R_A^T) \parallel \sqrt{\rho_B} \right\| \\ &= \sqrt{\text{Tr}_{CB^*} \left[ (R_C - R_B^T) \Pi_{CB^*} (R_C - R_B^T) \right]} \\ &\quad + \sqrt{\text{Tr}_{BA^*} \left[ (R_B - R_A^T) \Pi_{BA^*} (R_B - R_A^T) \right]} \\ &\quad + \left\| (R_B - R_B^T) \parallel \sqrt{\rho_B} \right\|,\end{aligned}\tag{56}$$

and the claim follows taking the infimum over  $\Pi_{BA^*} \in \mathcal{C}(\rho_A, \rho_B)$  and  $\Pi_{CB^*} \in \mathcal{C}(\rho_B, \rho_C)$ .  $\square$

## 4. Connection with the Wigner-Yanase metric

Let  $\mathcal{H}$  be a Hilbert space with finite dimension. The Wigner-Yanase metric [36, 37] is the Riemannian metric  $g$  on the manifold of quantum states of  $\mathcal{H}$

given by

$$g_\rho(X, Y) = 4 \langle\langle X \parallel (\sqrt{\rho} \otimes \mathbb{I}_{\mathcal{H}^*} + \mathbb{I}_{\mathcal{H}} \otimes \sqrt{\rho}^T) \rangle\rangle^{-2} \langle\langle Y \rangle\rangle \quad (57)$$

for any  $\rho \in \mathcal{S}(\mathcal{H})$  and any tangent vectors  $X$  and  $Y$  at the point  $\rho$ .

The following Proposition 6 connects the quantum Wasserstein square distance between a quantum state and itself with the Wigner-Yanase square norms of the tangent vectors induced by the commutators with the quadratures.

**Proposition 6 (Wasserstein – Wigner-Yanase connection).** *For any  $\rho \in \mathcal{S}(\mathcal{H})$ ,*

$$D(\rho, \rho)^2 = \frac{1}{4} \sum_{i=1}^N g_\rho(i[R_i, \rho], i[R_i, \rho]) . \quad (58)$$

*Proof.* With the help of Lemma 1, we have for any  $i = 1, \dots, N$

$$\begin{aligned} g_\rho(i[R_i, \rho], i[R_i, \rho]) &= -4 \operatorname{Tr}_{\mathcal{H}} \left[ [R_i, \sqrt{\rho}]^2 \right] \\ &= 4 \langle\langle \sqrt{\rho} \parallel (R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T) \rangle\rangle^2 , \end{aligned} \quad (59)$$

and the claim follows.  $\square$

## 5. The Wasserstein distance for quantum Gaussian systems

We now specialize to quantum Gaussian systems, which provide the mathematical model for the electromagnetic radiation in the quantum regime. Here we will just give a brief introduction to the required formalism. For a more comprehensive presentation of quantum Gaussian systems and their applications in quantum information, the reader can consult Refs. [38, 39, 46–49].

### 5.1. Introduction to quantum Gaussian systems

The Hilbert space of a quantum Gaussian system is  $\mathcal{H} = L^2(\mathbb{R}^m)$ , *i.e.*, the Hilbert space of  $m$  harmonic oscillators. Let  $Q_1, \dots, Q_m$  and  $P_1, \dots, P_m$  be the position and momentum operators of the  $m$  modes, which act on a wavefunction  $\psi \in L^2(\mathbb{R}^m)$  as

$$(Q_i \psi)(q) = q_i \psi(q), \quad (P_i \psi)(q) = -i \frac{\partial}{\partial q_i} \psi(q) \quad (60)$$

and satisfy the canonical commutation relations

$$[Q_i, P_j] = i \delta_{ij} \mathbb{I}_{\mathcal{H}}, \quad [Q_i, Q_j] = [P_i, P_j] = 0, \quad i, j = 1, \dots, m. \quad (61)$$

It is useful to define the quadratures

$$R_1 = Q_1, \quad R_2 = P_1, \quad \dots, \quad R_{2m-1} = Q_m, \quad R_{2m} = P_m, \quad (62)$$

which satisfy the commutation relations

$$[R_i, R_j] = i \Delta_{ij} \mathbb{I}_{\mathcal{H}}, \quad i, \dots, j = 1, \dots, 2m, \quad (63)$$



where

$$\Delta = \bigoplus_{k=1}^m \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (64)$$

is the symplectic form. We also define the ladder operators

$$a_i = \frac{Q_i + iP_i}{\sqrt{2}}, \quad i = 1, \dots, m, \quad (65)$$

satisfying the commutation relations

$$[a_i, a_j^\dagger] = \delta_{ij} \mathbb{I}_{\mathcal{H}}, \quad [a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0, \quad i, j = 1, \dots, m. \quad (66)$$

The first moments of a quantum state  $\rho$  are the expectation values of the quadratures

$$r_i = \text{Tr} [R_i \rho], \quad i = 1, \dots, 2m, \quad (67)$$

and its covariance matrix is

$$\sigma_{ij} = \frac{1}{2} \text{Tr} [\{R_i - r_i \mathbb{I}_{\mathcal{H}}, R_j - r_j \mathbb{I}_{\mathcal{H}}\} \rho], \quad i, j = 1, \dots, 2m, \quad (68)$$

where

$$\{X, Y\} = XY + YX \quad (69)$$

is the anticommutator. From the Robertson-Heisenberg uncertainty principle, the covariance matrix of any quantum state satisfies [47]

$$\sigma \geq \pm \frac{i}{2} \Delta. \quad (70)$$

A fundamental class of states of quantum Gaussian systems is the class of quantum Gaussian states. They are the Gibbs thermal states of quadratic Hamiltonians, and they are the easiest states to prepare in the laboratory. For this reason, they play a key role in several quantum information protocols, *e.g.*, in protocols for quantum key distribution, quantum teleportation or for communication of classical information [39, 46, 47].

**Definition 9 (quantum Gaussian state).** A quantum Gaussian state of  $\mathcal{H}$  is a quantum state proportional to the exponential of a quadratic polynomial in the quadratures:

$$\rho \propto \exp \left( -\frac{1}{2} \sum_{i,j=1}^{2m} (R_i - r_i \mathbb{I}_{\mathcal{H}}) h_{ij} (R_j - r_j \mathbb{I}_{\mathcal{H}}) \right), \quad (71)$$

where  $r \in \mathbb{R}^{2m}$  and  $h$  is a strictly positive  $2m \times 2m$  real matrix. A quantum Gaussian state is completely determined by its first moments and its covariance matrix: for any  $r \in \mathbb{R}^{2m}$  and any symmetric  $2m \times 2m$  real matrix  $\sigma$  satisfying (70), there exists a unique quantum Gaussian state with first moments  $r$  and covariance matrix  $\sigma$ .

$\mathcal{H}^*$  is the Hilbert space of the quantum Gaussian system with  $m$  modes and quadratures  $R_1^T, \dots, R_{2m}^T$ , which satisfy the commutation relations

$$[R_i^T, R_j^T] = -i \Delta_{ij} \mathbb{I}_{\mathcal{H}^*}. \quad (72)$$

$\mathcal{H} \otimes \mathcal{H}^*$  is the Hilbert space of a quantum Gaussian system with  $2m$  modes, and a quantum Gaussian state of  $\mathcal{H} \otimes \mathcal{H}^*$  is a quantum state proportional to the exponential of a quadratic polynomial in  $R_i$  and  $R_i^T$ .

A special class of quantum Gaussian states are the thermal quantum Gaussian states, for which both the covariance matrix and the matrix  $h$  in (71) are proportional to the identity. The thermal quantum Gaussian state with zero temperature is the vacuum state  $|0\rangle\langle 0|$ , which is the projector onto the ground state  $|0\rangle$  of the photon-number Hamiltonian

$$H = \frac{1}{2} \sum_{i=1}^{2m} \left( R_i^2 - \frac{\mathbb{I}_{\mathcal{H}}}{2} \right); \quad (73)$$

its covariance matrix is  $\frac{1}{2}I_{2m}$ .

The quantum Gaussian unitary operators are the unitary operators that preserve the set of quantum Gaussian states. The main quantum Gaussian unitary operators are the displacement operators, the beamsplitter and the squeezing. For any  $z \in \mathbb{C}^m$ , the displacement operator

$$D(z) = \exp \left( \sum_{i=1}^m (z_i a_i^\dagger - z_i^* a_i) \right) \quad (74)$$

is the unitary operator that acts on the ladder operators as [50]

$$D(z)^\dagger a_i D(z) = a_i + z_i \mathbb{I}_{\mathcal{H}}. \quad (75)$$

The beamsplitter and the squeezing are the quantum counterparts of the classical linear mixing of random variables, and are the main transformations in quantum optics. Let  $A$  and  $B$  be  $m$ -mode quantum Gaussian systems with Hilbert spaces  $\mathcal{H}_A$  and  $\mathcal{H}_B$  and ladder operators  $a_1 \dots a_m$  and  $b_1 \dots b_m$ , respectively. The *beamsplitter* of transmissivity  $0 \leq \eta \leq 1$  is implemented by the unitary operator

$$U(\eta) = \exp \left( \arccos \sqrt{\eta} \sum_{i=1}^m (a_i^\dagger b_i - b_i^\dagger a_i) \right), \quad (76)$$

and performs a linear rotation of the ladder operators [46, Section 1.4.2]:

$$\begin{aligned} U(\eta)^\dagger a_i U(\eta) &= \sqrt{\eta} a_i + \sqrt{1-\eta} b_i, \\ U(\eta)^\dagger b_i U(\eta) &= -\sqrt{1-\eta} a_i + \sqrt{\eta} b_i, \quad i = 1, \dots, m. \end{aligned} \quad (77)$$

We define for any  $0 \leq \eta \leq 1$  and any  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$

$$\mathcal{B}_\eta(\rho) = \text{Tr}_B \left[ U(\eta) \rho U(\eta)^\dagger \right]. \quad (78)$$

The *squeezing* [50] of parameter  $\kappa \geq 1$  is implemented by the unitary operator

$$U(\kappa) = \exp \left( \text{arccosh} \sqrt{\kappa} \sum_{i=1}^m (a_i^\dagger b_i^\dagger - a_i b_i) \right), \quad (79)$$

and acts on the ladder operators as

$$\begin{aligned} U(\kappa)^\dagger a_i U(\kappa) &= \sqrt{\kappa} a_i + \sqrt{\kappa - 1} b_i^\dagger, \\ U(\kappa)^\dagger b_i U(\kappa) &= \sqrt{\kappa - 1} a_i^\dagger + \sqrt{\kappa} b_i, \quad i = 1, \dots, m. \end{aligned} \quad (80)$$

Quantum Gaussian channels are the quantum channels that preserve the set of quantum Gaussian states, and provide the mathematical model for the attenuation and the noise that affect electromagnetic signals traveling through optical fibers and for their amplification. The most important families of quantum Gaussian channels are the quantum Gaussian attenuators and amplifiers. The *noiseless quantum Gaussian attenuator*  $\mathcal{E}_\eta$  [51, case (C) with  $k = \sqrt{\eta}$  and  $N = 0$ ] models the attenuation affecting electromagnetic signals traveling through optical fibers or free space and can be implemented mixing the input state  $\rho$  with the vacuum state through a beamsplitter of transmissivity  $0 \leq \eta \leq 1$ :

$$\mathcal{E}_\eta(\rho) = \mathcal{B}_\eta(\rho \otimes |0\rangle\langle 0|). \quad (81)$$

The *noiseless quantum Gaussian amplifier*  $\mathcal{A}_\kappa$  [51, case (C) with  $k = \sqrt{\kappa}$  and  $N = 0$ ] models the amplification of electromagnetic signals and can be implemented performing a squeezing of parameter  $\kappa \geq 1$  on the input state  $\rho$  and the vacuum state:

$$\mathcal{A}_\kappa(\rho) = \text{Tr}_{\mathcal{H}_B} \left[ U(\kappa) (\rho \otimes |0\rangle\langle 0|) U(\kappa)^\dagger \right]. \quad (82)$$

## 5.2. The Wasserstein distance

We are now ready to define the Wasserstein distance for quantum Gaussian systems. In analogy to the classical transport cost on  $\mathbb{R}^n$ , we consider the cost associated to the canonical quadratures (62), such that the cost operator is

$$C = \frac{1}{2} \sum_{i=1}^{2m} (R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T)^2. \quad (83)$$

(83) differs from (29) by a factor  $\frac{1}{2}$  to match the classical normalization and the GMPC normalization. Sometimes it will be useful to express the cost operator in terms of the ladder operators:

$$C = \sum_{i=1}^m (a_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes a_i^T)^\dagger (a_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes a_i^T). \quad (84)$$

Contrarily to the cost operator adopted by GMPC, our  $C$  does not have discrete eigenvalues, and its essential spectrum is the whole interval  $[0, \infty)$ . As in the classical case, since  $C$  is a quadratic polynomial in the quadratures, the cost of a quantum coupling is completely determined by its first moments and its covariance matrix. Therefore, as for the transport distance between classical Gaussian probability measures, the transport distance between quantum Gaussian states can be computed considering only Gaussian couplings:

**Proposition 7.** *Let  $\rho$  and  $\sigma$  be quantum Gaussian states of  $\mathcal{H}$ . Then, the infimum over  $\mathcal{C}(\rho, \sigma)$  in the definition of  $D(\rho, \sigma)$  can be restricted to the quantum Gaussian states of  $\mathcal{H} \otimes \mathcal{H}^*$  with marginals  $\sigma$  and  $\rho^T$ , respectively.*

*Proof.* If we replace a generic  $\Pi \in \mathcal{C}(\rho, \sigma)$  with the quantum Gaussian state with the same first moments and covariance matrix, both the marginals and the cost remain the same.  $\square$

The optimization over quantum Gaussian couplings can be performed analytically when  $\rho$  and  $\sigma$  are thermal quantum Gaussian states, and the optimal transport plans are noiseless quantum Gaussian attenuators or amplifiers:

**Theorem 3.** *For any  $\nu \geq \frac{1}{2}$ , let  $\omega(\nu)$  be the thermal quantum Gaussian state with covariance matrix  $\nu I_{2m}$ . Then, for any  $\frac{1}{2} \leq \nu \leq \nu'$ ,*

- *the optimal transport plan from  $\omega(\nu)$  to  $\omega(\nu')$  is the noiseless quantum Gaussian amplifier  $\mathcal{A}_\kappa$  with amplification parameter  $\kappa = \frac{2\nu'+1}{2\nu+1}$ ;*
- *the optimal transport plan from  $\omega(\nu)$  to  $\omega(\nu')$  is the noiseless quantum Gaussian attenuator  $\mathcal{E}_\eta$  with attenuation parameter  $\eta = \frac{2\nu-1}{2\nu'-1}$ ;*
- *in both cases, the transport distance is*

$$D(\omega(\nu), \omega(\nu')) = \sqrt{m} \left( \sqrt{\nu' + \frac{1}{2}} - \sqrt{\nu - \frac{1}{2}} \right). \quad (85)$$

*Proof.* Thanks to Proposition 5, it is sufficient to prove the claim for  $m = 1$ . Thanks to Proposition 7, we can assume that the coupling  $\Pi$  is a quantum Gaussian state. Its covariance matrix  $\sigma$  must have the form

$$\sigma(X) = \begin{pmatrix} \nu' I_2 & X \\ X^T & \nu I_2 \end{pmatrix}, \quad (86)$$

where  $X$  is a  $2 \times 2$  real matrix such that

$$\sigma(X) \geq \pm \frac{i}{2} \begin{pmatrix} \Delta & 0 \\ 0 & -\Delta \end{pmatrix}, \quad (87)$$

where the minus sign on the symplectic matrix of  $\mathcal{H}^*$  is due to the commutation relations (72). The cost associated to the coupling above is

$$\mathrm{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} [C \Pi(X)] = \nu' + \nu - \mathrm{tr} X. \quad (88)$$

For any  $X$  satisfying (87),  $X' = \frac{\mathrm{tr} X}{2} I_2$  still satisfies (87), and the associated coupling has the same cost. Therefore, we can assume that  $X = c I_2$  for some  $c \in \mathbb{R}$ . The condition (87) becomes  $c^2 \leq (\nu - \frac{1}{2})(\nu' + \frac{1}{2})$ , hence the optimal coupling has covariance matrix

$$\sigma^* = \begin{pmatrix} \nu' I_2 & \sqrt{(\nu - \frac{1}{2})(\nu' + \frac{1}{2})} I_2 \\ \sqrt{(\nu - \frac{1}{2})(\nu' + \frac{1}{2})} I_2 & \nu I_2 \end{pmatrix}, \quad (89)$$

and the claim (85) follows.

For any  $\nu \geq \frac{1}{2}$ , let

$$\gamma(\nu) = \left\| \left\langle \sqrt{\omega(\nu)} \right\rangle \right\| \left\| \left\langle \sqrt{\omega(\nu)} \right\rangle \right\|. \quad (90)$$

$\gamma(\nu)$  is the quantum Gaussian state with zero first moments and covariance matrix

$$\sigma(\nu) = \begin{pmatrix} \nu I_2 & \sqrt{(\nu - \frac{1}{2})(\nu + \frac{1}{2})} I_2 \\ \sqrt{(\nu - \frac{1}{2})(\nu + \frac{1}{2})} I_2 & \nu I_2 \end{pmatrix}. \quad (91)$$

From [52, Corollary 1], the quantum Gaussian states  $(\mathcal{A}_\kappa \otimes \mathbb{I}_{\mathcal{H}^*})(\gamma(\nu))$  and  $(\mathbb{I}_{\mathcal{H}} \otimes \mathcal{E}_\eta)(\gamma(\nu'))$  for  $\kappa = \frac{2\nu'+1}{2\nu+1}$  and  $\eta = \frac{2\nu-1}{2\nu'-1}$  both have covariance matrix equal to  $\sigma^*$ . Therefore, they coincide with the optimal coupling, and  $\mathcal{A}_\kappa$  and  $\mathcal{E}_\eta$  are the optimal plans.  $\square$

The following Theorem 4 states that the quantum Wasserstein distance is convex with respect to the mixing with the beamsplitter.

**Theorem 4 (beamsplitter convexity).** *Let  $\rho_0, \rho_1, \sigma_0, \sigma_1 \in \mathcal{S}(\mathcal{H})$  with finite energy such that for any  $i = 1, \dots, m$*

$$\mathrm{Tr}[(\sigma_0 - \rho_0) R_i] = \mathrm{Tr}[(\sigma_1 - \rho_1) R_i] = 0, \quad (92)$$

and for any  $0 \leq \eta \leq 1$  let

$$\rho_\eta = \mathcal{B}_\eta(\rho_1 \otimes \rho_0), \quad \sigma_\eta = \mathcal{B}_\eta(\sigma_1 \otimes \sigma_0), \quad (93)$$

where  $\mathcal{B}_\eta$  is the beamsplitter (78). Then,

$$D(\rho_\eta, \sigma_\eta)^2 \leq \eta D(\rho_1, \sigma_1)^2 + (1 - \eta) D(\rho_0, \sigma_0)^2. \quad (94)$$

*Proof.* For any  $\Pi_0 \in \mathcal{C}(\rho_0, \sigma_0)$  and any  $\Pi_1 \in \mathcal{C}(\rho_1, \sigma_1)$ , let

$$\Pi_\eta = \mathrm{Tr}_{\mathcal{H}_B \otimes \mathcal{H}_B^*} \left[ (U(\eta) \otimes U(\eta)^{T\dagger}) (\Pi_1 \otimes \Pi_0) (U(\eta)^\dagger \otimes U(\eta)^T) \right]. \quad (95)$$

Let  $C_A$  and  $C_B$  be the cost operators of the quantum Gaussian systems  $A$  and  $B$ , respectively. We have

$$\begin{aligned} & (U(\eta)^\dagger \otimes U(\eta)^T) C_A (U(\eta) \otimes U(\eta)^{T\dagger}) \\ &= \eta C_A + (1 - \eta) C_B + \sqrt{\eta(1 - \eta)} \sum_{i=1}^{2m} (R_i^A - R_i^{AT}) \otimes (R_i^B - R_i^{BT}). \end{aligned} \quad (96)$$

We have  $\Pi_\eta \in \mathcal{C}(\rho_\eta, \sigma_\eta)$ , hence

$$\begin{aligned} D(\rho_\eta, \sigma_\eta)^2 &\leq C(\Pi_\eta) = \eta C(\Pi_1) + (1 - \eta) C(\Pi_0) \\ &+ \sqrt{\eta(1 - \eta)} \sum_{i=1}^{2m} \mathrm{Tr}_{\mathcal{H}_A \otimes \mathcal{H}_A^*} [(\sigma_1 - \rho_1) R_i^A] \mathrm{Tr}_{\mathcal{H}_B \otimes \mathcal{H}_B^*} [(\sigma_0 - \rho_0) R_i^B], \end{aligned} \quad (97)$$

and the claim follows.  $\square$

The following Theorem 5 states that the quantum Wasserstein distance is subadditive with respect to the addition of classical noise.

**Theorem 5 (classical noise).** *Let  $\rho_0, \sigma_0 \in \mathcal{S}(\mathcal{H})$  with finite energy and let  $\mu, \nu$  be probability measures on  $\mathbb{C}^m$  such that*

$$\begin{aligned} & (\mathbb{E}_{Z \sim \mu} Z - \mathbb{E}_{W \sim \nu} W) \mathrm{Tr}[(\rho_0 - \sigma_0) R_i] = 0, \\ & \mathbb{E}_{Z \sim \mu} \|Z\|^2 < \infty, \quad \mathbb{E}_{W \sim \nu} \|W\|^2 < \infty. \end{aligned} \quad (98)$$

Let  $\rho_1$  and  $\sigma_1$  be the quantum states obtained adding classical noise distributed according to  $\mu$  and  $\nu$  to  $\rho_0$  and  $\sigma_0$ , respectively:

$$\rho_1 = \int_{\mathbb{C}^m} D(z) \rho_0 D(z)^\dagger d\mu(z), \quad \sigma_1 = \int_{\mathbb{C}^m} D(w) \sigma_0 D(w)^\dagger d\nu(w), \quad (99)$$

where  $D(z)$  is the displacement operator (74). Then,

$$D(\rho_1, \sigma_1)^2 \leq D(\rho_0, \sigma_0)^2 + W(\mu, \nu)^2, \quad (100)$$

where  $W$  denotes the classical Wasserstein distance.

*Proof.* Let  $\Pi_0 \in \mathcal{C}(\rho_0, \sigma_0)$  and  $\pi \in \mathcal{C}(\mu, \nu)$ , and let

$$\Pi_1 = \int_{\mathbb{C}^m \times \mathbb{C}^m} \left( D(w) \otimes D(z)^\dagger \right) \Pi_0 \left( D(w)^\dagger \otimes D(z)^T \right) d\pi(z, w) \in \mathcal{C}(\rho_1, \sigma_1). \quad (101)$$

With the help of (75), we have for any  $z, w \in \mathbb{C}^m$

$$\begin{aligned} & \left( D(w)^\dagger \otimes D(z)^T \right) C \left( D(w) \otimes D(z)^\dagger \right) \\ &= C + \sum_{i=1}^m \left( (w_i - z_i)^* (a_i - a_i^T) + (w_i - z_i) (a_i - a_i^T)^\dagger \right) + |w - z|^2, \end{aligned} \quad (102)$$

therefore

$$C(\Pi_1) = C(\Pi_0) + \mathbb{E}_{(Z,W) \sim \pi} |Z - W|^2, \quad (103)$$

and the claim follows taking the infimum over  $\Pi_0 \in \mathcal{C}(\rho_0, \sigma_0)$  and  $\pi \in \mathcal{C}(\mu, \nu)$ .  $\square$

**Corollary 2.** *The quantum Wasserstein distance is translation invariant, i.e., for any  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with finite energy and any  $z \in \mathbb{C}^m$ ,*

$$D \left( D(z) \rho D(z)^\dagger, D(z) \sigma D(z)^\dagger \right) = D(\rho, \sigma). \quad (104)$$

*Proof.* The claim follows from Theorem 5 with both  $Z$  and  $W$  equal to  $z$  with probability one.  $\square$

### 5.3. Semiclassical limit

The coherent states of a quantum Gaussian system are the pure quantum Gaussian states obtained applying a displacement operator to the vacuum state and are the eigenvectors of the ladder operators:

$$|z\rangle = D(z)|0\rangle, \quad a_i|z\rangle = z_i|z\rangle, \quad z \in \mathbb{C}^m, \quad i = 1, \dots, m. \quad (105)$$

They are the easiest state to realize in laboratory, and they are considered to be the most classical pure states of the system. Coherent states form an overcomplete set and satisfy the resolution of the identity [48]

$$\int_{\mathbb{C}^m} |z\rangle\langle z| \frac{dz}{\pi^m} = \mathbb{I}_{\mathcal{H}}, \quad (106)$$

where the integral converges weakly. Therefore, for any quantum state  $\rho$ , the function

$$Q(z) = \langle z|\rho|z\rangle, \quad z \in \mathbb{C}^m, \quad (107)$$

called Husimi Q representation of  $\rho$  [47, 50], defines a probability density on  $\mathbb{C}^m$  with normalization

$$\int_{\mathbb{C}^m} \langle z|\rho|z\rangle \frac{dz}{\pi^m} = 1. \quad (108)$$

A quantum state is completely determined by its Q representation. Moreover, the Q representation is the probability distribution of a particular measurement that can be performed on the state, which is called heterodyne measurement [47, 50], and is one of the main measurements in quantum optics.

A quantum state  $\rho$  is called semiclassical if it can be expressed as a convex mixture of coherent states, *i.e.*,

$$\rho = \int_{\mathbb{C}^m} |z\rangle\langle z| d\hat{\mu}(z) \quad (109)$$

for some probability measure  $\hat{\mu}$  on  $\mathbb{C}^m$ , where the integral converges in the trace norm. If this is the case, it can be proved that  $\hat{\mu}$  is uniquely determined, and is called the Glauber-Sudarshan P representation of  $\rho$  [47, 50].

From their definitions, the Q representation of a semiclassical state is equal to its P representation convolved with the Gaussian function  $e^{-|z|^2}$ .

We will prove that as the GMPC distance, our quantum Wasserstein distance is upper bounded by the Wasserstein distance between the P representations and lower bounded by the Wasserstein distance between the Q representations. In the semiclassical limit the P and Q representations become the same, hence both our distance and the GMPC distance recover the classical Wasserstein distance.

**Theorem 6 (P representation).** *Let  $\hat{\mu}$  and  $\hat{\nu}$  be probability measures on  $\mathbb{C}^m$  with  $\mathbb{E}_{Z \sim \hat{\mu}} \|Z\|^2 < \infty$  and  $\mathbb{E}_{W \sim \hat{\nu}} \|W\|^2 < \infty$ , and let*

$$\rho = \int_{\mathbb{C}^m} |z\rangle\langle z| d\hat{\mu}(z), \quad \sigma = \int_{\mathbb{C}^m} |w\rangle\langle w| d\hat{\nu}(w) \quad (110)$$

*be the associated semiclassical states, where both integrals converge in the trace norm. Then,*

$$D(\rho, \sigma)^2 \leq W(\hat{\mu}, \hat{\nu})^2 + m. \quad (111)$$

*Proof.* We define for any  $\hat{\pi} \in \mathcal{C}(\hat{\mu}, \hat{\nu})$

$$\Pi = \int_{\mathbb{C}^m \times \mathbb{C}^m} |w\rangle\langle w| \otimes |z\rangle\langle z|^T d\hat{\pi}(z, w) \in \mathcal{C}(\rho, \sigma), \quad (112)$$

where the integral converges in the trace norm. With the help of (105), we have for any  $z, w \in \mathbb{C}^m$

$$C(|w\rangle\langle w| \otimes |z\rangle\langle z|^T) = m + |z - w|^2, \quad (113)$$

therefore we get from Proposition 9

$$C(\Pi) = m + \int_{\mathbb{C}^m \times \mathbb{C}^m} |w - z|^2 d\hat{\pi}(z, w), \quad (114)$$

and the claim follows taking the inf over  $\hat{\pi} \in \mathcal{C}(\hat{\mu}, \hat{\nu})$ .  $\square$

**Corollary 3.** *Let  $\mu$  be a probability measure on  $\mathbb{C}^m$  with  $\mathbb{E}_{Z \sim \mu} \|Z\|^2 < \infty$ , and let*

$$\rho = \int_{\mathbb{C}^m} |z\rangle\langle z| d\mu(z) \quad (115)$$

*be a semiclassical state. Then,*

$$D(\rho, \rho)^2 \leq m. \quad (116)$$

*Remark 10.* In the same hypotheses of Corollary 3, [24, Theorem 2.3] implies that

$$D_{\text{GMPC}}(\rho, \rho) = \sqrt{m}. \quad (117)$$

On the contrary, our distance does not always saturate the upper bound of Corollary 3. Indeed, if  $\rho$  is the  $m$ -mode quantum thermal Gaussian state  $\omega(\nu)$  with covariance matrix  $\nu I_{2m}$ , we get from Theorem 3

$$D(\omega(\nu), \omega(\nu)) = \sqrt{m} \left( \sqrt{\nu + \frac{1}{2}} - \sqrt{\nu - \frac{1}{2}} \right), \quad (118)$$

which is strictly decreasing in  $\nu$ , is equal to  $\sqrt{m}$  for  $\nu = \frac{1}{2}$  and tends to 0 for  $\nu \rightarrow \infty$ .

**Theorem 7 (Q representation).** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  with finite energy, and let  $\mu$  and  $\nu$  be the probability measures on  $\mathbb{C}^m$  associated to their respective Husimi Q representations:*

$$d\mu(z) = \langle z|\rho|z\rangle \frac{dz}{\pi^m}, \quad d\nu(w) = \langle w|\sigma|w\rangle \frac{dw}{\pi^m}, \quad z, w \in \mathbb{C}^m. \quad (119)$$

*Then,*

$$D(\rho, \sigma)^2 \geq W(\mu, \nu)^2 - m. \quad (120)$$

*Proof.* Let  $\pi$  be the probability measure on  $\mathbb{C}^m \times \mathbb{C}^m$  given by

$$d\pi(z, w) = \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ \Pi \left( |w\rangle\langle w| \otimes |z\rangle\langle z|^T \right) \right] \frac{dz dw}{\pi^{2m}}. \quad (121)$$

The marginals of  $\pi$  are  $\mu$  and  $\nu$ , therefore  $\pi \in \mathcal{C}(\mu, \nu)$ . We have

$$\begin{aligned} W(\mu, \nu)^2 &\leq \int_{\mathbb{C}^m \times \mathbb{C}^m} |w - z|^2 d\pi(z, w) \\ &\stackrel{(a)}{=} \sum_{i=1}^m \int_{\mathbb{C}^m \times \mathbb{C}^m} \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ (a_i^\dagger \Pi a_i - a_i^\dagger a_i^T \Pi - \Pi a_i a_i^{\dagger T} + a_i^T \Pi a_i^{\dagger T}) \right. \\ &\quad \left. (|w\rangle\langle w| \otimes |z\rangle\langle z|^T) \right] \frac{dz dw}{\pi^{2m}} \\ &\stackrel{(b)}{=} \sum_{i=1}^m \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ a_i^\dagger \Pi a_i - a_i^\dagger a_i^T \Pi - \Pi a_i a_i^{\dagger T} + a_i^T \Pi a_i^{\dagger T} \right] \\ &\stackrel{(c)}{=} m + \sum_{i=1}^m \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ a_i^\dagger \Pi a_i - a_i^\dagger \Pi a_i^T - a_i^{\dagger T} \Pi a_i + a_i^{\dagger T} \Pi a_i^T \right] \\ &= m + \sum_{i=1}^m \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*} \left[ (a_i - a_i^T)^\dagger \Pi (a_i - a_i^T) \right] = m + C(\Pi), \end{aligned} \quad (122)$$

where (a) follows from (105), (b) follows from (106) and (c) follows from (66). The claim follows taking the inf over  $\Pi \in \mathcal{C}(\rho, \sigma)$ .  $\square$



#### 5.4. Quantum Stam inequality

The Wasserstein distance between a quantum state and itself satisfies the following Stam inequality.

**Theorem 8 (quantum Stam inequality).** *Let  $\rho_0, \rho_1 \in \mathcal{S}(\mathcal{H})$  with finite energy, and for any  $0 \leq \eta \leq 1$  let*

$$\rho_\eta = \mathcal{B}_\eta(\rho_1 \otimes \rho_0), \quad (123)$$

where  $\mathcal{B}_\eta$  is the beamsplitter (78). Then,

$$\frac{1}{D(\rho_\eta, \rho_\eta)^2} \geq \frac{\eta}{D(\rho_1, \rho_1)^2} + \frac{1-\eta}{D(\rho_0, \rho_0)^2}. \quad (124)$$

*Proof.* Let  $\rho \in \mathcal{S}(\mathcal{H})$  have finite energy. From Proposition 3 and Remark 3, also  $\|\sqrt{\rho}\rangle\rangle \langle\langle \sqrt{\rho}$  has finite energy, i.e.,  $[R_i, \sqrt{\rho}] \in \mathcal{T}_2(\mathcal{H})$  for any  $i = 1, \dots, 2m$ . Therefore, we can define the Fisher information matrix of  $\rho$  as

$$J(\rho)_{ab} = \langle\langle [R_a, \sqrt{\rho}] \|[R_b, \sqrt{\rho}] \rangle\rangle, \quad a, b = 1, \dots, 2m, \quad (125)$$

such that

$$D(\rho, \rho)^2 = \text{tr } J(\rho). \quad (126)$$

For any  $k \in \mathbb{R}^{2m}$ , let

$$\rho(k) = e^{-i \sum_{i=1}^{2m} k_i R_i} \rho e^{i \sum_{i=1}^{2m} k_i R_i}. \quad (127)$$

We have for any  $k, q \in \mathbb{R}^{2m}$

$$e^{i \sum_{i=1}^{2m} k_i R_i} e^{-i \sum_{i=1}^{2m} q_i R_i} \propto e^{i \sum_{i=1}^{2m} (k_i - q_i) R_i}, \quad (128)$$

therefore, since the Hilbert-Schmidt scalar product is unitarily invariant, we get

$$\langle\langle \sqrt{\rho(k)} \|\sqrt{\rho(q)} \rangle\rangle = \langle\langle \sqrt{\rho} \|\sqrt{\rho(q-k)} \rangle\rangle, \quad (129)$$

and

$$\begin{aligned} J(\rho)_{ab} &= \frac{\partial^2}{\partial k_a \partial q_b} \left. \langle\langle \sqrt{\rho(k)} \|\sqrt{\rho(q)} \rangle\rangle \right|_{k=q=0} \\ &= \frac{\partial^2}{\partial k_a \partial q_b} \left. \langle\langle \sqrt{\rho} \|\sqrt{\rho(q-k)} \rangle\rangle \right|_{k=q=0} \\ &= \frac{\partial^2}{\partial k_a \partial k_b} \left. \langle\langle \sqrt{\rho} \|\sqrt{\rho(k)} \rangle\rangle \right|_{k=0}. \end{aligned} \quad (130)$$

From Lieb's concavity theorem [44], the function  $(\rho, \sigma) \mapsto -\langle\langle \sqrt{\rho} \|\sqrt{\sigma} \rangle\rangle$  is jointly convex, and therefore contractive with respect to the joint application of a quantum channel to both arguments. Applying this property to the

beamsplitter we get that for any  $k, q \in \mathbb{R}^{2m}$

$$\begin{aligned}
& - \left\langle \left\| \sqrt{\rho_1} \otimes \rho_0 \right\| \left\| \sqrt{\rho_1(k) \otimes \rho_0(q)} \right\| \right\rangle \\
& \geq - \left\langle \left\| \sqrt{\mathcal{B}_\eta(\rho_1 \otimes \rho_0)} \right\| \left\| \sqrt{\mathcal{B}_\eta(\rho_1(k) \otimes \rho_0(q))} \right\| \right\rangle \\
& = - \left\langle \left\| \sqrt{\rho_\eta} \right\| \left\| \sqrt{\rho_\eta(\sqrt{\eta}k + \sqrt{1-\eta}q)} \right\| \right\rangle, \tag{131}
\end{aligned}$$

where the last equality follows from [53, Lemma 1], stating that

$$\mathcal{B}_\eta(\rho_1(k) \otimes \rho_0(q)) = (\mathcal{B}_\eta(\rho_1 \otimes \rho_0)) \left( \sqrt{\eta}k + \sqrt{1-\eta}q \right). \tag{132}$$

Since both sides of (131) are equal to  $-1$  for  $k = q = 0$ , which is their global minimum, the inequality (131) translates to the Hessian with respect to  $(k, q)$ :

$$\begin{pmatrix} \text{tr } J(\rho_1) & 0 \\ 0 & \text{tr } J(\rho_0) \end{pmatrix} \geq \begin{pmatrix} \frac{\sqrt{\eta}}{\sqrt{1-\eta}} & \\ & \end{pmatrix} \text{tr } J(\rho_\eta) \begin{pmatrix} \sqrt{\eta} & \sqrt{1-\eta} \end{pmatrix}. \tag{133}$$

Putting together (133) and (126) we get

$$\begin{pmatrix} D(\rho_1, \rho_1)^2 & 0 \\ 0 & D(\rho_0, \rho_0)^2 \end{pmatrix} \geq \begin{pmatrix} \frac{\sqrt{\eta}}{\sqrt{1-\eta}} & \\ & \end{pmatrix} D(\rho_\eta, \rho_\eta)^2 \begin{pmatrix} \sqrt{\eta} & \sqrt{1-\eta} \end{pmatrix}, \tag{134}$$

therefore

$$\begin{pmatrix} D(\rho_1, \rho_1)^2 - \eta D(\rho_\eta, \rho_\eta)^2 & -\sqrt{\eta(1-\eta)} D(\rho_\eta, \rho_\eta)^2 \\ -\sqrt{\eta(1-\eta)} D(\rho_\eta, \rho_\eta)^2 & D(\rho_0, \rho_0)^2 - (1-\eta) D(\rho_\eta, \rho_\eta)^2 \end{pmatrix} \geq 0, \tag{135}$$

and the claim follows from the positivity of the determinant of (135).  $\square$

## 6. Conclusions and perspectives

We have proposed a new quantum generalization of the Wasserstein distance that has the property that the transport plans are in one-to-one correspondence with quantum channels. This property allows for the first time to interpret the quantum transport plans as physical operations performed on the system. We have started to explore the properties of our distance, proving *e.g.* that it satisfies a modified triangle inequality and determining the optimal transport plans between thermal quantum Gaussian states.

The most natural application of our distance is the theory of quantum rate-distortion coding [54–61], whose goal is to determine the maximum achievable rates for the lossy compression of quantum states with a given distortion. The distortion measure is defined through the quantum state obtained applying in sequence the encoding and the decoding quantum channel to half of a purification of the source state. Therefore, the problem of determining the maximum achievable compression rates for a given distortion and a given source state can be related to a sequence of optimal transport problems where the transport plan is given by the composition of the encoding and decoding quantum channel and optimization is performed over both the transport plan and the target state. So far, most of the effort has focused

on the entanglement fidelity as quantum distortion measure. By contrast, the most common distortion measure in classical rate-distortion theory for signals with values in  $\mathbb{R}^n$  is the average square norm of the difference between original and distorted signal [62, Chapter 10], and coincides with the definition of the transport cost (5) of a classical coupling. Our quantum transport cost provides the most natural generalization of the classical distortion measure, and we will explore in future works its applications in this direction.

A further natural application of our distance is the field of quantum machine learning, and in particular the quantum version of the Generative Adversarial Networks (GANs) [30]. Quantum GANs provide an algorithm to learn a target quantum state with a parametric quantum circuit. A key role is played by the cost function that quantifies the quality of the approximation, *i.e.*, the distance between the target and the generated state. The distances induced by optimal mass transport have turned out to be the best choice for the cost function of classical GANs [63], and are showing extremely promising results also in the quantum setting [31, 64, 65]. The formulation of the Wasserstein distance with quantum couplings is fundamental to the application to classical GANs, since it allows for a fast algorithm to compute a regularized version of the distance [10]. The quantum Wasserstein distance proposed in this paper together with the GMPC distance are defined through a quantum counterpart of the couplings, and therefore constitute very promising candidates for the cost function of quantum GANs.

## Acknowledgements

GdP was supported by the USA Air Force Office of Scientific Research and the USA Army Research Office under the Blue Sky program. DT is a member of Gnampa group (INdAM). DT has been partially supported by the University of Pisa, Project PRA 2018-49 and Gnampa project 2019 “Proprietà analitiche e geometriche di campi aleatori”.

We thank Luigi Ambrosio for useful remarks on a preliminary version of this work and Emanuele Caglioti for fruitful discussions and for a careful reading of the paper.

## Appendix A. Continuity of the quantum Wasserstein distance

### A.1. Compactness of quantum plans

Recall that a sequence  $\{A_n\}_{n=0}^{\infty} \subset \mathcal{B}(\mathcal{H})$  weakly converges towards  $A \in \mathcal{B}(\mathcal{H})$  if, for every  $|\phi\rangle, |\psi\rangle \in \mathcal{H}$ ,

$$\lim_{n \rightarrow \infty} \langle \phi | A_n | \psi \rangle = \langle \phi | A | \psi \rangle. \quad (136)$$

**Lemma 4.** *Given  $\{\rho_n\}_{n=0}^{\infty} \subseteq \mathcal{S}(\mathcal{H})$ , there exists  $\rho \in \mathcal{T}_1(\mathcal{H})$ ,  $\rho \geq 0$ ,  $\text{Tr}_{\mathcal{H}}[\rho] \leq 1$  and a subsequence  $\{\rho_{n_k}\}_{k=0}^{\infty}$  weakly converging towards  $\rho$ . If  $\text{Tr}_{\mathcal{H}}[\rho] = 1$  then convergence holds in the trace norm.*

*Proof.* The first statement is a consequence of Banach-Alaoglu theorem and lower semicontinuity of the trace, see also the proof of [35, Theorem 11.2]. The second statement is [35, Lemma 11.1].  $\square$

Because of the result above, we may consider weak convergence or equivalently in the trace norm for a sequence of quantum states  $\{\rho_n\}_{n=0}^\infty$ , provided that the limit  $\rho$  is a quantum state. Hence we often omit to specify which type of convergence we consider on quantum states in the results below.

**Proposition 8.** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  be quantum states and let  $\{\Pi_n\}_{n=0}^\infty \subset \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*)$  be a sequence of quantum couplings  $\Pi_n \in \mathcal{C}(\rho_n, \sigma_n)$  such that  $\lim_n \rho_n = \rho$ ,  $\lim_n \sigma_n = \sigma$ . Then, there exists a subsequence  $\{\Pi_{n_k}\}_{k=0}^\infty$  converging towards  $\Pi \in \mathcal{C}(\rho, \sigma)$ .*

*Proof.* Existence of a weak limit  $\Pi = \lim_k \Pi_{n_k} \in \mathcal{T}_1(\mathcal{H})$ ,  $\Pi \geq 0$ , follows from Lemma 4. To argue that  $\text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[\Pi] = 1$ , let  $P, Q \in \mathcal{B}(\mathcal{H})$  be projectors with  $P + Q = \mathbb{I}_{\mathcal{H}}$  and  $P$  with finite rank. We have

$$\begin{aligned} \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[\Pi] &\geq \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes P^T)\Pi(P \otimes P^T)] \\ &= \lim_{n \rightarrow \infty} \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes P^T)\Pi_n(P \otimes P^T)], \end{aligned} \quad (137)$$

where the limit holds because  $P \otimes P^T$  has finite rank. Then

$$\begin{aligned} &\text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes P^T)\Pi_n(P \otimes P^T)] \\ &= \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[\Pi_n] - \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes Q^T)\Pi_n(P \otimes Q^T)] \\ &\quad - \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(Q \otimes \mathbb{I}_{\mathcal{H}^*})\Pi_n(Q \otimes \mathbb{I}_{\mathcal{H}^*})] \\ &= 1 - \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes Q^T)\Pi_n(P \otimes Q^T)] - \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(Q \otimes \mathbb{I}_{\mathcal{H}^*})\Pi_n(Q \otimes \mathbb{I}_{\mathcal{H}^*})]. \end{aligned} \quad (138)$$

We have the inequality

$$\text{Tr}_{\mathcal{H}}[(P \otimes Q^T)\Pi_n(P \otimes Q^T)] \leq \text{Tr}_{\mathcal{H}}[(\mathbb{I}_{\mathcal{H}} \otimes Q^T)\Pi_n(\mathbb{I}_{\mathcal{H}} \otimes Q^T)] \leq Q^T \rho_n^T Q^T. \quad (139)$$

Taking the partial trace with respect to  $\mathcal{H}^*$ ,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(P \otimes Q^T)\Pi_n(P \otimes Q^T)] &\leq \limsup_{n \rightarrow \infty} \text{Tr}_{\mathcal{H}^*}[Q^T \rho_n^T Q^T] \\ &= \text{Tr}_{\mathcal{H}^*}[Q^T \rho^T Q^T] = \text{Tr}_{\mathcal{H}}[Q \rho Q]. \end{aligned} \quad (140)$$

Similarly,  $\limsup_n \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[(Q \otimes \mathbb{I}_{\mathcal{H}^*})\Pi_n(Q \otimes \mathbb{I}_{\mathcal{H}^*})] \leq \text{Tr}_{\mathcal{H}}[Q \sigma Q]$ . It follows that

$$\text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[\Pi] \geq 1 - \text{Tr}_{\mathcal{H}}[Q \rho Q] - \text{Tr}_{\mathcal{H}}[Q \sigma Q]. \quad (141)$$

Letting  $P \rightarrow \mathbb{I}_{\mathcal{H}}$  weakly, both  $Q \rho Q$  and  $Q \sigma Q$  converge to 0 in the trace norm, hence the thesis.  $\square$

A similar argument gives the following result.

**Lemma 5.** *Given quantum states  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  let  $\{P_n\}_{n=0}^\infty, \{Q_n\}_{n=0}^\infty \subset \mathcal{B}(\mathcal{H})$  be such that  $0 \leq P_n \leq \mathbb{I}_{\mathcal{H}}, 0 \leq Q_n \leq \mathbb{I}_{\mathcal{H}}$ , and  $\lim_n P_n \rho P_n = \rho, \lim_n Q_n \sigma Q_n = \sigma$ . Then, for any  $\Pi \in \mathcal{C}(\rho, \sigma)$ ,*

$$\lim_{n \rightarrow \infty} (Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T) = \Pi \quad (142)$$

in the trace norm.

*Proof.* The assumption  $\lim_n P_n \rho P_n = \rho$  implies that  $\{P_n\}_{n=0}^\infty$  weakly converges to the identity operator on  $\text{supp } \rho$ . By a density argument, it is sufficient to prove that, for any  $|\phi\rangle$  eigenvector for  $\rho, \rho|\phi\rangle = p|\phi\rangle$ , with  $p > 0$ , one has  $\lim_n P_n |\phi\rangle = |\phi\rangle$ . Since  $p|\phi\rangle\langle\phi| \leq \rho$ , we have indeed

$$\begin{aligned} \|(\mathbb{I}_{\mathcal{H}} - P_n)|\phi\rangle\|^2 &= \text{Tr}[(\mathbb{I}_{\mathcal{H}} - P_n)|\phi\rangle\langle\phi|(\mathbb{I}_{\mathcal{H}} - P_n)] \\ &\leq p^{-1} \text{Tr}[(\mathbb{I}_{\mathcal{H}} - P_n)\rho(\mathbb{I}_{\mathcal{H}} - P_n)] \rightarrow 0. \end{aligned} \quad (143)$$

Similarly, we have that  $\{Q_n\}_{n=0}^\infty$  weakly converges to the identity operator on  $\text{supp } \sigma$ . We deduce that  $\{Q_n \otimes P_n^T\}_{n=0}^\infty$  weakly converges to the identity operator on  $\text{supp } \sigma \otimes \text{supp } \rho^T$ , which contains  $\text{supp } \Pi$ . Hence (142) holds in the weak sense.

We prove next that

$$\lim_{n \rightarrow \infty} \text{Tr}[(Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T)] = 1. \quad (144)$$

We write

$$\begin{aligned} &\text{Tr}[(Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T)] \\ &= 1 - \text{Tr}[(Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T)) \Pi (Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T))] \\ &\quad - \text{Tr}[(\mathbb{I}_{\mathcal{H}} - Q_n) \otimes \mathbb{I}_{\mathcal{H}^*}] \Pi [(\mathbb{I}_{\mathcal{H}} - Q_n) \otimes \mathbb{I}_{\mathcal{H}^*}]. \end{aligned} \quad (145)$$

If we take the partial trace with respect to  $\mathcal{H}$ , then

$$\begin{aligned} &\text{Tr}_{\mathcal{H}}[(Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T)) \Pi (Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T))] \\ &\leq (\mathbb{I}_{\mathcal{H}^*} - P_n^T) \text{Tr}_{\mathcal{H}}[\Pi] (\mathbb{I}_{\mathcal{H}^*} - P_n^T) = (\rho - P_n \rho P_n)^T, \end{aligned} \quad (146)$$

so that

$$\lim_{n \rightarrow \infty} \text{Tr}[(Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T)) \Pi (Q_n \otimes (\mathbb{I}_{\mathcal{H}^*} - P_n^T))] = 0. \quad (147)$$

Arguing similarly, taking the partial trace with respect to  $\mathcal{H}^*$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Tr}[(\mathbb{I}_{\mathcal{H}} - Q_n) \otimes \mathbb{I}_{\mathcal{H}^*}] \Pi [(\mathbb{I}_{\mathcal{H}} - Q_n) \otimes \mathbb{I}_{\mathcal{H}^*}] &\leq \lim_{n \rightarrow \infty} \text{Tr}_{\mathcal{H}}[\sigma - Q_n \sigma Q_n] \\ &= 0, \end{aligned} \quad (148)$$

and (144) follows.

To conclude that (142) holds in the trace norm, define for  $n$  sufficiently large the quantum state

$$\Pi_n = (Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T) / \text{Tr}[(Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T)], \quad (149)$$

so that  $\lim_n \Pi_n = \Pi$  weakly, hence in the trace norm. By (144) this implies convergence in the trace norm also in (142).  $\square$

## A.2. Energy

We prove in this subsection some results that are needed to deal with the energy of quantum states with respect to unbounded operators.

**Lemma 6.** *Let  $\mathcal{H}_A$ ,  $\mathcal{H}_B$  be Hilbert spaces and let  $A$  be a self-adjoint operator on  $\mathcal{H}_A$ . Then, there exists a sequence of bounded self-adjoint operators  $\{A_n\}_{n=0}^\infty \subset \mathcal{B}(\mathcal{H})$  such that, for every for every quantum state  $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ ,  $E_{A_n \otimes \mathbb{I}_B}(\rho)$  increases towards  $E_{A \otimes \mathbb{I}_B}(\rho)$ .*

*Proof.* It is sufficient to consider the case of a pure state  $\rho = |\psi\rangle\langle\psi|$ , because the general case follows by Remark 6 and the monotone convergence theorem for series.

Define by functional calculus [66, Theorem VIII.5]  $A_n = A\chi_{(-n,n)}(A)$ , where  $\chi_{(-n,n)}(x) = 1$  if  $x \in (-n, n)$ ,  $\chi_{(-n,n)}(x) = 0$  otherwise. Let us argue without the Hilbert space  $\mathcal{H}_B$  first (i.e.,  $\mathcal{H}_B = \mathbb{C}$ ). Then,

$$E_{A_n}(\rho) = \|A_n|\psi\rangle\|^2 = \int_{-n}^n \lambda^2 d\langle\psi|P_\lambda|\psi\rangle, \quad (150)$$

where  $P_\lambda = \chi_{(-\infty, \lambda]}(A)$  denotes the resolution of the identity associated to  $A$ . The expression is increasing with respect to  $n$  and  $\sup_n \|A_n|\psi\rangle\|^2 < \infty$  if and only if  $\psi$  belongs to the domain of  $A$ , because, defining  $|\psi_n\rangle = \chi_{(-n,n)}(A)|\psi\rangle$ , the sequence  $\{|\psi_n\rangle\}_{k=0}^\infty \subset \mathcal{H}$  converges towards  $|\psi\rangle$  with  $A|\psi_n\rangle$  bounded and  $A$  is a closed operator. In this case, we also have  $\lim_n A_n|\psi\rangle = A|\psi\rangle$  by [66, Theorem VIII.5, c)], hence the conclusion in this case. The thesis with the additional Hilbert space  $\mathcal{H}_B$  follows from repeating the argument by noticing that

$$A_n \otimes \mathbb{I}_B = (A \otimes \mathbb{I}_B)\chi_{(-n,n)}(A \otimes \mathbb{I}_B). \quad (151)$$

□

The following result shows that  $E_A(\rho)$  does not depend on the ensemble that generates  $\rho$ .

**Proposition 9.** *If  $\rho = \sum_{k=0}^\infty |\phi_k\rangle\langle\phi_k|$  with  $\{|\phi_k\rangle\}_{k=0}^\infty \subset \mathcal{H}$ , then*

$$E_A(\rho) = \sum_{k=0}^\infty \|A|\phi_k\rangle\|^2, \quad (152)$$

where we let  $\|A|\phi\rangle\|^2 = \infty$  if  $|\phi\rangle$  does not belong to the domain of  $A$ .

Analogously, let  $\mu$  be a probability measure on  $\mathcal{H}$ , and let

$$\rho = \int_{\mathcal{H}} |\psi\rangle\langle\psi| d\mu(\psi). \quad (153)$$

Then,

$$E_A(\rho) = \int_{\mathcal{H}} \|A|\psi\rangle\|^2 d\mu(\psi). \quad (154)$$

*Proof.* If we define  $|\hat{\phi}_k\rangle = |\phi_k\rangle/\| |\phi_k\rangle \|$ , then the thesis can be rewritten as

$$E_A(\rho) = \sum_{k=0}^\infty \| |\phi_k\rangle \|^2 E_A(|\hat{\phi}_k\rangle\langle\hat{\phi}_k|), \quad (155)$$

hence, by Lemma 6 and the monotone convergence theorem for series, it is sufficient to prove the result for  $A \in \mathcal{B}(\mathcal{H})$ .

Arguing as in [34, Theorem 2.6] (noticing that the proof holds also in the infinite dimensional case), there exists a unitary map  $U : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ ,  $U^\dagger U = \mathbb{I}_{\ell(\mathbb{N})}$ , i.e.,  $(u_{ij})_{i,j=0}^\infty$  with  $\sum_{k=0}^\infty u_{ki}^* u_{kj} = \delta_{ij}$ , such that

$$|\phi_k\rangle = \sum_{i=0}^\infty u_{ki} \sqrt{p_i} |\psi_i\rangle. \quad (156)$$

Then,

$$\begin{aligned} \sum_{k=0}^\infty \|A|\phi_k\rangle\|^2 &= \sum_{k=0}^\infty \langle \phi_k | A^2 | \phi_k \rangle = \sum_{k=0}^\infty \sum_{i,j=0}^\infty \sqrt{p_i p_j} u_{ki}^* u_{kj} \langle \psi_i | A^2 | \psi_j \rangle \\ &= \sum_{i,j=0}^\infty \sqrt{p_i p_j} \langle \psi_i | A^2 | \psi_j \rangle \left( \sum_{k=0}^\infty u_{ki}^* u_{kj} \right) = \sum_{i=0}^\infty p_i \langle \psi_i | A^2 | \psi_i \rangle = E_A(\rho), \end{aligned} \quad (157)$$

and the claim (152) follows.

The integral version (154) of the claim can be proved along the same lines employing Beppo Levi's monotone convergence theorem [67, Theorem 2.8.2] and Fubini's theorem [67, Theorem 3.4.4].  $\square$

**Proposition 10.**  $E_A$  is lower semicontinuous, i.e., if  $\{\rho_n\}_{n=0}^\infty \subset \mathcal{S}(\mathcal{H})$  converge towards  $\rho \in \mathcal{S}(\mathcal{H})$ , then  $E_A(\rho) \leq \liminf_n E_A(\rho_n)$ .

*Proof.* If  $A$  is bounded then, by Remark 7,  $E_A(\rho) = \text{Tr}_{\mathcal{H}} [A \rho A]$ , which is continuous with respect to the trace norm. The general case follows by Lemma 6, since  $E_A(\rho) = \sup_n E_{A_n}(\rho)$  and a pointwise supremum of continuous functions is lower semicontinuous.  $\square$

**Proposition 11.** Let  $A, B$  be respectively self-adjoint operators on Hilbert spaces  $\mathcal{H}_A, \mathcal{H}_B$  and let  $\Pi \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$  with  $\rho = \text{Tr}_A[\Pi]$ ,  $\sigma = \text{Tr}_B[\Pi]$ . Then

$$E_{A \otimes \mathbb{I}_B}(\Pi) = E_A(\rho), \quad E_{\mathbb{I}_A \otimes B}(\Pi) = E_B(\sigma). \quad (158)$$

If both energies above are finite, then

$$E_{A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes B}(\Pi) + E_{A \otimes \mathbb{I}_B - \mathbb{I}_A \otimes B}(\Pi) = 2E_A(\rho) + 2E_B(\sigma). \quad (159)$$

*Proof.* We prove the first identity in (158), the second one being similar. By Lemma 6, it is sufficient to argue when  $A$  is bounded. Then, by Remark 7,

$$\begin{aligned} E_{A \otimes \mathbb{I}_B}(\Pi) &= \text{Tr}_{AB} [(A \otimes \mathbb{I}_B) \Pi (A \otimes \mathbb{I}_B)] \\ &= \text{Tr}_A [A \text{Tr}_B [\Pi] A] = \text{Tr}_A [A \sigma A] = E_A(\sigma). \end{aligned} \quad (160)$$

To prove (159), by Remark 6, is it sufficient to consider the case of a pure state  $\Pi = |\psi\rangle\langle\psi|$ , so that it reads

$$\begin{aligned} &\| (A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes B) |\psi\rangle \|^2 + \| (A \otimes \mathbb{I}_B - \mathbb{I}_A \otimes B) |\psi\rangle \|^2 \\ &= 2\| (A \otimes \mathbb{I}_B) |\psi\rangle \|^2 + 2\| (\mathbb{I}_A \otimes B) |\psi\rangle \|^2 \end{aligned} \quad (161)$$

assuming that both terms in the right hand side are finite, *i.e.*,  $|\psi\rangle$  belongs to the domain of  $A \otimes \mathbb{I}_B$  and  $\mathbb{I}_A \otimes B$ . But then

$$(A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes B)|\psi\rangle = (A \otimes \mathbb{I}_B)|\psi\rangle + (\mathbb{I}_A \otimes B)|\psi\rangle, \quad (162)$$

$$(A \otimes \mathbb{I}_B - \mathbb{I}_A \otimes B)|\psi\rangle = (A \otimes \mathbb{I}_B)|\psi\rangle - (\mathbb{I}_A \otimes B)|\psi\rangle, \quad (163)$$

hence the thesis by straightforward computation.  $\square$

### A.3. Convergence of the quantum Wasserstein distance

As in section 3, we fix quadratures  $\{R_1, \dots, R_N\}$ , *i.e.*, self-adjoint operators on  $\mathcal{H}$ . Then the cost associated to a quantum plan  $\Pi \in \mathcal{C}(\rho, \sigma)$  (Definition 5) reads

$$C(\Pi) = \sum_{i=1}^N E_{R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T}(\Pi) \quad (164)$$

and a quantum state  $\rho \in \mathcal{S}(\mathcal{H})$  has finite energy (Definition 6) if

$$\sum_{i=1}^N E_{R_i}(\rho) < \infty. \quad (165)$$

We are in a position to give a proof of Proposition 3.

**Proposition 3** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  have finite energy. Then, any plan  $\Pi \in \mathcal{C}(\rho, \sigma)$  has finite cost.*

*Proof.* For  $i = 1, \dots, N$ , we apply Proposition 11 with  $A = R_i$ ,  $B = R_i^T$  so that

$$\begin{aligned} C(\Pi) &= \sum_{i=1}^N E_{R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T}(\Pi) \\ &= \sum_{i=1}^N \left( 2E_{R_i}(\rho) + 2E_{R_i}(\sigma) - E_{R_i \otimes \mathbb{I}_{\mathcal{H}^*} - \mathbb{I}_{\mathcal{H}} \otimes R_i^T}(\Pi) \right) < \infty, \end{aligned} \quad (166)$$

where we also used that  $E_{R_i^T}(\rho^T) = E_{R_i}(\rho)$ .  $\square$

**Proposition 12.** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  be quantum states with finite energy and let  $\{\Pi_n\}_{n=0}^\infty \subset \mathcal{S}(\mathcal{H} \otimes \mathcal{H}^*)$  be a sequence of quantum couplings  $\Pi_n \in \mathcal{C}(\rho_n, \sigma_n)$  converging towards  $\Pi \in \mathcal{C}(\rho, \sigma)$ . Then  $C(\Pi) \leq \liminf_n C(\Pi_n)$ . If moreover  $\lim_n E_{R_i}(\rho_n) = E_{R_i}(\rho)$  and  $\lim_n E_{R_i}(\sigma_n) = E_{R_i}(\sigma)$  for  $i = 1, \dots, N$ , then  $C(\Pi) = \lim_n C(\Pi_n)$ .*

*Proof.* The first inequality follows from Proposition 10 applied to each term in (164). Assuming convergence of the energies of both marginals, it is sufficient to show that  $C(\Pi) \geq \limsup_n C(\Pi_n)$ . To this aim we use (166) and again Proposition 10 to argue that

$$-E_{(R_i \otimes \mathbb{I}_{\mathcal{H}^*} + \mathbb{I}_{\mathcal{H}} \otimes R_i^T)}(\Pi_n) \geq \limsup_{n \rightarrow \infty} -E_{(R_i \otimes \mathbb{I}_{\mathcal{H}^*} + \mathbb{I}_{\mathcal{H}} \otimes R_i^T)}(\Pi_n). \quad (167)$$

$\square$

**Proposition 13.** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  be quantum states with finite energy. Then, there exists  $\Pi \in \mathcal{C}(\rho, \sigma)$  such that  $C(\Pi) = D(\rho, \sigma)^2$ .*



*Proof.* Let  $\Pi_n \in \mathcal{C}(\rho, \sigma)$  be such that  $\lim_n C(\Pi_n) = D(\rho, \sigma)^2$ . By Proposition 8 we can assume, up to extracting a sub-sequence, that  $\{\Pi_n\}_{n=0}^\infty$  converge towards  $\Pi \in \mathcal{C}(\rho, \sigma)$ . By Proposition 12,  $D(\rho, \sigma)^2 \leq C(\Pi) \leq \liminf_n C(\Pi_n) = D(\rho, \sigma)^2$ , hence  $\Pi$  is optimal.  $\square$

**Theorem 9.** *Let  $\rho, \sigma \in \mathcal{S}(\mathcal{H})$  be quantum states with finite energy.*

- *If  $\{\rho_n\}_{n=0}^\infty, \{\sigma_n\}_{n=0}^\infty \subset \mathcal{S}(\mathcal{H})$  have finite energy and converge respectively towards  $\rho, \sigma$ , then*

$$D(\rho, \sigma) \leq \liminf_{n \rightarrow \infty} D(\rho_n, \sigma_n). \quad (168)$$

- *Let  $\{P_n\}_{n=0}^\infty, \{Q_n\}_{n=0}^\infty \subset \mathcal{B}(\mathcal{H})$  be such that  $0 \leq P_n \leq \mathbb{I}_{\mathcal{H}}, 0 \leq Q_n \leq \mathbb{I}_{\mathcal{H}}$ , and  $\lim_n P_n \rho P_n = \rho, \lim_n Q_n \sigma Q_n = \sigma$  in the trace norm. Define, for  $n$  sufficiently large,*

$$\rho_n = \frac{P_n \rho P_n}{\text{Tr}[P_n \rho P_n]}, \quad \sigma_n = \frac{Q_n \sigma Q_n}{\text{Tr}[Q_n \sigma Q_n]}, \quad (169)$$

*and assume that  $\lim_n E_{R_i}(\rho_n) = E_{R_i}(\rho), \lim_n E_{R_i}(\sigma_n) = E_{R_i}(\sigma)$  for  $i = 1, \dots, N$ . Then*

$$D(\rho, \sigma) = \lim_{n \rightarrow \infty} D(\rho_n, \sigma_n). \quad (170)$$

*Proof.* By Proposition 13, choose  $\Pi_n \in \mathcal{C}(\rho_n, \sigma_n)$  such that

$$C(\Pi_n) = D(\rho_n, \sigma_n)^2. \quad (171)$$

By Proposition 8 and Proposition 12 we obtain  $\Pi \in \mathcal{C}(\rho, \sigma)$  such that  $C(\Pi) \leq \liminf_n C(\Pi_n)$ , hence

$$D(\rho, \sigma)^2 \leq C(\Pi) \leq \liminf_n D(\rho_n, \sigma_n)^2. \quad (172)$$

To prove the second statement, let  $\Pi \in \mathcal{C}(\rho, \sigma)$  be such that  $C(\Pi) = D(\rho, \sigma)^2$ . Write

$$\begin{aligned} \tilde{\Pi}_n &= (Q_n \otimes P_n^T) \Pi (Q_n \otimes P_n^T), \\ \tilde{\rho}_n^T &= \text{Tr}_{\mathcal{H}}[\tilde{\Pi}_n] = P_n^T \text{Tr}_{\mathcal{H}}[(Q_n \otimes \mathbb{I}_{\mathcal{H}^*}) \Pi (Q_n \otimes \mathbb{I}_{\mathcal{H}^*})] P_n^T, \\ \tilde{\sigma}_n &= \text{Tr}_{\mathcal{H}^*}[\tilde{\Pi}_n] = Q_n \text{Tr}_{\mathcal{H}^*}[(\mathbb{I}_{\mathcal{H}} \otimes P_n^T) \Pi (\mathbb{I}_{\mathcal{H}} \otimes P_n^T)] Q_n, \\ m_n &= \text{Tr}_{\mathcal{H} \otimes \mathcal{H}^*}[\tilde{\Pi}_n] = \text{Tr}_{\mathcal{H}^*}[\tilde{\rho}_n^T] = \text{Tr}_{\mathcal{H}}[\tilde{\sigma}_n], \end{aligned} \quad (173)$$

and notice that  $\tilde{\rho}_n^T \leq P_n^T \rho^T P_n^T = \rho_n^T$  and  $\tilde{\sigma}_n \leq Q_n \sigma Q_n = \sigma_n$ . We define

$$\Pi_n = m_n \tilde{\Pi}_n + \tilde{\sigma}_n \otimes (\rho_n^T - \tilde{\rho}_n^T) + (\sigma_n - \tilde{\sigma}_n) \otimes \tilde{\rho}_n^T + (\sigma_n - \tilde{\sigma}_n) \otimes (\rho_n^T - \tilde{\rho}_n^T). \quad (174)$$

It holds  $\Pi_n \geq 0$ ,

$$\text{Tr}_{\mathcal{H}}[\Pi_n] = m_n \tilde{\rho}_n^T + m_n (\rho_n^T - \tilde{\rho}_n^T) + (1 - m_n) \tilde{\rho}_n^T + (1 - m_n) (\rho_n^T - \tilde{\rho}_n^T) = \rho_n^T, \quad (175)$$

and similarly  $\text{Tr}_{\mathcal{H}^*}[\Pi_n] = \sigma_n$ , so that  $\Pi_n \in \mathcal{C}(\rho_n, \sigma_n)$ .

We claim that  $\lim_n \Pi_n = \Pi$ . Since  $\lim_n P_n \rho P_n = \rho$  in the trace norm,  $\lim_n \text{Tr}[P_n \rho P_n] = 1$ , so  $\lim_n \rho_n = \rho$ . Similarly,  $\lim_n \sigma_n = \sigma$ . Lemma 5 gives  $\lim_n \tilde{\Pi}_n = \Pi$  in the trace norm and, by continuity of partial trace,  $\lim_n \tilde{\rho}_n = \rho, \lim_n \tilde{\sigma}_n = \sigma$  and  $\lim_n m_n = 1$ , hence the claim is proved.

By Proposition 12, because the energies of the marginals converge, it follows that  $C(\Pi) = \lim_n C(\Pi_n)$ . We conclude that

$$D(\rho, \sigma)^2 = C(\Pi) \geq \limsup_{n \rightarrow \infty} D(\rho_n, \sigma_n)^2. \quad (176)$$

□

*Remark 11.* In Theorem 9 we may choose  $P_n = \sum_{i=0}^n |\psi_i\rangle\langle\psi_i|$  where  $\rho$  has the eigendecomposition  $\rho = \sum_{k=1}^{\infty} p_k |\psi_k\rangle\langle\psi_k|$ , and similarly for  $Q_n$ , by considering an eigendecomposition of  $\sigma$ . Indeed, by Proposition 9 it follows that  $E_{R_i}(\rho_n)$  converge towards  $E_{R_i}(\rho)$ , and similarly for  $\sigma$ . We deduce that

$$\lim_{n \rightarrow \infty} D(\rho_n, \sigma_n) = D(\rho, \sigma). \quad (177)$$

## Appendix B.

**Lemma 7.** *For any  $X \in \mathcal{T}_2(\mathcal{H})$  and any self-adjoint operator  $R$  on  $\mathcal{H}$  such that  $\|[R, X]\|_2 < \infty$  we have*

$$\begin{aligned} \|[R, X]\|_2^2 \geq \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R - \sqrt{X^\dagger X} R \sqrt{X^\dagger X} R \right. \\ \left. - \sqrt{X X^\dagger} R \sqrt{X X^\dagger} R \right]. \end{aligned} \quad (178)$$

*Proof.* Let us consider the singular-value decomposition of  $X$ :

$$X = \sum_{i=0}^{\infty} x_i |\psi_i\rangle\langle\phi_i|, \quad x_i \geq 0, \quad \langle\psi_i|\psi_j\rangle = \langle\phi_i|\phi_j\rangle = \delta_{ij}, \quad (179)$$

where the series converges in the Hilbert-Schmidt norm. We get

$$\begin{aligned} \|[R, X]\|_2^2 &= \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R - 2 X^\dagger R X R \right] \\ &\geq \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R \right] - 2 |\operatorname{Tr}_{\mathcal{H}} [X^\dagger R X R]| \\ &= \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R \right] - 2 \left| \sum_{i,j=0}^{\infty} x_i x_j \langle\psi_i|R|\psi_j\rangle\langle\phi_j|R|\phi_i\rangle \right| \\ &\geq \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R \right] - \sum_{i,j=0}^{\infty} x_i x_j \left( |\langle\psi_i|R|\psi_j\rangle|^2 + |\langle\phi_j|R|\phi_i\rangle|^2 \right) \\ &= \operatorname{Tr}_{\mathcal{H}} \left[ R(X^\dagger X + X X^\dagger) R - \sqrt{X^\dagger X} R \sqrt{X^\dagger X} R - \sqrt{X X^\dagger} R \sqrt{X X^\dagger} R \right], \end{aligned} \quad (180)$$

and the claim follows. □

## References

- [1] C. Villani. *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2008.
- [2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

- [3] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.
- [4] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.
- [5] A. Figalli, F. Maggi, and A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent. Math.*, 182(1):167–211, 2010.
- [6] Karl-Theodor Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl. (9)*, 84(2):149–168, 2005.
- [7] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.
- [8] Luigi Ambrosio, Giuseppe Savaré, and Lorenzo Zambotti. Existence and stability for Fokker-Planck equations with log-concave reference measure. *Probab. Theory Related Fields*, 145(3-4):517–564, 2009.
- [9] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. Scaling hypothesis for the euclidean bipartite matching problem. *Phys. Rev. E*, 90:012118, Jul 2014.
- [10] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [11] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.
- [12] L. Kantorovitch. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [13] V. I. Bogachev. *Measure theory. Vol. I, II*. Springer-Verlag, Berlin, 2007.
- [14] Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987.
- [15] Eric A. Carlen and Jan Maas. An Analog of the 2-Wasserstein Metric in Non-Commutative Probability Under Which the Fermionic Fokker–Planck Equation is Gradient Flow for the Entropy. *Commun. Math. Phys.*, 331(3):887–926, November 2014.
- [16] Eric A. Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance. *J. Funct. Anal.*, 273(5):1810–1869, 2017.
- [17] Eric A Carlen and Jan Maas. Non-commutative calculus, optimal transport and functional inequalities in dissipative quantum systems. *Journal of Statistical Physics*, 178(2):319–378, 2020.
- [18] Cambyse Rouzé and Nilanjana Datta. Concentration of quantum states from quantum functional and transportation cost inequalities. *Journal of Mathematical Physics*, 60(1):012202, 2019.
- [19] Nilanjana Datta and Cambyse Rouzé. Relating relative entropy, optimal transport and fisher information: A quantum hwi inequality. In *Annales Henri Poincaré*, pages 1–36. Springer, 2020.
- [20] Yongxin Chen, Tryphon T. Georgiou, and Allen R. Tannenbaum. Matrix optimal mass transport: A quantum mechanical approach. *IEEE Trans. Automat. Contr.*, 63(8):2612–2619, 2018.
- [21] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Wasserstein geometry of quantum states and optimal transport of matrix-valued measures.

- In *Emerging Applications of Control and Systems Theory*, pages 139–150. Springer, 2018.
- [22] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [23] Giacomo De Palma and Stefan Huber. The conditional Entropy Power Inequality for quantum additive noise channels. *Journal of Mathematical Physics*, 59(12):122201, 2018.
- [24] François Golse, Clément Mouhot, and Thierry Paul. On the mean field and classical limits of quantum mechanics. *Communications in Mathematical Physics*, 343(1):165–205, 2016.
- [25] Emanuele Caglioti, François Golse, and Thierry Paul. Towards optimal transport for quantum densities. working paper or preprint, December 2018.
- [26] François Golse. The quantum N-body problem in the mean-field and semiclassical regime. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2118):20170229, 2018.
- [27] François Golse and Thierry Paul. The Schrödinger equation in the mean-field and semiclassical regime. *Archive for Rational Mechanics and Analysis*, 223(1):57–94, 2017.
- [28] François Golse and Thierry Paul. Wave packets and the quadratic Monge–Kantorovich distance in quantum mechanics. *Comptes Rendus Mathématique*, 356(2):177–197, 2018.
- [29] Emanuele Caglioti, François Golse, and Thierry Paul. Quantum Optimal Transport is Cheaper. *Journal of Statistical Physics*, 2020.
- [30] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Physical review letters*, 121(4):040502, 2018.
- [31] Shouvanik Chakrabarti, Huang Yiming, Tongyang Li, Soheil Feizi, and Xiaodi Wu. Quantum Wasserstein generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 6781–6792. Curran Associates, Inc., 2019.
- [32] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017.
- [33] H.P. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. OUP Oxford, 2007.
- [34] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [35] A.S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction*. Texts and Monographs in Theoretical Physics. De Gruyter, 2019.
- [36] Paolo Gibilisco and Tommaso Isola. Wigner–Yanase information on quantum state space: the geometric approach. *Journal of Mathematical Physics*, 44(9):3752–3762, 2003.
- [37] I. Bengtsson and K. Życzkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2017.
- [38] Samuel L Braunstein and Peter Van Loock. Quantum information with continuous variables. *Reviews of Modern Physics*, 77(2):513, 2005.

- [39] Christian Weedbrook, Stefano Pirandola, Raul Garcia-Patron, Nicolas J Cerf, Timothy C Ralph, Jeffrey H Shapiro, and Seth Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84(2):621, 2012.
- [40] M. Schuld. *Supervised Learning with Quantum Computers*. Quantum science and technology. Springer, 2018.
- [41] Antoine Royer. Wigner function in liouville space: a canonical formalism. *Physical Review A*, 43(1):44, 1991.
- [42] GM D’Ariano, P Lo Presti, and MF Sacchi. Bell measurements and observables. *Physics Letters A*, 272(1-2):32–38, 2000.
- [43] Andreas Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347(1):291–313, 2016.
- [44] Elliott H Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, 19:0–35, 1973.
- [45] Eliakim H Moore. On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26:394–395, 1920.
- [46] A. Ferraro, S. Olivares, and M. Paris. *Gaussian States in Quantum Information*. Napoli series on physics and astrophysics. Bibliopolis, 2005.
- [47] Alessio Serafini. *Quantum Continuous Variables: A Primer of Theoretical Methods*. CRC Press, 2017.
- [48] Alexander Semenovich Holevo. Gaussian optimizers and the additivity problem in quantum information theory. *Russian Mathematical Surveys*, 70(2):331, 2015.
- [49] Giacomo De Palma, Dario Trevisan, Vittorio Giovannetti, and Luigi Ambrosio. Gaussian optimizers for entropic inequalities in quantum information. *Journal of Mathematical Physics*, 59(8):081101, 2018.
- [50] S. Barnett and P.M. Radmore. *Methods in Theoretical Quantum Optics*. Oxford Series in Optical and Imaging Sciences. Clarendon Press, 2002.
- [51] Alexander S Holevo. One-mode quantum Gaussian channels: Structure and quantum capacity. *Problems of Information Transmission*, 43(1):1–11, 2007.
- [52] Giacomo De Palma. The squashed entanglement of the noiseless quantum Gaussian attenuator and amplifier. *Journal of Mathematical Physics*, 60(11):112201, 2019.
- [53] Giacomo De Palma and Dario Trevisan. The conditional entropy power inequality for bosonic quantum systems. *Communications in Mathematical Physics*, 360(2):639–662, Jun 2018.
- [54] Howard Barnum. Quantum rate-distortion coding. *Phys. Rev. A*, 62:042309, Sep 2000.
- [55] Igor Devetak and Toby Berger. Quantum rate-distortion theory for iid sources. In *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No. 01CH37252)*, page 276. IEEE, 2001.
- [56] Igor Devetak and Toby Berger. Quantum rate-distortion theory for memoryless sources. *IEEE Transactions on Information Theory*, 48(6):1580–1589, 2002.

- [57] X. Chen and W. Wang. Entanglement information rate distortion of a quantum gaussian source. *IEEE Transactions on Information Theory*, 54(2):743–748, Feb 2008.
- [58] N. Datta, M. Hsieh, and M. M. Wilde. Quantum Rate Distortion, Reverse Shannon Theorems, and Source-Channel Separation. *IEEE Transactions on Information Theory*, 59(1):615–630, Jan 2013.
- [59] Nilanjana Datta, Min-Hsiu Hsieh, Mark M. Wilde, and Andreas Winter. Quantum-to-classical rate distortion coding. *Journal of Mathematical Physics*, 54(4):042201, 2013.
- [60] Mark M Wilde, Nilanjana Datta, Min-Hsiu Hsieh, and Andreas Winter. Quantum rate-distortion coding with auxiliary resources. *IEEE Transactions on Information Theory*, 59(10):6755–6773, 2013.
- [61] Sina Salek, Daniela Cadamuro, Philipp Kammerlander, and Karoline Wiesner. Quantum rate-distortion coding of relevant information. *IEEE Transactions on Information Theory*, 65(4):2603–2613, 2018.
- [62] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. A Wiley-Interscience publication. Wiley, 2006.
- [63] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 214–223, 2017.
- [64] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. The quantum Wasserstein distance of order 1. *arXiv preprint arXiv:2009.04469*, 2020.
- [65] Bobak Kiani, Giacomo De Palma, Milad Marvian, Zi-Wen Liu, and Seth Lloyd. Quantum Earth Mover’s Distance: A New Approach to Learning Quantum Data. *arXiv preprint arXiv:2101.03037*, 2021.
- [66] M. Reed and B. Simon. *Methods of Modern Mathematical Physics: Functional analysis*. Number v. 1 in Methods of Modern Mathematical Physics. Academic Press, 1980.
- [67] V.I. Bogachev. *Measure Theory*. Number v. 1 in Measure Theory. Springer Berlin Heidelberg, 2007.

Giacomo De Palma  
Massachusetts Institute of Technology  
Research Laboratory of Electronics and Department of Mechanical Engineering  
77 Massachusetts Avenue  
02139 Cambridge MA  
USA  
e-mail: [gdepalma@mit.edu](mailto:gdepalma@mit.edu)

Dario Trevisan  
Università degli Studi di Pisa  
Dipartimento di Matematica  
56127 Pisa  
Italy  
e-mail: [dario.trevisan@unipi.it](mailto:dario.trevisan@unipi.it)