

# Geometry of Grassmannians and optimal transport of quantum states

Paolo Antonini\* and Fabio Cavalletti†

## Abstract

Let  $\mathbf{H}$  be a separable Hilbert space. Building on the metric geometry of the Grassmannian  $\mathbb{P}_c(\mathbf{H})$  of finite dimensional subspaces of  $\mathbf{H}$ , we develop a theory of optimal transport for the normal states of the von Neumann algebra of linear and bounded operators  $\mathbf{B}(\mathbf{H})$ . Seeing density matrices as discrete probability measures on  $\mathbb{P}_c(\mathbf{H})$  (via the spectral theorem) we define an optimal transport cost and the Wasserstein distance for normal states. We prove that the transport cost induces the  $w^*$ -topology and satisfies the triangular inequality on a dense subset of normal states.

Our construction is compatible with the quantum mechanics approach of composite systems as tensor products. We provide a natural interpretation of the pure normal states of  $\mathbf{B}(\mathbf{H} \otimes \mathbf{H})$  as families of transport maps. In this way we assign a Wasserstein cost to each pure normal state of  $\mathbf{B}(\mathbf{H} \otimes \mathbf{H})$  consistently with the transport cost defined via  $\mathbb{P}_c(\mathbf{H})$ .

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
<b>3</b>	<b>Spectral-projections measures</b>	<b>14</b>
<b>4</b>	<b>The Wasserstein distance between normal states</b>	<b>18</b>
<b>5</b>	<b>Kantorovich duality for <math>W_p</math> and consequences</b>	<b>23</b>
<b>6</b>	<b>Tensor product interpretation: a generalization</b>	<b>28</b>
<b>A</b>	<b>Homogeneous spaces and principal bundles</b>	<b>30</b>
<b>B</b>	<b>Some basic facts in operator theory</b>	<b>31</b>

## 1 Introduction

In this paper we develop a basic theory of Optimal transport for the  $C^*$ -algebra of compact operators acting on a separable Hilbert space. The construction will be based on the metric geometry of the Grassmannian of all the finite dimensional subspaces of the Hilbert space.

The general optimal transport problem is simple to state: a certain amount of mass (a probability measure) has to be moved to a different configuration in the most efficient way (minimizing a cost). Though so immediate to formulate, this problem has generated a number of deep results furnishing an arsenal of powerful techniques suitable to be applied to a vastity of different areas in mathematics. Let's briefly recall the classical setting. Consider two probability measures  $\sigma_1$  and  $\sigma_2$  over a Polish space  $X$ . A transport plan consists in a probability measure  $\pi$  over  $X \times X$  having  $\sigma_i$  as marginal laws. Denoting with  $\Pi(\sigma_1, \sigma_2)$  the set of transport plans, one looks for the infimum

$$\inf_{\pi \in \Pi(\sigma_1, \sigma_2)} \int_{X \times X} c(x, y) \pi(dx, dy),$$

---

\*Dipartimento di Matematica e Fisica “E. De Giorgi”, Università del Salento, Lecce, Italia, email: paolo.antonini@unisalento.it

†Dipartimento di Matematica “Federigo Enriques”, Università di Milano, Milano, Italia, email: fabio.cavalletti@unimi.it

where the cost function  $c : X \times X \rightarrow \mathbb{R}^+$  measures how expensive is the movement of one unit of mass from  $x$  to  $y$ . A natural choice is  $c = d^p$ , where  $d$  is the distance in  $X$  and  $p \geq 1$  is a real exponent. In this case, the resulting infimum (provided only measures with finite  $p$ -moments are considered) is the  $p$ -th power of a distance  $W_p(\sigma_1, \sigma_2)$  on the space of all the probability measures on  $X$ . This is called the  $p$ -Wasserstein distance.

It is an important natural problem to find the appropriate generalization of the optimal transport scheme together with the Wasserstein distance to the *quantum world*. This is meant in general sense: we are given a  $C^*$ -algebra of operators with its space of states (the probability measures) and we look for a generalization of the optimal transport theory. The resulting Wasserstein distance should have some special properties tailored on the applications in mind. However, it is in general highly non trivial, to produce a Wasserstein distance retaining even the most characterizing feature of a distance function, the triangular inequality.

Before going into details explaining our work we are going to give a brief description of the (large) literature on the subject.

One of the first constructions goes back to the work of Slomczynski and Zyczkowski [50, 51] where a Monge–Wasserstein distance is defined using (generalized) coherent states and the Husimi representation of quantum states. Their metric satisfies an important semiclassical property.

Carlen and Maas [17, 18, 19], used the noncommutative differential calculus to build a 2–Wasserstein quantum distance in the spirit of the dynamical Benamou–Brenier approach [11] to optimal transport. Such metric is induced by a Riemannian structure on the manifold of all the quantum states and enjoys the notable property of making the heat semigroup the gradient flow of the relative von Neumann entropy. Moreover this metric is related to the rate of convergence of the quantum Ornstein–Uhlenbeck semigroup [19]. We also mention that a definition of a quantum Wasserstein distance together with a detailed study, in dimension 2, of all the properties that such a quantum generalization should satisfy can be found in [1].

Motivated by the study of the mean field and semiclassical limit of quantum mechanics, Golse, Mouhot and Paul [36], have defined a quantum analog of the 2–Wasserstein distance relying on the notion of *couplings* seen as density operators acting on the tensor product Hilbert space. As shown in [16], the resulting cost is cheaper than the classical one in the case of the matching problem for two particles with different masses.

Starting from the distance defined by Golse et al., De Palma and Trevisan [27] have recently defined a quantum Wasserstein distance considering transport plans that are in one to one correspondence with the *quantum channels*, the quantum analog of the classical stochastic maps. In [28], a quantum 1–Wasserstein distance as the quantum version of the Hamming distance is also proposed. In the framework of spectral triples considered as noncommutative manifolds, where a noncommutative algebra  $\mathcal{A}$  interacts with a Dirac operator, Connes [24] defined a 1–Wasserstein distance on the space of states of  $\mathcal{A}$ . This is thought as the dual distance in the spirit of Monge–Kantorovich, defined in terms of Lipschitz functions (or their noncommutative analog). Connes’ distance and the Kantorovich duality have been the subject of many works by Rieffel, D’Andrea, Martinetti and collaborators. We refer in a non exhaustive way to the papers [49, 29, 30] and the references therein.

In the realm of free probability, Biane and Voiculescu defined an analog of the Wasserstein distance on the space of the trace-states of a  $C^*$ -algebra [12]. Their metric extends the classical Wasserstein metric.

Other approaches that are worth mentioning include: the one of Wirth [59], based on the non-commutative Dirichlet forms of Cipriani and Sauvageot [21] and the work of Hornshow [39].

Our contribution goes in the direction of studying the static formulation of the optimal transport problem between quantum states.

Let  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  denote the convex set of normal states of  $\mathbf{B}(\mathbf{H})$ , the von Neumann algebra of linear bounded operators on  $\mathbf{H}$ . For more details we refer to Section 2.2. Any such state  $\varphi$  is identified with its density matrix  $\rho_\varphi$  satisfying

$$\rho_\varphi^* = \rho_\varphi, \quad \rho_\varphi \geq 0 \quad \text{and} \quad \text{tr}(\rho_\varphi) = 1.$$

We introduce a distance between density matrices relying on the optimal transport problem between probability measures over the Grassmanian of  $\mathbf{H}$ . The latter is denoted by  $\mathbf{P} = \mathbf{P}(\mathbf{H})$  and defined as the collection of all orthogonal projections of  $\mathbf{H}$ . Its connected components are labelled by the dimension of the ranges of the projections.

The map between density matrices and non-negative measures over  $\mathbf{P}$  is induced by the Spectral Theorem (for selfadjoint and compact operators) providing a natural correspondence:

$$\rho_\varphi = \sum_i \lambda_i P_{V_i} \quad \implies \quad \mu_\varphi := \sum_i \lambda_i \delta_{P_{V_i}}.$$

Here  $P_{V_i}$  stands for the orthogonal projection with range the finite dimensional eigenspace  $V_i$  with eigenvalue  $\lambda_i > 0$ . The spectral decomposition is understood without repetitions. Notice that the projection onto the kernel does not belong to the support of the associate measure  $\mu_\varphi$ . Since  $\text{tr}(\rho_\varphi) = 1$ , it follows that  $\text{tr}(\cdot)\mu_\varphi$  is a probability measure over the Polish space  $\mathbb{P}_c = \mathbb{P}_c(\mathbb{H})$ , the submanifold of  $\mathbb{P}$  of finite rank orthogonal projections.

The Polish structure of  $\mathbb{P}_c$ , making it amenable to standard measure theory techniques, is the one inherited as a Finsler submanifold of  $\mathbb{B}(\mathbb{H})$ . However  $\mathbb{P}_c$  admits a more convenient geometric structure induced infinitesimally by viewing  $\mathbb{P}_c$  as a submanifold of the space of the Hilbert–Schmidt operators. It follows that each connected component (where the trace is constant) of  $\mathbb{P}_c$  is an Alexandrov space of non-negative curvature. This fact puts the investigation of the geometric relations between normal states and optimal transport in a natural setting, see Section 2.1.

Denoting by  $\mathbf{d}$  the geodesic distance of  $\mathbb{P}_c$  as a submanifold of the Hilbert–Schmidt operators, for any  $p \geq 1$ , the  $p$ –Wasserstein distance between two normal states  $\varphi, \psi$  can be then defined as the  $p$ –Wasserstein distance between the normalized spectral measures  $\mu_\varphi$  and  $\mu_\psi$ :

$$W_p(\varphi, \psi) := W_p^{\mathbb{P}_c}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi). \quad (1.1)$$

In this formula  $W_p^{\mathbb{P}_c}$  denotes the classical  $p$ –Wasserstein distance defined over the Polish space  $(\mathbb{P}_c, \mathbf{d})$ . Because of the presence of different connected components,  $W_p(\cdot, \cdot)$  might easily become infinite making  $W_p$  an extended distance.

We overcome this issue by considering a larger family of discrete measures representing density matrices. In particular for each normal state  $\varphi$  we consider the set  $\Lambda_\varphi^\perp$  of discrete measures  $\mu = \sum_i \lambda_i \delta_{P_i}$  with  $\lambda_i \geq 0$  such that  $\rho_\varphi = \sum_i \lambda_i P_i$  and  $P_i \perp P_j$  whenever  $i \neq j$ . In contrast with the representations considered before the eigenvalues now admit repetitions. Then the natural extension of  $W_p(\varphi, \psi)$  is obtained by defining the *cost* between  $\varphi$  and  $\psi$  as

$$\mathcal{C}_p(\varphi, \psi) := \inf_{\substack{\mu_0 \in \Lambda_\varphi^\perp \\ \mu_1 \in \Lambda_\psi^\perp}} W_p^{\mathbb{P}_c}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1), \quad (1.2)$$

the Wasserstein distance between the two (compact) sets of the various associated measures representing the states. We prove that  $\mathcal{C}_p$  has the following main properties.

**Existence of optimal configurations:** for any couple of normal states  $\varphi$  and  $\psi$ , the infimum in (1.2) can be replaced by the minimum (Proposition 3.5). Moreover optimal couplings always exist (Proposition 4.9).

**Projections of dimension 1:** The optimal configurations  $\mu_0, \mu_1$  can always be taken with support contained inside the connected component  $\mathbb{P}_1$ , the space of projections with one dimensional rank (Proposition 2.6). This is nothing but  $\mathbb{P}(\mathbb{H})$  the projective space of  $\mathbb{H}$ , the space of the pure states of the  $C^*$ -algebra of the compact operators  $\mathbb{K}$ .

**Topology:**  $\mathcal{C}_p$  is a semi-distance inducing the weak topology over  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  (Theorem 4.10).

**Triangular inequality:**  $\mathcal{C}_p$  verifies the triangular inequality when restricted to the open dense subset of  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  of those density matrices with only simple eigenvalues (Remark 4.12).

We also obtain a duality formula for  $W_p$  with the Kantorovich potentials represented by densely defined operators (Theorem 5.4 and Corollary 5.6). Relying on the geodesic structure of  $(\mathbb{P}_c, \mathbf{d})$ , we finally study, in Section 5.2, the  $W_p$ -geodesics of  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$ .

In the last section we study tensor product Hilbert spaces  $\mathbb{H} \otimes \mathbb{H}$  corresponding in quantum mechanics to composite systems. A natural way to match two normal states  $\varphi, \psi$  of  $\mathbb{B}(\mathbb{H})$  would be via a normal state  $\Xi \in \mathcal{S}_n(\mathbb{B}(\mathbb{H} \otimes \mathbb{H}))$  satisfying the partial trace conditions  $J_b^1 \Xi = \varphi$  and  $J_b^2 \Xi = \psi$  (for the notation see Section 2.2.2). In Section 6 we reconcile this point of view with the one presented in Section 4. In particular we prove the following (Theorem 6.3).

**Pure normal states of the tensor product as natural families of transport plans:** given any element  $\omega_\zeta$  of  $\mathcal{P}\mathcal{S}_n(\mathbb{B}(\mathbb{H} \otimes \mathbb{H}))$ , i.e. any pure normal state of  $\mathbb{B}(\mathbb{H} \otimes \mathbb{H})$  with partial traces  $\varphi$  and  $\psi$ , we associate a family of admissible transport plans between admissible representations of  $\varphi$  and  $\psi$ . In particular this permits to assign a well-defined optimal transport cost to any pure normal state of  $\mathbb{B}(\mathbb{H} \otimes \mathbb{H})$  (Remark 6.5).

We tried to keep the presentation as self-contained as possible by including in Section 2 some known facts used in the paper. In particular Section 2.1 collects, and in some cases re-proves, many of the known geometric properties of  $\mathbb{P}_c$  that were distributed through different references with often different notations and settings.

**Acknowledgements.** The authors wish to thank Antonio Lerario for a number of interesting discussions.

## 1.1 Notations

In this paper we will switch freely from the standard notation for vectors in a Hilbert space to the Dirac notation with Bra and Kets. In particular we consider the inner product  $\langle \cdot, \cdot \rangle$  or  $\langle \cdot | \cdot \rangle$  with the *Physicists convention*: antilinear in the first entry. For a linear operator  $T$  on a vector space we denote  $N(T)$  for its Kernel and  $R(T)$  for its image.

Projection means orthogonal projection i.e.  $P = P^*$  and  $P^2 = P$ ; when  $P^2 = P$  we say idempotent. Also for two projections we write  $Q \leq P$  if and only if  $QH \subset PH$ . This is equivalent to  $PQ = Q$  or  $QP = Q$ .

## 2 Preliminaries

### 2.1 Geometry of the space of projections

Let us fix an Hilbert space  $H$ ; then  $B(H)$  will be the space of bounded linear operators in  $H$  and  $B_h(H)$  the subspace of the self-adjoint ones (Hermitian); also denote by  $B_{sa}(H) = \{X \in B(H) : X = -X^*\}$  the skew adjoints.

The Grassmannian of  $H$ , denoted with  $P$  is the space of all the projections:

$$P = \{P \in B(H) : P = P^* \text{ and } P^2 = P\}.$$

We describe its geometry mainly following [5, 6, 7, 26, 55]. Fundamental is the natural action of the unitary group  $U(H)$  by conjugation  $g \cdot P = gPg^*$  for  $g \in U(H)$ .

We recall here few but important facts about the group  $U(H)$ . This is a Banach–Lie group, closed inside  $B(H)$  with Lie algebra identified with the skew adjoint operators  $\mathfrak{u} := T_1U(H) = B_{sa}(H)$  having the operators commutator as Lie bracket. The exponential map  $\exp : \mathfrak{u} \rightarrow U(H)$  is the operators exponentiation. It is surjective because in  $B(H)$  we may form Borel functions of normal operators; this gives a logarithm for every skew-adjoint operator.

All the curves in the form

$$[-1, 1] \ni t \mapsto ue^{itX} \in U(H)$$

with  $X \in B_h(H)$  i.e. the translations of one parameter groups are called the *group geodesics* of  $U(H)$ . The name is legitimated by the fact that we can find a natural class of linear connections on  $U(H)$  creating such geodesics. Moreover one can show that these are minimal curves inside  $U(H)$  with respect to the natural Finsler structure inherited by the embedding  $U(H) \subset B(H)$  (see [3]). We are now ready to discuss the geometry of  $P$ .

1. **Manifold structure.**  $P$  is a submanifold of  $B_h(H)$  with complemented tangent. Its tangent space at  $P$ , as a submanifold is naturally identified in the following way:

$$T_P P = \{Y \in B_h(H) : PY + YP = Y\}; \quad (2.1)$$

or equivalently with all the selfadjoint operators  $Y$  satisfying  $PYP = (1 - P)Y(1 - P) = 0$ . Indeed  $P$  induces a block decomposition for the whole  $B_h(H)$

$$A \mapsto \begin{pmatrix} PAP & PA(1 - P) \\ (1 - P)AP & (1 - P)A(1 - P) \end{pmatrix}, \quad (2.2)$$

so that can give the following.

**Definition 2.1.** The selfadjoint operators which are off-diagonal in the decomposition (2.2) are called *co-diagonal* with respect to  $P$ . The space of all the co-diagonal operators with respect to  $P$  is denoted by  $\mathcal{C}_P$ .

In symbols

$$\mathcal{C}_P = \{Y \in B_h(H) : PY + YP = Y\}. \quad (2.3)$$

Let us prove the (2.1). The first inclusion comes from differentiating the relation  $\gamma^2(t) = \gamma(t)$  for a smooth curve in  $P$  with  $\gamma(0) = P$ . For the reversed inclusion we make use of (2.3) and we observe first that any  $X \in \mathcal{C}_P$  satisfies  $X = [[X, P], P]$ . This also means (every commutator with  $P$  is codiagonal) that  $\mathcal{C}_P = \{i[X, P] : X \in B_h(H)\}$ . Now if  $X$  is codiagonal,  $e^{t[X, P]}$  is a one parameter group of unitaries ( $[X, P]$  is skew-adjoint) and the path  $\gamma(t) = e^{t[X, P]}Pe^{-t[X, P]}$  satisfies  $\dot{\gamma}(0) = [[X, P], P] = X$ .

We will see later that curves in the form of  $\gamma$  are exactly the geodesics through  $P$  with respect to a family of natural connections. Summing up:

$$T_P P = \mathcal{C}_P = \{i[X, P] : X \in B_h(H)\}.$$

If we denote by  $\mathcal{D}_P$  the selfadjoint operators which are diagonal in the decomposition (2.2) we have a linear splitting

$$B_h(H) = \mathcal{C}_P \oplus \mathcal{D}_P. \quad (2.4)$$

## 2. Homogeneous space structure of the connected components.

The  $U(\mathbb{H})$  action on  $\mathbb{P}$  is locally transitive for if  $\|P - Q\| < 1$  then  $Q = g \cdot P$  for some unitary  $g$ . Using this fact one shows that  $P$  and  $Q$  are in the same connected component if and only if there exists a path of unitaries  $g_t$  with  $g_0 = 1$  and  $P = g_1 Q g_1^*$  (a proof in [60, Corollary 5.2.9]). In other words the  $U(\mathbb{H})$ -orbits, i.e. the conjugacy classes are the connected components in  $\mathbb{P}$  :

$$\mathcal{O}_P := \{gPg^* : g \in U(\mathbb{H})\} = \text{connected component of } P.$$

These connected components are easily found:  $P$  and  $Q$  are connected iff  $\dim N(P) = \dim N(Q)$  and  $\dim R(P) = \dim R(Q)$ .

Let's now fix a reference point  $P \in \mathbb{P}$  (for the rest of this section). The stabiliser  $I_P = \{g : g \cdot P = P\}$  coincides with the subgroup  $\{g \in U(\mathbb{H}) : [g, P] = 0\}$  and the quotient  $U(\mathbb{H})/I_P$  is diffeomorphic to  $\mathcal{O}_P$ . More precisely, with the canonical projection

$$U(\mathbb{H}) \longrightarrow U(\mathbb{H})/I_P \cong \mathcal{O}_P, \quad (2.5)$$

we have a principal bundle with equivariant projection. In other words  $\mathcal{O}_P$  is an homogeneous space [5, Proposition 2.2].

The decomposition diagonal/codiagonal (2.4) defines on the principal bundle (2.5) a canonical connection (indeed the homogeneous space structure is *reductive*). This connection induces in the customary way a notion of parallel translation, covariant derivative and geodesics for  $\mathbb{P}$ . We don't construct them explicitly here because we will consider in a while, a second, more direct connection on  $T\mathbb{P}$  sharing the same geodesics.

3. **Connection on  $T\mathbb{P}$ .** To any  $X \in \mathbb{B}_h(\mathbb{H})$  we can associate its co-diagonal part with respect to  $P$  using the projection onto the codiagonals

$$E_P : \mathbb{B}_h(\mathbb{H}) \longrightarrow T_P\mathbb{P}, \quad E_P(X) := PX(1 - P) + (1 - P)XP. \quad (2.6)$$

This induces a connection (in the usual sense) on  $T\mathbb{P}$ . If  $X$  is a tangent field (i.e.  $X : \mathbb{P} \rightarrow \mathbb{B}_h(\mathbb{H})$  with  $X(P) \in T_P\mathbb{P}$  for every  $P$ ) and  $\gamma : I \rightarrow \mathbb{P}$  is a curve, then  $X \circ \gamma$  is a vector field along  $\gamma$  with covariant derivative

$$\frac{DX}{dt} = E_{\gamma(t)} \left( \frac{d}{dt} X(\gamma(t)) \right). \quad (2.7)$$

4. **Geodesics.** A curve  $\gamma : I \rightarrow \mathbb{P}$  is a geodesic if, by definition

$$\frac{D\dot{\gamma}}{dt} = 0, \quad \forall t \in I.$$

All the geodesics starting at  $P \in \mathbb{P}$  are in the form  $\gamma(t) = e^{itZ} P e^{-itZ}$  with  $Z \in T_P(\mathbb{P})$  [6, 26]. As anticipated we can prove that these are also all the geodesics with respect to the connection induced by the natural connection in  $\mathbb{P}$  as an homogeneous reductive space.

To check that the geodesic equation is satisfied for  $\gamma(t) = e^{itZ} P e^{-itZ}$  we take the opportunity to discuss the manifold of symmetries  $\mathbb{S} := \{S \in \mathbb{B}_h(\mathbb{H}) : S^2 = 1\}$ , diffeomorphic to  $\mathbb{P}$  via the map

$$\mathcal{F} : \mathbb{P} \longrightarrow \mathbb{S}, \quad P \longmapsto 2P - 1. \quad (2.8)$$

The tangent space at  $S \in \mathbb{S}$  consists in all the self-adjoint  $X \in \mathbb{B}_h(\mathbb{H})$  anticommuteing with  $S$  i.e.

$$T_S\mathbb{S} = \left\{ X \in \mathbb{B}_h(\mathbb{H}) : SX + XS = 0 \right\}.$$

We have a corresponding projection on the tangent space taking the form

$$\text{Pr}_S : \mathbb{B}_h(\mathbb{H}) \longrightarrow \mathbb{B}_h(\mathbb{H}), \quad \text{Pr}_S(Z) = (1 - P)ZP + PZ(1 - P); \quad 2P - 1 = S,$$

also inducing a connection on  $\mathbb{S}$ . This is given by the same formula as (2.6). On the other hand the map  $\mathcal{F} : \mathbb{P} \rightarrow \mathbb{S}$  is compatible with the two connections on the domain and target thus sending a geodesic to a geodesic. In fact  $\mathcal{F}$  is the restriction of a map defined on the whole of  $\mathbb{B}_h(\mathbb{H})$  and its differential  $d_P\mathcal{F}(X) = 2X$  intertwines the two projections onto  $\mathbb{P}$  and  $\mathbb{S}$ .

Now thanks to the inclusion  $\mathbb{S} \subset U(\mathbb{H})$  some formulas simplify when passing to  $\mathbb{S}$ . Start with the curve  $\gamma(t) = e^{itZ} P e^{-itZ}$  in  $\mathbb{P}$  with  $Z \in T_P\mathbb{P}$ . Since  $Z$  is  $P$ -codiagonal, it anticommutes with  $S = \mathcal{F}(P)$  so that  $e^{itZ} \mathcal{F}(P) = \mathcal{F}(P) e^{-itZ}$ . We can now transform  $\gamma$  under  $\mathcal{F}$ :

$$\mathcal{F}(e^{itZ} P e^{-itZ}) = e^{itZ} \mathcal{F}(P) e^{-itZ} = \mathcal{F}(P) e^{-2itZ} = \mathcal{F}(P) e^{-itd\mathcal{F}(Z)}.$$

It is immediate to check that this is a geodesic in  $\mathbf{S}$  and by the properties of  $\mathcal{F}$  we see that  $\gamma$  is a geodesic too. Moreover  $\mathcal{F}(\gamma)$  is also a geodesic in  $\mathbf{U}(\mathbf{H})$  (a traslation of a one parameter group). In other words  $\mathbf{S}$  is totally geodesic inside  $\mathbf{U}(\mathbf{H})$ .

Put  $Y := -iSZ/2 \in T_S\mathbf{S}$  then the geodesic in  $\mathbf{S}$  can also be written as  $t \mapsto e^{tXS/2}Se^{-tXS/2}$ . Indeed the exponential map is the restriction of the family of analytic mappings

$$\mathbf{B}(\mathbf{H}) \longrightarrow \mathbf{B}(\mathbf{H}), \quad Z \longmapsto e^{ZS/2}Se^{-ZS/2}.$$

The exponential map for  $\mathbf{P}$  follows using  $\mathcal{F}$ . We note also the formula  $\frac{d}{dt} e^{tXS/2}Se^{-tXS/2} = e^{tXS/2}Xe^{-tXS/2}$ .

### 2.1.1 Metric aspects

The Grassmannian  $\mathbf{P}$  has a natural non-smooth reversible Finsler structure induced by the operator norm via the embedding  $\mathbf{P} \subset \mathbf{B}_h(\mathbf{H})$ . However the submanifold

$$\mathbf{P}_c := \mathbf{P} \cap \mathbb{K},$$

of the compact and then finite rank projections is contained in the Hilbert space  $\mathbf{HS}(\mathbf{H})$  of the (selfadjoint) Hilbert–Schmidt operators with metric  $(A, B) \mapsto \Re \operatorname{tr}(A^*B)$  and inherits a riemannian structure. Any point  $P \in \mathbf{P}_c$  is finite rank so that the co-diagonal operators at  $P$  are finite rank too and we have the induced metric<sup>(a)</sup>

$$g(X, Y) := \operatorname{tr}(XY), \quad X, Y \in T_P\mathbf{P}_c,$$

generalising the familiar riemannian (Kähler) structure on the finite dimensional Grassmann manifold. We summarise some of its basic properties (see [8, 40]) :

- the topology on  $\mathbf{P}_c$  induced by the embedding  $\mathbf{P}_c \subset \mathbf{B}_h(\mathbf{H})$  where  $\mathbf{B}_h(\mathbf{H})$  is given with the norm topology coincides the topology induced by the embedding  $\mathbf{P}_c \subset \mathbf{HS}(\mathbf{H})$ . This is clear for if  $T$  and  $S$  are finite rank operators with range of dimension at most  $n$  then:

$$\|T - S\| \leq \|T - S\|_2 \leq \sqrt{2n} \|T - S\|$$

with  $\|\cdot\|_2$  the Hilbert–Schmidt norm.

- The connection (2.7) is exactly the Levi–Civita connection. We can compute an explicit formula following [32]. We have orthogonal projections on the tangent space and on the normal space to  $\mathbf{P}_c$  and the theory of submanifolds presents no differences with the finite dimensional case. The orthogonal projection is exactly the projection on the codiagonals that we have already used.

Now let  $P \in \mathbf{P}_c$  and  $X, Y$  vector fields tangent to  $\mathbf{P}_c$ ; if we denote with  $D_X Y$  the covariant derivative in the flat space  $\mathbf{HS}(\mathbf{H})$ , we have at  $P$ :

$$D_X Y = (PD_X Y(1 - P) + (1 - P)D_X YP) + (XY + YX)(1 - 2P).$$

The first addendum is tangential to  $\mathbf{P}_c$  while the second one is normal. Therefore

$$\nabla_X Y = PD_X Y(1 - P) + (1 - P)D_X YP, \quad \text{the connection of } \mathbf{P}_c \text{ at } P,$$

$$\sigma(X, Y) = (XY + YX)(1 - 2P) \quad \text{the second fundamental form at } P.$$

- The geodesics that we have already discussed are geodesics for the metric in  $\mathbf{P}_c$  too. In particular  $t \mapsto e^{t[X, P]}Pe^{-t[X, P]}$  is the unique geodesic starting from  $P$  with initial speed  $X$ .
- The curvature tensor is

$$R(X, Y)Z = [[X, Y], Z], \quad X, Y, Z \in T_P\mathbf{P}_c$$

as follows immediately from the Gauss formula (the ambient space is flat)

$$\langle R(X, Y)Z, W \rangle = \langle \sigma(X, W), \sigma(Y, Z) \rangle - \langle \sigma(X, Z), \sigma(Y, W) \rangle.$$

From the Cauchy–Schwartz inequality it follows the sectional curvature is non negative.

- The length of a smooth or Lipschitz, curve  $\gamma : I \longrightarrow \mathbf{P}_c$  is defined by  $L(\gamma) = \int_I \|\dot{\gamma}\| dt$ . The geodesic distance  $\mathbf{d}$  follows by minimisation over all the paths. If  $P$  and  $Q$  satisfy  $\mathbf{d}(P, Q) < \pi/2$ , they are joined by a unique geodesic with length  $L(\gamma) = \mathbf{d}(P, Q)$ . The metric space  $(\mathbf{P}_c, \mathbf{d})$  is complete. It follows that (for  $\mathbf{H}$  separable) it is Polish.

<sup>(a)</sup>since the operators are codiagonal the trace of  $XY$  is real valued

To describe in more details the geometry of  $P_c$  it is useful to follow the techniques of [37] presented in the real case. The extension to our, complex case, is straightforward as we will show in the following.

To start with, we depict  $P_c$  as the base of a second principal bundle with fiber  $U_r$  (a compact Lie group). Firstly we introduce a notation for the connected components of  $P_c$

$$P_r := \{P \in P_c : \dim R(P) = r\}. \quad (2.9)$$

Keeping the rank  $r$  fixed, let  $\text{St}(r, \mathbf{H})$  be the (complex) Stiefel manifold. It is the manifold of all the Hilbert space embeddings  $\varphi : \mathbb{C}^r \rightarrow \mathbf{H}$ . Thus  $\varphi^* \varphi = \text{Id}_r$ . Any  $\varphi \in \text{St}(r, \mathbf{H})$  is specified by a collection of  $r$ -orthonormal vectors in  $\mathbf{H}$ , the columns of the finite dimensional matrix of  $\varphi$ . We have in this way a natural embedding

$$\text{St}(r, \mathbf{H}) \subset \underbrace{\mathbf{H} \times \cdots \times \mathbf{H}}_{r \text{ times}} \quad (2.10)$$

with tangent space

$$T_\varphi \text{St}(r, \mathbf{H}) = \{X \in \mathbf{B}(\mathbb{C}^r, \mathbf{H}) : X^* \varphi + \varphi^* X = 0\}.$$

This is the space of the linear maps  $X : \mathbb{C}^r \rightarrow \mathbf{H}$  such that  $X^* \varphi$  is skew-adjoint. Indeed the inclusion  $\subset$  is straightforward. To see the second one first solve the o.d.e.  $\frac{d}{dt}(\gamma^* \gamma) = \dot{\gamma}^* \gamma + \gamma^* \dot{\gamma} = 0$  in the space  $\mathbf{B}(\mathbb{C}^r, \mathbf{H})$  of the finite rank maps with initial data satisfying:  $\gamma(0) = \varphi \in \text{St}(r, \mathbf{H})$ ,  $\dot{\gamma}(0) = X$  with  $X^* \varphi + \varphi^* X = 0$ . It follows  $\gamma(t) \in \text{St}(r, \mathbf{H})$ .

The embedding (2.10) induces a riemannian metric on the Stiefel manifold:  $(X, Y) \mapsto \Re \text{tr}(X^* Y)$  for  $X, Y \in T_\varphi \text{St}(r, \mathbf{H})$  and we shall consider its rescaled version

$$g(X, Y) := 2\Re \text{tr}(X^* Y) \quad X, Y \in T_\varphi \text{St}(r, \mathbf{H}).$$

We compute the orthogonal projection on the tangent space of  $\text{St}(r, \mathbf{H})$ . In fact the orthogonal decomposition

$$\mathbf{H}^r \cong \mathbf{B}(\mathbb{C}^r, \mathbf{H}) = T_\varphi \text{St}(r, \mathbf{H}) \oplus N_\varphi \text{St}(r, \mathbf{H})$$

at  $\varphi$  is obtained combining the decomposition

$$\mathbf{H} = R(\varphi) \oplus R(\varphi)^\perp \quad (2.11)$$

induced by the projection  $\varphi \varphi^*$  together with the orthogonal decomposition in  $\mathbf{B}(\mathbb{C}^r)$  by Hermitian and Skew-Hermitian matrices (with projections denoted by  $\text{He}$  and  $\text{Sk}$ ). For any vector  $X \in \mathbf{B}(\mathbb{C}^r, \mathbf{H})$  we write

$$X = \varphi \varphi^* X + (1 - \varphi \varphi^*) X = [\varphi (\text{Sk} \varphi^* X) + (1 - \varphi \varphi^*) X] + \varphi (\text{He} \varphi^* X).$$

It is easy to check that these are respectively the tangent and normal components with:  $X \mapsto \varphi (\text{Sk} \varphi^* X) + (1 - \varphi \varphi^*) X$  the tangent projection and  $X \mapsto \varphi (\text{He} \varphi^* X)$  the normal one. In particular we see that  $N_\varphi \text{St}(r, \mathbf{H}) = \{\varphi S : S \in \mathbf{B}(\mathbb{C}^r), S = S^*\}$ .

There are two commuting left and right actions

$$U(\mathbf{H}) \curvearrowright \text{St}(r, \mathbf{H}) \curvearrowleft U_r = U(\mathbb{C}^r),$$

corresponding to post and pre composition

$$u \cdot \varphi := u \circ \varphi \quad \text{and} \quad \varphi \cdot g := \varphi \circ g, \quad u \in U(\mathbf{H}), g \in U_r.$$

The  $U(\mathbf{H})$  action is transitive while the  $U_r$  one is free. Two points  $\varphi$  and  $\psi$  are in the same  $U_r$ -orbit if and only if they have the same range. It follows the quotient is  $P_r$  with bundle projection

$$\pi^{\text{St}} : \text{St}(r, \mathbf{H}) \longrightarrow \text{St}(r, \mathbf{H})/U_r \cong P_r, \quad \varphi \longmapsto \varphi U_r \longmapsto \varphi \varphi^*. \quad (2.12)$$

The vertical space at  $\varphi$  is  $V_\varphi \text{St}(r, \mathbf{H}) = \{\varphi X : X \in \mathbf{B}(\mathbb{C}^r), X^* + X = 0\}$  and we may choose as horizontal space its orthogonal complement

$$\mathcal{H}_\varphi \text{St}(r, \mathbf{H}) = V_\varphi \text{St}(r, \mathbf{H})^\perp = \{X \in T_\varphi \text{St}(r, \mathbf{H}) : g(X, Y) = 0, \forall Y \in V_\varphi\}.$$

Therefore  $X$  is horizontal if and only if  $\Re \text{tr}(X^* \varphi Y) = 0$  for every  $Y \in \mathbf{B}_{sa}(\mathbb{C}^r)$ . Since  $X^* \varphi$  is skew-adjoint too this happens if and only if  $X^* \varphi = 0$ .

Let us check that the projection (2.12) is a riemannian submersion i.e. its differential induces an isometry from the horizontal space to the tangent space of  $\mathbf{P}_r$ . For horizontal vectors  $X, Y \in T_\varphi \mathbf{St}(r, \mathbf{H})$  we have

$$\begin{aligned} g(d_\varphi \pi^{\mathbf{St}}(X), d_\varphi \pi^{\mathbf{St}}(Y)) &= \operatorname{tr}((X\varphi^* + \varphi X^*)(Y\varphi^* + \varphi Y^*)) \\ &= 2\Re \operatorname{tr}(X^*Y) + \Re \operatorname{tr}(X\varphi^*Y\varphi^* + \varphi X^*\varphi Y^*) \\ &= 2\Re \operatorname{tr}(X^*Y) = g(X, Y). \end{aligned}$$

We have used the properties of the trace and the fact that  $X$  and  $Y$  are horizontal.

Following [34] we derive the geodesic equation

$$\ddot{\gamma} + \gamma(\dot{\gamma}^*\dot{\gamma}) = 0. \quad (2.13)$$

Starting with the condition  $\gamma^*\gamma = \operatorname{Id}_r$  and differentiating two times we get  $\ddot{\gamma}^*\gamma + 2\dot{\gamma}^*\dot{\gamma} + \gamma^*\ddot{\gamma} = 0$ . If  $\gamma$  is a geodesic, the normal component of the second derivative is zero, i.e.  $\ddot{\gamma} = -\gamma S$  for some curve  $S(t) = S(t)^* \in \mathbf{B}(r, \mathbf{H})$ . Inserting this condition in the previous equation we get (2.13). On the other hand any curve  $t \mapsto \mathbf{St}(r, \mathbf{H})$  satisfying (2.13) is a geodesic because the normal component of its second derivative is zero.

We take from [41, Section 3.4.1] a closed formula for the geodesics starting from  $\varphi_0 \in \mathbf{St}(r, \mathbf{H})$ . We continue to use the splitting (2.11) induced by  $\varphi_0$  so that operators in  $\mathbf{H}$  are  $2 \times 2$  block-matrices. For any skew-adjoint operator in the form

$$\mathcal{M} = \begin{pmatrix} A & B \\ -B^* & 0 \end{pmatrix}$$

with skew-adjoint  $A : R(\varphi_0) \rightarrow R(\varphi_0)$ , put  $\mathcal{Q} := \begin{pmatrix} A/2 & 0 \\ 0 & 0 \end{pmatrix}$ . Then  $\mathcal{Q}^* = -\mathcal{Q}$  and we have a curve

$$t \mapsto \gamma(t) := e^{t\mathcal{M}} e^{-t\mathcal{Q}} \varphi_0 \in \mathbf{St}(r, \mathbf{H}).$$

**Proposition 2.2.** *The curve  $\gamma$  is the geodesic in  $\mathbf{St}(r, \mathbf{H})$  satisfying the initial conditions:  $\gamma(0) = \varphi_0$  and  $\dot{\gamma}(0) = \begin{pmatrix} A/2 & B \\ -B^* & 0 \end{pmatrix} \varphi_0$ . Since every tangent vector  $X \in T_{\varphi_0} \mathbf{St}(r, \mathbf{H})$  can be put in the form  $X = \begin{pmatrix} A/2 & B \\ -B^* & 0 \end{pmatrix} \varphi_0$  (with skew-adjoint  $A$ ) this exhausts all the geodesics. Concretely take*

$$A = 2(\varphi_0 \varphi_0^*) X \varphi_0^* \Big|_{R(\varphi_0)} \quad \text{and} \quad B = \varphi_0 X^* (\varphi_0 \varphi_0^* - \operatorname{Id}) \Big|_{R(\varphi_0)^\perp}.$$

*Proof.* The proof that  $\gamma$  is a geodesic is the computation in [41, Section 3.4.1] that we write for definiteness. Since we already know that  $\gamma(t) \in \mathbf{St}(r, \mathbf{H})$  at every time  $t$ , let's check that (2.13) is satisfied i.e.  $\dot{\gamma} = Y$  and  $\dot{Y} = -\gamma(Y^*Y)$ . Put  $\gamma(t) = g(t)\varphi_0$  with  $g(t) = e^{t\mathcal{M}} e^{-t\mathcal{Q}}$ . We also define

$$\mathcal{P} = \begin{pmatrix} A/2 & B \\ -B^* & 0 \end{pmatrix} = \dot{\gamma}(0), \quad \text{and} \quad U(t) := e^{t\mathcal{Q}} \mathcal{P} e^{-t\mathcal{Q}}.$$

It follows  $\mathcal{P} + \mathcal{Q} = \mathcal{M}$  and  $\dot{g}(t) = g(t)U(t)$ . We compute  $Y = g(t)U(t)\varphi_0$  and

$$\begin{aligned} \dot{Y} &= \dot{g}(t)U(t)\varphi_0 + g(t)\dot{U}(t)\varphi_0 = g(t)U^2\varphi_0 + g(t)\dot{U}(t)\varphi_0 \\ &= g(t)e^{t\mathcal{Q}}(\mathcal{P}^2 + [\mathcal{P}, \mathcal{Q}]e^{-t\mathcal{Q}})\varphi_0 \\ &= g(t) \begin{pmatrix} A^2/4 - e^{tA/2} B B^* e^{-tA/2} & 0 \\ 0 & 0 \end{pmatrix} \varphi_0. \end{aligned}$$

Before comparing this result with  $-\gamma(Y^*Y)$  we notice that  $U(t)^* = -U(t)$  and  $g(t)^*g(t) = \operatorname{Id}$ . Finally

$$\begin{aligned} -\gamma(Y^*Y) &= g(t) \varphi_0 \varphi_0^* U(t)^* g(t)^* g(t) U(t) \varphi_0 = -g(t) \varphi_0 \varphi_0^* U^2 \varphi_0 \\ &= -g(t) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} U^2 \varphi_0 \\ &= -g(t) \begin{pmatrix} A^2/4 - e^{tA/2} B B^* e^{-tA/2} & 0 \\ 0 & 0 \end{pmatrix} \varphi_0. \end{aligned}$$

It follows that  $\gamma$  is a geodesic. The remaining statement is straightforward using the decomposition

$$X = \begin{pmatrix} (\varphi_0 \varphi_0^*) X \varphi_0^* & \varphi_0 X^* (\varphi_0 \varphi_0^* - 1) \\ (1 - \varphi_0 \varphi_0^*) X \varphi_0^* & 0 \end{pmatrix} \varphi_0,$$

where all the entries are meant to be restricted to  $R(\varphi_0)$  or  $R(\varphi_0)^\perp$ .  $\square$

**Corollary 2.3.** *For a geodesic  $\gamma : [0, 1] \rightarrow \text{St}(r, \mathbf{H})$ , the image of the map  $\gamma(t) : \mathbb{C}^r \rightarrow \mathbf{H}$  (for every  $t$ ) is contained in the subspace of  $\mathbf{H}$  spanned by  $(\gamma(0), \dot{\gamma}(0))$ . Of course its dimension is bounded by  $2r$  and it follows that if  $\gamma(0)$  and  $\gamma(1)$  are independent then  $\gamma(t)$  and  $\dot{\gamma}(t)$  belong to  $\text{span}(\gamma(0), \gamma(1))$  for every  $t \in [0, 1]$ . The geodesic moves inside a finite dimensional subspace of  $\mathbf{H}$ .*

*Proof.* From the formula of the geodesics we just have to examine the image of the operator  $e^{t\mathcal{M}}$  taking into account that  $X = \dot{\gamma}(0)$ . Then:

$$\mathcal{M} = \begin{pmatrix} 2(\varphi_0\varphi_0^*)\dot{\gamma}(0)\varphi_0^* & \varphi_0\dot{\gamma}(0)^*(\varphi_0\varphi_0^* - \text{Id}) \\ (\text{Id} - \varphi_0\varphi_0^*)\dot{\gamma}(0)\varphi_0^* & 0 \end{pmatrix}.$$

But  $R(\mathcal{M}) \subset \text{Span}(\varphi_0, \dot{\gamma}(0))$  and  $\text{Span}(\varphi_0, \dot{\gamma}(0))$  is stable under  $\mathcal{M}$ .  $\square$

An embedding  $\iota : \mathbf{K} \hookrightarrow \mathbf{H}$  of Hilbert spaces induces embeddings  $\iota_* : \text{St}(r, \mathbf{K}) \hookrightarrow \text{St}(r, \mathbf{H})$  and  $\iota_* : \text{P}_r(\mathbf{K}) \hookrightarrow \text{P}_r(\mathbf{H})$  where we make a slight abuse of notation for using the same symbol for the two maps. Also the notation used for the Grassmannians of different Hilbert spaces is self-explanatory. Indeed we define  $\iota_*\varphi = \iota \circ \varphi$ . This is  $\mathbf{U}_r$ -equivariant and induces the map at the level of the Grassmannians. These embeddings are very useful according to the following.

**Theorem 2.4.** [37]. *Let  $\mathbf{K}$  be a Hilbert space; for every embedding  $\iota : \mathbf{K} \rightarrow \mathbf{H}$  the corresponding  $\iota_* : \text{St}(r, \mathbf{K}) \hookrightarrow \text{St}(r, \mathbf{H})$  is an isometric embedding with totally geodesic image. Moreover:*

1. *When  $\dim \mathbf{K} \geq 2r$ , if we denote with  $\mathbf{d}_{\mathbf{H}}$  and  $\mathbf{d}_{\mathbf{K}}$  the respective distances then  $\mathbf{d}_{\mathbf{H}}(\iota_*(x), \iota_*(y)) = \mathbf{d}_{\mathbf{K}}(x, y)$  for every  $x, y \in \text{St}(r, \mathbf{K})$ .*
2. *Let again  $\dim \mathbf{K} \geq 2r$  and let  $\gamma$  be a minimal geodesic inside  $\text{St}(r, \mathbf{K})$ . Then  $\iota_* \circ \gamma$  is a minimal geodesic.*
3. *The diameter of  $\text{St}(r, \mathbf{H})$  equals the diameter of  $\text{St}(r, \mathbb{C}^{2r})$ .*
4. *Any two points in  $\text{St}(r, \mathbf{H})$  can be joined by a minimal geodesic. Every minimal geodesic  $\gamma$  lies inside some submanifold  $\text{St}(r, V)$  where  $V \subset \mathbf{H}$  is a  $2r$ -dimensional subspace depending on  $\gamma$ .*
5. *Fix two points  $x, y \in \text{St}(r, \mathbf{H})$ ; then  $y$  is in the cut locus of  $x$  if and only if there is a  $2r$ -dimensional subspace  $V \subset \mathbf{H}$  such that  $x = \iota_*(\tilde{x})$ ,  $y = \iota_*(\tilde{y})$  and  $\tilde{y}$  is in the cut locus of  $\tilde{x}$ .*

*All these properties hold for the Grassmannian manifold  $\text{P}_r(\mathbf{H})$  too. In particular any two points  $x, y \in \text{P}_r(\mathbf{H})$  are joined by a minimal geodesic.*

*Proof.* As already mentioned, the proof in [37] is performed for the real Stiefel and Grassmannian manifolds. The key being the fundamental property of the geodesics in Corollary 2.3. One checks immediately that every argument is transferred without changes to the complex case. We write here the proof in loc. cit. in a somewhat sketchy way for the first statement of the Theorem and of properties 1., 2. and 4. both for the Stiefel and the Grassmannians manifolds. We will use these in the proof of Theorem 2.5 below.

First one checks the following fact:

a). Fixed  $y \in \text{St}(r, \mathbf{H})$  the set of all the  $x$  such that the columns of  $x, y$  are independent is dense in the Stiefel manifold.

Then the proof follows the steps:

**Step 1.** The first statement of the Theorem (for the Stiefel manifold) and points 1., 2. and 4. hold when  $\mathbf{H}$  is finite dimensional.

**Step 2.** The statements in **Step 1** hold in the infinite dimensional case.

**Step 3.** Every statement also holds for the Grassmannian.

**Proof of Step 1.** For  $\iota : \mathbf{K} \hookrightarrow \mathbf{H}$  let  $\mathbf{U}(\iota(\mathbf{K})^\perp)$  be the unitary group of the complement. it is included (diagonally) in  $\mathbf{U}(\mathbf{H})$  and acts by isometries on  $\text{St}(r, \mathbf{H})$  with fixed points being exactly  $\iota_*(\text{St}(r, \mathbf{K}))$ . Therefore  $\iota_*(\text{St}(r, \mathbf{K}))$  is totally geodesic being the fixed point set of a set of an isometry.

For the statement 1. we prove it only for those couple of points  $x, y$  of the Stiefel manifold with independent images. Then by Lipschitz continuity of the distances and by the fact a). it will hold for every couple of points. Now  $\mathbf{d}_{\mathbf{K}}(x, y) \geq \mathbf{d}_{\mathbf{H}}(\iota_*(x), \iota_*(y))$  because  $\iota_*(\text{St}(r, \mathbf{K}))$  is totally geodesic. For the reversed inclusion, let  $\gamma \subset \text{St}(r, \mathbf{H})$  be a minimal geodesic (Hopf-Rinow in finite dimensions) joining  $\iota_*(x)$  with  $\iota_*(y)$ . Then since the images of  $\iota_*(x)$  and  $\iota_*(y)$  are independent, by Corollary 2.3 we have that the image of  $\gamma(t)$  is contained in the span of the images of  $\iota_*(x)$  and  $\iota_*(y)$  which is contained in  $\mathbf{K}$ . In other words  $\gamma = \iota_* \circ \tilde{\gamma}$  for a geodesic  $\tilde{\gamma} \subset \text{St}(r, \mathbf{K})$ . Using  $\tilde{\gamma}$  the inequality  $\mathbf{d}_{\mathbf{H}}(\iota_*(x), \iota_*(y)) \geq \mathbf{d}_{\mathbf{K}}(x, y)$  immediately follows.

Point 2. is a direct consequence of point 1. Point 4 is already known from the Corollary 2.3.

**Proof of Step 2.** The unique point which has a different proof in the infinite dimensional case is point 1. Here of course  $d_K(x, y) \geq d_H(\iota_*(x), \iota_*(y))$ . To prove the converse, one takes any smooth path  $\zeta$  connecting  $\iota_*(x)$  and  $\iota_*(y)$ . We can divide  $\zeta$  in subpaths  $\zeta|_{[t_i, t_{i+1}]}$  ( $i = 1, \dots, n$ ) each one contained in a normal neighborhood and using the exponential map each couple  $\zeta(t_i)$  and  $\zeta(t_{i+1})$  can be joined by a minimising geodesic. We get a piecewise smooth path  $\eta(t)$  joining  $\iota_*(x)$  and  $\iota_*(y)$  with  $\ell(\eta) \leq \ell(\zeta)$ . Moreover from all the extreme points  $(\zeta_{t_i})_{i=1, \dots, n-1}$  and the velocities  $(\dot{\eta}_{t_i})_{i=1, \dots, n-1}$  we manufacture a finite dimensional vector space  $\tilde{K}$  which contains every image of the map  $\eta(t)$  for every  $t$ . Of course we can enlarge it to ensure  $K \subset \tilde{K}$ . Now we apply the finite dimensional case (in  $\tilde{K}$ ) to estimate

$$d_K(x, y) = d_{\tilde{K}}(\iota_*(x), \iota_*(y)) \leq \ell(\eta) \leq \ell(\xi)$$

and we are done.

**Proof of step 3.** We check just point 1. and 2. in the finite dimensional case because the infinite dimensional extension is similar to the one performed for the Stiefel case. First point: we have  $d_{P_r(K)}(x, y) \geq d_{P_r(H)}(\iota_*(x), \iota_*(y))$  as before. Also assume that the subspaces  $x$  and  $y$  in the Grassmannian are independent and they generate a  $2r$ -dimensional space. Of course the corresponding fact a). also holds for the Grassmannian. Now let  $\gamma$  be a minimal geodesic in  $P_r(H)$  joining  $\iota_*(x)$  and  $\iota_*(y)$ . Lift this to a curve  $\zeta(t)$  in the Stiefel manifold  $St(r, H)$ . The images of the maps  $\zeta(0)$  and  $\zeta(1)$  are exactly  $x$  and  $y$ . This means that  $\zeta$  belongs to the image of  $St(r, K)$  and in turn that  $\gamma$  belongs to the image of the embedding  $\iota_* : P_r(K) \hookrightarrow P_r(H)$ . It follows  $d_{P_r(K)}(x, y) \leq d_{P_r(H)}(\iota_*(x), \iota_*(y))$ . As before this fact implies the point 2.  $\square$

**Theorem 2.5.** *Every connected component  $P_r$  of the finite rank Grassmannian is an Alexandrov space with non negative curvature.*

*Proof.* According to [52] a complete metric space  $\mathcal{X}$  with intrinsic metric i.e. the metric derived from the length of curves is Alexandrov with non negative scalar curvature if and only if any four points  $p, x, y, z \in \mathcal{X}$  satisfy the inequality

$$d(p, x)^2 + d(p, y)^2 + d(p, z)^2 \geq 1/3(d(x, y)^2 + d(y, z)^2 + d(z, x)^2).$$

For a finite dimensional manifold this condition is equivalent to the non negativity of the sectional curvature. But in our case such a configuration of four points is always included in a finite dimensional totally geodesic submanifold of non negative sectional curvature.  $\square$

Now we prove a simple fact that will be useful later.

**Proposition 2.6.** *Let  $P, Q \in \mathbf{P}$  then  $Q \leq P \implies T_P \mathbf{P} \subset T_Q \mathbf{P}$ . Let moreover  $Q \leq P$  be projections in  $P_c$  and let  $\gamma : [0, 1] \rightarrow P_c$  be the geodesic  $\gamma(t) = e^{itZ} P e^{-itZ}$  starting from  $P$ . Then*

$$Q_1 := e^{iZ} Q e^{-iZ} \leq \gamma(1) =: P_1 \quad \text{and} \quad d(Q, Q_1) \leq \ell(\gamma).$$

*In particular taking  $\gamma$  minimal we have  $d(Q, Q_1) \leq d(P, P_1)$ .*

*Proof.* Let  $Z \in T_P \mathbf{P}$ ; we have to show that  $QZQ = (1 - Q)Z(1 - Q) = 0$ . This is immediate to check under the block decomposition induced by  $P$  where:

$$Q = \begin{pmatrix} QP & 0 \\ 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}. \quad (2.14)$$

Now  $Q(t) := e^{itZ} Q e^{-itZ}$  is a geodesic from  $Q$  to  $Q_1$  with  $\dot{Q}(0) = i[Q, Z]$  and  $\|\dot{Q}(0)\|_{T_{P_c}}^2 = \text{tr}(\dot{Q}(0)^2)$ . Using (2.14) we easily compute

$$\dot{Q}(0)^2 = \begin{pmatrix} QX^*XQ & 0 \\ 0 & XPQPX^* \end{pmatrix}.$$

From the properties of the trace we get  $\|\dot{Q}(0)\|_{T_{P_c}}^2 = 2 \text{tr}(QX^*XQ)$ . In the same way  $\|\dot{\gamma}(0)\|_{T_{P_c}}^2 = 2 \text{tr}(X^*X)$ . The result is clear from the positivity of  $X^*X$ .  $\square$

## 2.2 Normal States

Let  $\mathcal{A}$  be a  $C^*$ -algebra. A linear functional  $\varphi : \mathcal{A} \rightarrow \mathbb{C}$  is positive if  $\varphi(a^*a) \geq 0$  for every  $a \in \mathcal{A}$ . Then  $\varphi$  is automatically bounded; if  $\|\varphi\| = 1$  it is called a state. When the algebra is unital this normalisation is equivalent to the condition  $\varphi(1) = 1$ . We denote then by  $\mathcal{S}(\mathcal{A})$ , the space of the states of  $\mathcal{A}$  included in the dual  $\mathcal{A}'$  and considered with the subspace topology induced by the  $w^*$ -topology. For convenience of the reader we include a sketch of the proof of the following well-know fact.

**Proposition 2.7.** *The space of states is always convex. When  $\mathcal{A}$  is unital it is compact.*

*Proof.* When  $\mathcal{A}$  is unital the convexity is immediate. In general every  $C^*$ -algebra has an *approximate unit*: an increasing net  $(u_j)_{j \in J}$  of positive elements with  $\|u_j\| \leq 1$  for every  $j \in J$  such that

$$\lim_{j \in J} \|a - u_j a\| = 0, \quad \lim_{j \in J} \|a - a u_j\| = 0, \quad \forall a \in \mathcal{A}.$$

If the algebra is separable we can take a sequence for  $(u_j)$ . Now for a linear bounded functional  $\varphi : \mathcal{A} \rightarrow \mathbb{C}$  positivity implies  $\lim_{j \in J} \varphi(u_j) = \|\varphi\|$  (the converse statement also holds but we don't need it). It follows that convex combinations of states are states. The rest of the proof is just the theorem of Banach–Alaoglu.  $\square$

We will denote by  $\mathcal{PS}(\mathcal{A})$  the set of *pure states* that is the boundary of  $\mathcal{S}(\mathcal{A})$  i.e. the subset of extremal points of the boundary of the convex set  $\mathcal{S}(\mathcal{A})$ .

Our object of study will be the space of states of  $\mathbb{K} = \mathbb{K}(\mathbf{H})$ , the  $C^*$ -algebra of compact operators. We have an identification

$$\mathbb{K}' \cong \mathcal{L}^1 \quad (\text{Banach dual}) \quad (2.15)$$

with the Banach space of the trace class operators  $\mathcal{L} = \mathcal{L}^1(\mathbf{H}) = \{A \in \mathbf{B}(\mathbf{H}) : \text{tr}|A| < \infty\}$  endowed with the norm  $\|A\|_1 := \text{tr}|A|$ . Here  $A \in \mathcal{L}^1$  defines the functional  $T \mapsto \text{tr}(AT)$  for  $T \in \mathbb{K}$ . One can also prove that  $\mathcal{L}^1$  is the *predual* of  $\mathbf{B}(\mathbf{H})$  in the sense that  $(\mathcal{L}^1)' = \mathbf{B}(\mathbf{H})$ . Restricting to the positive and norm one functionals we immediately see that for any state  $\varphi \in \mathcal{S}(\mathbb{K}(\mathbf{H}))$  there exists a unique *density matrix*, a positive operator  $\rho \in \mathcal{L}^1$  with

$$\text{tr}(\rho) = 1, \quad \varphi(B) = \text{tr}(\rho B), \quad \text{for every } B \in \mathbb{K}.$$

Viceversa all the density matrices give states on  $\mathbb{K}$ . We denote such space of density matrices by  $\mathcal{C}(\mathbf{H})$  or just  $\mathcal{C}$ , if the context is clear:

$$\mathcal{C}(\mathbf{H}) := \{\rho \in \mathcal{L}^1 : \rho \geq 0, \text{tr}(\rho) = 1\}, \quad (2.16)$$

with the identification denoted by

$$\mathcal{C}(\mathbf{H}) \ni \rho \mapsto \varphi_\rho \in \mathcal{S}(\mathbb{K}). \quad (2.17)$$

We also may, sometimes use the notation  $\rho_\varphi$  or  $\rho(\varphi)$  for the density matrix of  $\varphi$ .

**Example 1.** Every unit vector  $\xi \in \mathbf{H}$  defines a state  $\omega_\xi : \mathbf{B}(\mathbf{H}) \rightarrow \mathbb{R}$  by  $\omega_\xi(B) = \langle \xi, B\xi \rangle$ . The density matrix of  $\omega_\xi$  is the rank one operator  $\rho : \mathbf{H} \rightarrow \mathbf{H}$  with  $\rho(\eta) = \langle \xi, \eta \rangle \xi$ . This follows from:  $\text{tr}(\rho B) = \text{tr}(B\rho) = \langle \xi, B\xi \rangle$ . In Dirac notation our vector is  $|\xi\rangle$  so that

$$\rho = |\xi\rangle\langle \xi|.$$

Density matrices define states of  $\mathbb{K}$  that extend to states of  $\mathbf{B}(\mathbf{H})$  but there are many states on  $\mathbf{B}(\mathbf{H})$  which are not in this form. Precisely a state  $\varphi \in \mathcal{S}(\mathbf{B}(\mathbf{H}))$  comes from a density matrix if and only if it satisfies one of the following equivalent properties (see [46, Theorem 4.12], [43, Theorem 7.1.8] and [33, Theorem 1, Part I, Chapter 4]):

1. it is *normal*:  $\varphi(T) = \sup_{\mathcal{F}} \varphi(F)$  for every directed family  $\mathcal{F} \subset \mathbf{B}(\mathbf{H})^+$  of positive operators with  $T = \sup_{\mathcal{F}} F$ .
2. The state is *completely additive*: for every orthogonal family  $(p_j)_j$  of projections ( $p_j^* = p_j$  and  $p_j p_i = \delta_{ij} p_j$ ) then

$$\varphi\left(\sum_j p_j\right) = \sum_j \varphi(p_j).$$

The sum  $\sum_j p_j$  is defined as the projection on the closure of the smallest subspace in  $\mathbf{H}$  containing all the  $p_j \mathbf{H}$ . This is exactly the operation of forming  $\sup_j p_j$  in the partially ordered set of all the projections in  $\mathbf{B}(\mathbf{H})$  with the order given by the inclusion  $e \leq f$  iff  $e\mathbf{H} \subset f\mathbf{H}$  (see [46]). Also  $\sum_j p_j$  is the limit of all the finite sums in the strong operator topology.

3. There is a sequence of vectors  $(\xi_n)_n$  with  $\sum_{n=1}^{\infty} \|\xi_n\|^2 = 1$  such that

$$\varphi = \sum_{n=1}^{\infty} \omega_{\xi_n}$$

with norm convergence. These vectors can be taken pairwise orthogonal [43, Theorem 7.1.9].

By the spectral theorem we see that a pure state  $\varphi$  of  $\mathbb{K}$  is necessarily a vector state i.e. in the form  $\varphi = \omega_{\xi}$  for a unit vector  $\xi$ . Of course  $\omega_{\xi} = \omega_{\eta}$  if and only if  $\xi = \lambda\eta$  for a *phase*, a scalar  $\lambda \in \mathbb{C}$  with  $|\lambda| = 1$ . Thus  $\mathcal{PS}(\mathbb{K}) \cong \mathbb{P}(\mathbb{H})$ . On the right we have the projective space of  $\mathbb{H}$ , the quotient of the unit sphere by the  $U(1)$ -action by scalar multiplication.

We conclude with a basic useful fact.

**Lemma 2.8.** *For a normal state in the form  $\varphi = \sum_{n=1}^{\infty} \omega_{\xi_n}$  with  $\sum_{n=1}^{\infty} \|\xi_n\|^2 = 1$ , let  $P_n$  be the projection onto  $[\xi_n]$ , the line spanned by the vector. Then the density matrix of  $\varphi$  is:*

$$\rho(\varphi) = \sum_{n=1}^{\infty} \|\xi_n\|^2 P_n \quad \text{norm convergence of operators.}$$

*Proof.* This fact is more general. A proof can be found in [43, Theorem 7.1.9] (see also the following remark therein).  $\square$

We will denote by  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  the collection of all the normal states. To summarize we have recalled that

$$\mathcal{S}(\mathbb{K}) = \mathcal{S}_n(\mathbb{B}(\mathbb{H})), \quad \mathcal{S}_n(\mathbb{B}(\mathbb{H})) \simeq \mathcal{C}(\mathbb{H}),$$

where the symbol  $\simeq$  denotes an isomorphism between two convex sets inside a vector space that is induced by a linear map. This isomorphism maps extremals to extremals: any pure state  $\omega$  on  $\mathbb{K}$  has a unique extension to a normal state  $\omega'$  on  $\mathbb{B}(\mathbb{H})$  given by the same density operator which is extremal for  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$ . We refer to [46] for more details. Based on this we give the following:

**Definition 2.9.** We denote with  $\mathcal{PS}_n(\mathbb{B}(\mathbb{H}))$  the set of the *pure normal states of  $\mathbb{B}(\mathbb{H})$* . These are precisely the extremals of  $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$ . This space is identifiable with  $\mathbb{P}(\mathbb{H})$  the projective space of  $\mathbb{H}$ .

### 2.2.1 Topology on the space of states

We discuss now the various topologies that can be considered on  $\mathcal{S}(\mathbb{K})$  according to the inclusion  $\mathcal{S}(\mathbb{K}) \subset \mathbb{K}'$ .

- The uniform topology is the metric topology induced by the Banach dual structure on  $\mathbb{K}'$ . In terms of density matrices:

$$\|\varphi_{\rho} - \varphi_{\mu}\| = \sup_{B \in \mathbb{K}, \|B\|=1} |\operatorname{tr}(\rho - \mu)B| = \operatorname{tr}|\rho - \mu| = \|\rho - \mu\|_1,$$

because by the Kaplansky density Theorem the supremum can be computed over the unit ball of  $\mathbb{B}(\mathbb{H})$  leading immediately to the trace norm.

- The weak\* topology  $\sigma(\mathcal{S}(\mathbb{K}), \mathbb{K})$  is induced by the weak\* topology on  $\mathbb{K}'$ . In particular  $\varphi_{\rho_n} \xrightarrow{w^*} \varphi_{\rho}$  if  $\varphi_{\rho_n}(B) = \operatorname{tr}(\rho_n B) \rightarrow \varphi_{\rho}(B) = \operatorname{tr}(\rho B)$  for every  $B \in \mathbb{K}$ .
- Instead of evaluating against every  $B \in \mathbb{K}$  in the above convergence we can take all the tests  $B \in \mathbb{B}(\mathbb{H})$ . This defines  $\sigma(\mathcal{S}(\mathbb{K}), \mathbb{B}(\mathbb{H}))$  called the weak topology by virtue of the identification  $\mathbb{B}(\mathbb{H}) = (\mathcal{L}^1)'$ .

Importantly Robinson proved that all the above topologies are the same [56, Theorem 1]:

**Theorem 2.10.** *The three topologies above described all coincide. In particular for a sequence  $\rho_n \in \mathcal{L}^1$  we have*

$$\rho_n \xrightarrow{\mathcal{L}^1} \rho \quad \text{iff} \quad \varphi_{\rho_n} \xrightarrow{w^*} \varphi_{\rho}.$$

### 2.2.2 Partial traces and marginals

Let now  $\mathbb{H}$  and  $\mathbb{K}$  be two Hilbert spaces. The use of  $\mathbb{K}$  does not create confusion with the notation  $\mathbb{K}$  designated for the compact operators. The tensor product Hilbert space  $\mathbb{H} \otimes \mathbb{K}$  corresponds, in quantum mechanics, to a composite system. The isomorphism  $\mathbb{B}(\mathbb{H} \otimes \mathbb{K}) \cong \mathbb{B}(\mathbb{H}) \otimes \mathbb{B}(\mathbb{K})$  induces two maps with the meaning of taking marginals:

$$J_b^{\mathbb{H}} : \mathcal{S}(\mathbb{B}(\mathbb{H} \otimes \mathbb{K})) \rightarrow \mathcal{S}(\mathbb{B}(\mathbb{H})),$$

with a corresponding map  $J_b^K$ . We give the formula of the first one by dualising the inclusion  $J^H : \mathcal{B}(\mathcal{H}) \longrightarrow \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{K})$ ,  $T \longmapsto T \otimes \text{Id}_{\mathcal{K}}$ ; then:

$$J_b^H \varphi(T) = \varphi(T \otimes \text{Id}_{\mathcal{K}}).$$

Let us describe now partial traces. We follow closely the lecture notes [4] where all the proofs can be found.

Assume that  $\mathcal{H}$  and  $\mathcal{K}$  are separable. Every vector  $\xi \in \mathcal{K}$  defines linear bounded operators

$$R_\xi : \mathcal{H} \longrightarrow \mathcal{H} \otimes \mathcal{K}, \quad R_\xi^* : \mathcal{H} \otimes \mathcal{K} \longrightarrow \mathcal{H},$$

uniquely specified on simple tensors by

$$R_\xi \eta = \eta \otimes \xi \quad R_\xi^* \zeta \otimes \eta = \langle \xi, \eta \rangle \zeta.$$

It is immediate to verify that  $\|R_\xi\| = \|R_\xi^*\| = \|\xi\|$ . If  $T \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$  then we get a bounded operator on  $\mathcal{H}$  via

$${}_\xi T_\xi := R_\xi^* T R_\xi.$$

By definition:  $\langle \zeta, {}_\xi T_\xi \eta \rangle = \langle \zeta \otimes \xi, T \eta \otimes \xi \rangle$  for every  $\zeta, \eta \in \mathcal{H}$ . One proves

$$T \in \mathcal{L}^1(\mathcal{H} \otimes \mathcal{K}) \implies {}_\xi T_\xi \in \mathcal{L}^1(\mathcal{H}).$$

**Theorem 2.11.** *Let  $T \in \mathcal{L}^1(\mathcal{H} \otimes \mathcal{K})$  be a trace class operator; there is a unique trace class operator  $\text{Tr}_{\mathcal{K}}(T) \in \mathcal{L}^1(\mathcal{H})$ , called the partial trace with respect to  $\mathcal{K}$ , such that*

$$\text{tr}(\text{Tr}_{\mathcal{K}}(T) B) = \text{tr}(T(B \otimes \text{Id}_{\mathcal{K}})) \quad (2.18)$$

for every  $B \in \mathcal{B}(\mathcal{H})$ . Concretely  $\text{Tr}_{\mathcal{K}}(T)$  can be determined taking any orthonormal basis  $(\xi_n)_n$  of  $\mathcal{K}$ :

$$\text{Tr}_{\mathcal{K}}(T) = \sum_n \xi_n T \xi_n \quad (\text{series convergent in } \mathcal{L}^1(\mathcal{H})).$$

We have the following properties

- $\text{Tr}_{\mathcal{K}}(T) = \text{tr}(B)A$  if  $T = A \otimes B$  with  $A \in \mathcal{L}^1(\mathcal{H})$  and  $B \in \mathcal{L}^1(\mathcal{K})$ ,
- $\text{tr}(\text{Tr}_{\mathcal{K}}(T)) = \text{tr} T$
- $\text{Tr}_{\mathcal{K}}((A \otimes \text{Id}_{\mathcal{K}})T(B \otimes \text{Id}_{\mathcal{K}})) = A \text{Tr}_{\mathcal{K}}(T)B$  for every  $A, B \in \mathcal{B}(\mathcal{H})$ .

Switching the roles of  $\mathcal{H}$  and  $\mathcal{K}$  we define in the same way the partial trace  $\text{Tr}_{\mathcal{H}}$ . If  $\mathcal{K} = \mathcal{H}$ , the unique case we shall treat, we denote with  $\text{Tr}_1$  and  $\text{Tr}_2$  the two partial traces. For instance for  $\xi \otimes \eta \in \mathcal{H} \otimes \mathcal{H}$ :

$$\text{Tr}_1(|\xi \otimes \eta\rangle\langle \xi \otimes \eta|) = \|\xi\|^2 |\eta\rangle\langle \eta|, \quad \text{Tr}_2(|\xi \otimes \eta\rangle\langle \xi \otimes \eta|) = \|\eta\|^2 |\xi\rangle\langle \xi|.$$

Let now  $\varphi \in \mathcal{S}(\mathcal{B}(\mathcal{H} \otimes \mathcal{K}))$  be a state with density matrix  $\rho_\varphi$ . The defining property of the partial trace (2.18) means that, for the density matrix of the first marginal we have:

$$\rho(J_b^H \varphi) = \text{Tr}_{\mathcal{K}}(\rho_\varphi).$$

The density matrices of the partial traces are usually called *reduced density matrices*.

Given normal states  $\varphi \in \mathcal{S}_n(\mathcal{B}(\mathcal{H}))$  and  $\psi \in \mathcal{S}_n(\mathcal{B}(\mathcal{K}))$  the tensor product  $\varphi \otimes \psi$  is a normal state on  $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$ . We say that  $\varphi \otimes \psi$  is *separable*. More generally we agree with [35] on the following.

**Definition 2.12** (Separable and entangled states). A normal state  $\varphi$  on  $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$  is separable if it is limit in the trace norm of a sequence  $\varphi_k$  of normal states each of them is an infinite convex combination of states:

$$\varphi_k = \sum_i p_i^{(k)} \eta_{\mathcal{H}}^{(k,i)} \otimes \eta_{\mathcal{K}}^{(k,i)},$$

with the coefficients  $\{p_i^{(k)}\}_{i=1}^\infty$  forming a probability measure. The trace norm is referred in the above sum to the corresponding density matrices. A normal state on  $\mathcal{B}(\mathcal{H} \otimes \mathcal{K})$  is *entangled* if it is not separable.

Notice in particular that a pure state  $\omega_\zeta$  with  $\zeta \in \mathcal{H} \otimes \mathcal{K}$  is separable if and only if  $\zeta$  is a simple tensor product, i.e.  $\zeta = \xi \otimes \eta$ .

**Notations 2.13.** Summing up the notation we are using:  $\rho$  is a generic density matrix,  $\rho_\varphi$  or  $\rho(\varphi)$  is the density matrix of the normal state  $\varphi$ . If instead we start with  $\rho$ , then  $\varphi_\rho$  is the associated state. Finally vector states defined by  $\xi$  are called  $\omega_\xi$  with density matrix  $\rho = \rho(\omega_\xi) = |\xi\rangle\langle \xi|$ .

### 3 Spectral-projections measures

To any density matrix  $\rho \in \mathcal{C}(\mathbf{H})$  we can associate its unique spectral decomposition for self-adjoint and compact operators

$$\rho = \sum_i \lambda_i P_{V_i}, \quad V_i \subset \mathbf{H}, \quad \text{tr}(\rho) = \sum_i \lambda_i \dim(V_i) = 1, \quad (3.1)$$

where  $\lambda_i > 0$  are the eigenvalues of  $\rho$  and  $P_{V_i} \in \mathbf{P}_c$  is the projection onto the corresponding finite dimensional eigenspace  $V_i$ . In (3.1) the eigenvalues are meant to be listed without repetitions so that:

$$i \neq j \implies \lambda_i \neq \lambda_j \quad \text{and} \quad V_i \perp V_j.$$

Then it is natural to identify the spectral decomposition (3.1) with a discrete, finite and non-negative measure over  $\mathbf{P}_c$ . Before going into details we fix the notation:  $\mathcal{P}(\mathbf{P}_c)$  will denote the space of Borel probability measures (i.e. non-negative and total mass 1) defined over the Polish space  $(\mathbf{P}_c, d)$  while  $\mathcal{M}_+(\mathbf{P}_c)$  is the space of non-negative Radon measures. We now introduce the following set

$$\mathcal{D}(\mathbf{P}_c) := \left\{ \mu = \sum_i \lambda_i \delta_{P_i} : P_i \in \mathbf{P}_c, \lambda_i \geq 0 \right\}, \quad (3.2)$$

where  $\mathcal{D}(\mathbf{P}_c)$  is mnemonic for “discrete” measures. Then we consider the following subsets

$$\mathcal{D}_1(\mathbf{P}_c) := \left\{ \mu \in \mathcal{D}(\mathbf{P}_c) : \text{tr}(\cdot) \mu \in \mathcal{P}(\mathbf{P}_c) \right\}, \quad (3.3)$$

playing the role of probability measures and

$$\mathcal{D}_1^+(\mathbf{P}_c) := \left\{ \mu \in \mathcal{D}_1(\mathbf{P}_c) : PQ = 0, \text{ for all } P \neq Q \in \text{supp}(\mu) \right\}, \quad (3.4)$$

for the space of all the measures supported on orthogonal collections of projections. Of course the defining condition for  $\mathcal{D}_1(\mathbf{P}_c)$  means  $\sum_i \lambda_i \text{tr}(P_i) = 1$ .

We are ready to define the injection:

$$\Phi_0 : \mathcal{C}(\mathbf{H}) \longrightarrow \mathcal{D}_1(\mathbf{P}_c) \subset \mathcal{D}(\mathbf{P}_c), \quad \Phi_0(\rho) = \Phi_0 \left( \sum_i \lambda_i P_{V_i} \right) := \sum_i \lambda_i \delta_{P_{V_i}}. \quad (3.5)$$

For consistency, we will also write:

$$\mu(\rho) := \Phi_0(\rho).$$

Notice that  $\text{tr}(\cdot) \mu_\rho(\mathbf{P}_c) = 1$  follows from  $\text{tr}(\rho) = 1$ . The spectral Theorem implies that  $\Phi_0(\rho) \subset \mathcal{D}_1^+(\mathbf{P}_c)$  for every  $\rho \in \mathcal{C}(\mathbf{H})$  (pairwise orthogonal projections). Moreover, for every  $\rho \in \mathcal{C}(\mathbf{H})$ , as no repetition of eigenvalues is present in  $\Phi_0(\rho)$ , we have:

$$\Phi_0(\rho)(P) \neq \Phi_0(\rho)(Q), \quad \text{for all } P \neq Q \in \text{supp}(\Phi_0(\rho)). \quad (3.6)$$

This property actually characterizes the image  $\Phi_0(\mathcal{C}(\mathbf{H}))$ . Now, we use the identification  $\mathcal{S}_n(\mathbf{B}(\mathbf{H})) \cong \mathcal{C}(\mathbf{H})$  associating  $\varphi \mapsto \rho_\varphi$  to define a map at the level of the normal states.

**Definition 3.1.** The map  $\Phi : \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \longrightarrow \mathcal{D}_1(\mathbf{P}_c)$  is defined by

$$\Phi(\varphi) := \Phi_0(\rho_\varphi) = \sum_i \lambda_i \delta_{P_{V_i}}, \quad (3.7)$$

for every  $\varphi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ . The notation  $\mu_\varphi$  in place of  $\Phi(\varphi)$  will sometimes be preferred.

**Remark 3.2.** The support of  $\mu_\varphi$  is  $\{P_{V_i} : i \in \mathbb{N}\} \subset \mathbf{P}_c$ , a totally disconnected set: notice indeed that by the orthogonality of the eigenspaces,  $\|P_{V_i} - P_{V_j}\|_{\mathbf{B}(\mathbf{H})} = 1$  whenever  $i \neq j$ . Hence  $\{P_{V_i} : i \in \mathbb{N}\}$  is discrete and then closed. Notice also that the projection onto the possibly infinite dimensional subspace  $N(\rho_\varphi)$  does not belong to  $\text{supp}(\mu_\varphi)$ .

We define now the converse correspondence.

**Definition 3.3.** To each element of  $\mathcal{D}_1(\mathbf{P}_c)$  we associate a density matrix in the following way:

$$\Psi : \mathcal{D}_1(\mathbf{P}_c) \longrightarrow \mathcal{C}(\mathbf{H}), \quad \Psi(\mu) = \Psi \left( \sum_i \lambda_i \delta_{P_i} \right) := \sum_i \lambda_i P_i. \quad (3.8)$$

Indeed  $\rho = \sum_i \lambda_i P_i$  converges in the trace norm to a well defined self-adjoint operator having  $\text{tr}(\rho) = \int \text{tr}(P) \mu(dP) = 1$ ; hence  $\Psi(\mathcal{D}_1(\mathbf{P}_c)) \subset \mathcal{C}(\mathbf{H})$ .

By the spectral Theorem again one notices that

$$\Psi(\Phi_0(\rho)) = \rho.$$

Hence  $\Psi$  is the left-inverse of  $\Phi_0$  while, in general, it fails to satisfy  $\Phi_0(\Psi(\mu)) = \mu$  because  $\Phi_0$  is not surjective by (3.6). For instance, let  $\mathbf{H}$  be a finite dimensional Hilbert space, say  $\dim \mathbf{H} = n$  with  $\mathbf{H} = V_1 \oplus V_2$ ,  $V_1 \perp V_2$  and  $\dim V_1 = \dim V_2$ ; then the operator  $\rho := 1/n P_{V_1} + 1/n P_{V_2} = 1/n \text{Id}_{\mathbf{H}} \in \mathcal{C}(\mathbf{H})$  and violates (3.6). Indeed

$$\Phi_0(\rho) = \frac{1}{n} P_{V_1} + \frac{1}{n} P_{V_2}.$$

Particularly relevant for us will be the sets

$$\Lambda_\varphi^\perp := \Psi^{-1}(\varphi) \cap \mathcal{D}_1^\perp(\mathbf{P}_c) = \left\{ \mu \in \mathcal{D}_1^\perp(\mathbf{P}_c) : \Psi(\mu) = \rho_\varphi \right\}. \quad (3.9)$$

These are the sets of those measures concentrated on pairwise orthogonal projections whose corresponding operator is the density matrix of  $\varphi$ . In particular any element of  $\Lambda_\varphi^\perp$  represents a spectral decomposition of  $\rho_\varphi$  admitting repeated eigenvalues.

Coming to the topological properties of these sets, we recall that a sequence of probability measures  $\mu_n \in \mathcal{P}(\mathbf{P}_c)$  weakly converges to  $\mu \in \mathcal{P}(\mathbf{P}_c)$  if

$$\int_{\mathbf{P}_c} f(P) \mu_n(dP) \longrightarrow \int_{\mathbf{P}_c} f(P) \mu(dP), \quad \forall f \in C_b(\mathbf{P}_c).$$

Here  $C_b(\mathbf{P}_c)$  denotes the algebra of continuous bounded functions.

It is well-known that, for Polish spaces, the Lévy–Prokhorov metric metrizes the weak convergence making  $\mathcal{P}(\mathbf{P}_c)$  complete and separable. It will be therefore enough to describe topologically the subsets of  $\mathcal{P}(\mathbf{P}_c)$  only using weakly converging sequences. Moreover we recall the following classical fact about compact subsets of probability measures: if  $(\mathcal{X}, \mathbf{d})$  is a metric space (considered with its Borel  $\sigma$ -algebra), a set  $\mathcal{S} \subset \mathcal{P}(\mathcal{X})$  of probability measures is *tight* whether for every  $\varepsilon > 0$  there is a compact  $K_\varepsilon \subset \mathcal{X}$  such that  $\mu(K_\varepsilon) \geq 1 - \varepsilon$  for every  $\mu \in \mathcal{S}$ . The Prohorov Theorem states that every tight family is relatively compact. If  $\mathcal{X}$  is Polish the converse is true: every relatively compact family is tight.

**Lemma 3.4.** *The map  $\Psi : \mathcal{D}_1(\mathbf{P}_c) \longrightarrow \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  is continuous in the following sense: if  $\text{tr}(\cdot)\mu_n \rightarrow \text{tr}(\cdot)\mu$ , then  $\Psi(\mu_n) \rightarrow \Psi(\mu)$ .*

*Proof.* For each  $B \in \mathcal{L}^1$  we consider the function

$$f_B : \mathbf{P}_c \longrightarrow \mathbb{R}, \quad f_B(P) = \text{tr}(BP)/\text{tr}(P)$$

and zero on  $0 \in \mathbf{P}_c$ . The function is easily seen to be continuous on every connected component of  $\mathbf{P}_c$  and bounded by  $|f_B(P)| = \|B\|$ . Then the following identities

$$\int_{\mathbf{P}_c} f_B(P) \text{tr}(P) \mu_n(dP) = \int_{\mathbf{P}_c} \text{tr}(BP) \mu_n(dP) = \Psi(\mu_n)(B),$$

imply that  $\Psi(\mu_n)(B) \rightarrow \Psi(\mu)(B)$ , for all  $B \in \mathcal{L}^1$ . By density in the norm sense, this is enough to conclude that  $\Psi(\mu_n)(B) \rightarrow \Psi(\mu)(B)$ , for all  $B \in \mathbb{K}$  and the conclusion comes from Theorem 2.10.  $\square$

Then we analyse the topological properties of subsets of discrete measures. In particular, the next result will be crucial in the study of the optimal transport problem between normal states.

**Proposition 3.5.** *The set*

$$\text{tr}(\cdot)\mathcal{D}_1^\perp(\mathbf{P}_c) := \left\{ \text{tr}(\cdot)\mu : \mu \in \mathcal{D}_1^\perp(\mathbf{P}_c) \right\}$$

*is closed. Moreover for any  $\varphi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ , the set  $\text{tr}(\cdot)\Lambda_\varphi^\perp$  is compact.*

*Proof. Step 1.* Consider a sequence  $\mu_n \in \mathcal{D}_1^\perp(\mathbf{P}_c)$  and  $\eta \in \mathcal{P}(\mathbf{P}_c)$  such that  $\text{tr}(\cdot)\mu_n \rightarrow \eta$ . Then for any  $P \in \text{supp}(\eta)$  there exists a sub-sequence  $n_k$  and  $P_k \in \text{supp}(\mu_{n_k})$  such that  $P_k \rightarrow P$ . Any two distinct projections  $P, Q \in \text{supp}(\mu_n)$  verify  $\|P - Q\| = 1$ , then necessarily  $\eta$  is a discrete measure, i.e.  $\eta = \sum_i \lambda_i \delta_{P_i}$ . For the same reason,  $\text{tr}(P_i P_j) = 0$  whenever  $i \neq j$ . Since by assumption  $\eta(\mathbf{P}_c) = 1$ , to have the claim is enough to define  $\mu := \eta/\text{tr}(\cdot)$ . In this way  $\eta \in \text{tr}(\cdot)\mathcal{D}_1^\perp(\mathbf{P}_c)$ .

**Step 2.** We fix the following notation:  $\Phi(\varphi) = \sum_i \lambda_i P_{V_i}$ , where  $P_{V_i}$  denotes the projection onto  $V_i$ . Given any  $\varepsilon > 0$  there exists  $m_\varepsilon \in \mathbb{N}$  such that

$$\sum_{i \geq m_\varepsilon} \lambda_i \operatorname{tr}(P_{V_i}) \leq \varepsilon, \quad \text{with} \quad \lambda_i > \lambda_{i+1}.$$

Let  $N := \max_{i \leq m_\varepsilon} \dim V_i$ . For every  $i \leq m_\varepsilon$  we say that a decomposition of  $P_{V_i}$  is a  $N$ -tuple  $(Q_1, \dots, Q_N) \in \mathcal{P}_c^N$  such that the  $Q_j$  different from zero are mutually orthogonal and satisfy  $\sum_{j=1}^N P_{Q_j} = P_{V_i}$ . If we call  $\mathcal{Q}_i$  the set of all such decompositions we have  $N$  projections  $q_j^{(i)} : \mathcal{Q}_i \rightarrow \mathcal{P}_c$ . Define

$$F_{V_i} := \bigcup_{j=1}^N q_j^{(i)}(\mathcal{Q}_i)$$

the set of all the projections appearing in at least one decomposition of  $V_i$ . Let  $\mathbb{G}(V_i, d)$  be the Grassmann manifold of all the subspaces of dimension  $d$  inside  $V_i$ . We may embed  $\mathcal{Q}_i$  into the union of all products  $\mathbb{G}(V_i, \operatorname{tr}(Q_1)) \times \dots \times \mathbb{G}(V_i, \operatorname{tr}(Q_N))$  with the union running over the finite set of all the possible ways of writing  $N = s_1 + \dots + s_N$  with  $s_j \in \mathbb{N}$  (including zero). We adopt the convention that  $\mathbb{G}(V_i, 0) = \bullet$ , the space with a point. Since the Grassmannians are compact we see that  $F_{V_i}$  and also  $\bigcup_{i=1}^{m_\varepsilon} F_{V_i}$  are relatively compact inside  $\mathcal{P}_c$ .

Now pick any  $\mu \in \Lambda_\varphi^\perp$ . We write it in the form  $\mu = \sum_i \lambda_i \delta_{P_{Z_i}}$  where the eigenvalues are the same as the ones of  $\Phi(\varphi)$  but here they may be repeated. Therefore,

$$\operatorname{tr}(\cdot)\mu \left( \bigcup_{i=1}^{m_\varepsilon} F_{V_i} \right) = \sum_{i \leq m_\varepsilon} \lambda_i \operatorname{tr}(P_{Z_i}) = \sum_{i \leq m_\varepsilon} \lambda_i \operatorname{tr}(P_{V_i}) \geq 1 - \varepsilon,$$

where the second identity is established collecting different projections with the same eigenvalue. This proves that  $\operatorname{tr}(\cdot)\Lambda_\varphi^\perp$  is tight. To prove the compactness it is enough to recall that tightness is equivalent to precompactness in  $\mathcal{P}(\mathcal{P}_c)$ . Moreover by Lemma 3.4 and the previous part of the proof  $\operatorname{tr}(\cdot)\Lambda_\varphi^\perp$  is closed. The claim follows.  $\square$

### 3.1 Weak\* topology and convergence of projections

We now relate the weak\* convergence of normal states with the spectral decomposition of their associated density matrices.

**Lemma 3.6.** *Let  $(P_n)_n$  be a sequence of projections inside  $\mathcal{L}^1$ ; assume moreover  $P_n \rightarrow P$  in the  $w$ -topology to some  $P \in \mathcal{L}^1$ , i.e. in duality with  $\mathcal{B}(\mathcal{H})$ . Then  $P$  is a projection:  $P^2 = P$  and  $P^* = P$ .*

*Proof.* To check  $P = P^*$  it is sufficient to notice that  $\langle Px, x \rangle \in \mathbb{R}$ , being the limit of  $\langle P_n x, x \rangle \in \mathbb{R}$ .

To prove  $P^2 = P$ , first we can assume that  $P \neq 0$  otherwise the claim is trivial. From  $P_n \rightarrow P$  in the weak topology we deduce that

$$\|P_n\|_{\mathcal{L}^1} = \operatorname{tr}(P_n) \rightarrow \operatorname{tr}(P) = \|P\|_{\mathcal{L}^1},$$

implying (see Theorem 2.10) that  $P_n \rightarrow P$  in  $\mathcal{L}^1$ . Then for any  $K \in \mathbb{K}$

$$\operatorname{tr}(P_n K P_n) = \operatorname{tr}(P_n^2 K) = \operatorname{tr}(P_n K) \rightarrow \operatorname{tr}(PK);$$

on the other hand, since  $K P_n \rightarrow K P$  in  $\mathcal{L}^1$ , it follows that  $\operatorname{tr}(P_n K P_n) \rightarrow \operatorname{tr}(PKP)$ . Therefore  $P^2 = P$ .  $\square$

**Lemma 3.7.** *Let  $\varphi_n, \varphi \in \mathcal{S}_n(\mathcal{B}(\mathcal{H}))$  such that  $\varphi_n \rightarrow \varphi$ . Consider the corresponding density matrices  $\rho_n = \rho_{\varphi_n}$  and  $\rho = \rho_\varphi$ , for which we consider any spectral decompositions (in the sense of (3.9))*

$$\rho_n = \sum_i \lambda_i^n P_i^n, \quad \rho = \sum_i \lambda_i P_i.$$

*In particular repetitions of eigenvalues are allowed. Let  $(\lambda_i^{n_k})_{k \in \mathbb{N}}$  be a subsequence converging to  $\tilde{\lambda}_i \neq 0$  and let  $\tilde{P}_i$  be any  $w^*$ -limit of the corresponding subsequence of projections  $(P_i^{n_k})_{k \in \mathbb{N}}$ .*

*Then  $P_i^{n_k} \rightarrow \tilde{P}_i$  in  $\mathcal{L}^1$  and therefore  $\tilde{P}_i$  is a projection. Moreover there exists  $j \in \mathbb{N}$  such that*

$$\tilde{P}_i \leq P_j, \quad \tilde{\lambda}_i = \lambda_j. \quad (3.10)$$

*Proof.* We start noticing the following: for any  $B \in \mathbf{B}(\mathbf{H})$  it holds true that  $\mathrm{tr}(\rho_{n_k} P_i^{n_k} B) \rightarrow \mathrm{tr}(\rho \tilde{P}_i B)$ . Indeed

$$\mathrm{tr}(\rho_{n_k} P_i^{n_k} B) - \mathrm{tr}(\rho \tilde{P}_i B) = \mathrm{tr}(\rho_{n_k} P_i^{n_k} B) - \mathrm{tr}(\rho P_i^{n_k} B) + \mathrm{tr}(\rho P_i^{n_k} B) - \mathrm{tr}(\rho \tilde{P}_i B);$$

then the first term goes to zero from  $\rho_{n_k} \rightarrow \rho$  in  $\mathcal{L}^1$  while the second one converges to zero from the  $w^*$ -convergence of  $P_i^{n_k}$  to  $\tilde{P}_i$  and by the compactness of  $\rho$ . Moreover by the orthogonality of the projections it follows that

$$\rho_{n_k} P_i^{n_k} = \lambda_i^{n_k} P_i^{n_k}.$$

Hence by  $\lambda_i^{n_k} \rightarrow \tilde{\lambda}_i \neq 0$ ,  $P_i^{n_k}$  is  $w$ -converging to  $\rho \tilde{P}_i / \tilde{\lambda}_i$ .

Then, since the  $w$  and  $w^*$  limits coincide, we deduce that  $\tilde{P}_i = \rho \tilde{P}_i / \tilde{\lambda}_i$  and that  $P_i^{n_k} \rightarrow \tilde{P}_i$  weakly. Therefore following the proof of Lemma 3.6,  $P_i^{n_k} \rightarrow \tilde{P}_i$  in  $\mathcal{L}^1$  and  $\tilde{P}_i$  is a projection, proving the first part of the claim.

To obtain the second part we observe that the previous identity  $\rho \tilde{P}_i = \tilde{\lambda}_i \tilde{P}_i$  implies the claim together with the uniqueness of the spectral decomposition of the compact and self-adjoint operator  $\rho$ .  $\square$

**Proposition 3.8.** *Let  $\varphi_n, \varphi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  such that  $\varphi_n \rightarrow \varphi$ . Then for any sequence  $\mu_n \in \Lambda_{\varphi_n}^\perp$  there exist a subsequence  $\mu_{n_k}$  and  $\mu \in \Lambda_\varphi^\perp$  such that  $\mathrm{tr}(\cdot)\mu_{n_k} \rightarrow \mathrm{tr}(\cdot)\mu$ , i.e. in duality with continuous and bounded functions  $C_b(\mathbf{P}_c)$ .*

*Proof. Step 1.* Consider the sequences  $(\lambda_i^n)_{i \in \mathbb{N}}$  and  $(\lambda_i)_{i \in \mathbb{N}}$  of eigenvalues of  $\rho_n$  and  $\rho$ , respectively, arranged in decreasing order and repeated according to the multiplicity; in particular both sequences have norm 1 in  $\ell^1$ . Then [56, Theorem 2] proves that

$$\sum_i |\lambda_i^n - \lambda_i| \leq \mathrm{tr}|\rho_n - \rho|,$$

giving that  $(\lambda_i^n)_{i \in \mathbb{N}} \rightarrow (\lambda_i)_{i \in \mathbb{N}}$  in the  $\ell^1$ -norm as  $n \rightarrow \infty$ . As a straightforward consequence for each  $\varepsilon$  there exist  $n_\varepsilon, M \in \mathbb{N}$  such that

$$\sum_{i \geq M} \lambda_i^n \leq \varepsilon, \quad \forall n \geq n_\varepsilon, \quad \sum_{i \geq M} \lambda_i \leq \varepsilon. \quad (3.11)$$

It is not restrictive to assume  $\lambda_i > 0$  for each  $i < M$  for if this is not the case we can simply lower  $M$  without changing the validity of (3.11). Then the  $\ell^1$ -convergence will imply (3.11) for  $(\lambda_i^n)_i$  as well.

**Step 2.** Consider now any sequence  $\mu_n \in \Lambda_{\varphi_n}^\perp$ . To fix the notations we write

$$\mu_n = \sum_j \lambda_{i_j}^n \delta_{P_j^n}, \quad \rho_n = \sum_j \lambda_{i_j}^n P_j^n.$$

Then we proceed as follows: denote with  $m \in \mathbb{N}$  the first number such that  $\lambda_m = 0$ , with  $(\lambda_i)_{i \in \mathbb{N}}$  seen as an element of  $\ell^1(\mathbb{N})$ ; in particular  $m \geq M$ . If  $\lambda_i > 0$  for all  $i \in \mathbb{N}$ , we put  $m = \infty$ .

From the  $\ell^1$ -convergence we have  $\lambda_i^n \rightarrow \lambda_i$  as  $n \rightarrow \infty$ . Hence for each  $j \in \mathbb{N}$  such that  $i_j < m$ , the sequence of projections  $(P_j^n)_{n \in \mathbb{N}}$  has trace uniformly bounded; hence the  $w^*$ -precompactness together with Lemma 3.7 imply the existence of a subsequence  $n_k$  and of a projection  $P_j$  such that

$$P_j^{n_k} \rightarrow P_j, \quad \rho P_j = \lambda_j P_j.$$

Via the usual diagonal argument, we deduce the existence of a subsequence, still denoted by  $n_k$ , such that for each  $j \in \mathbb{N}$  such that  $i_j < m$ , then:

$$P_j^{n_k} \rightarrow P_j, \quad \rho P_j = \lambda_j P_j$$

as  $k \rightarrow \infty$ . We define therefore  $\mu := \sum_j \lambda_j \delta_{P_j}$ . By the norm convergence, if  $j_1 \neq j_2$  then

$$\mathrm{tr}(P_{j_1} P_{j_2}) = 0,$$

and by  $\rho P_j = \lambda_j P_j$  it follows that  $\rho \geq \sum_j \lambda_j P_j$ . Moreover, since

$$\sum_{j: i_j \geq M} \lambda_{i_j}^{n_k} \mathrm{tr}(P_j^{n_k}) = \sum_{i \geq M} \lambda_i^n \leq \varepsilon,$$

it follows that

$$\sum_j \lambda_j \mathrm{tr}(P_j) \geq \limsup_n \sum_{j: i_j \leq M} \lambda_{i_j}^n \mathrm{tr}(P_j^{n_k}) \geq 1 - \varepsilon.$$

Since  $\varepsilon$  was arbitrarily chosen and did not play any role in the construction of  $\mu$ , it follows that  $\sum_j \lambda_j \operatorname{tr}(P_j) = 1$ , giving,  $\mu \in \Lambda_\varphi^\perp$ . As a byproduct we have also shown that the sequence  $(\lambda_{i_j^{n_k}} \operatorname{tr}(P_j^{n_k}))$  converges to  $(\lambda_j \operatorname{tr}(P_j))$  in  $\ell^1(\mathbb{N})$ .

**Step 3.** The claim is now equivalent to proving that for any  $f \in C_b(\mathbf{P}_c)$

$$\lim_{k \rightarrow \infty} \sum_j \lambda_{i_j^{n_k}} \operatorname{tr}(P_j^{n_k}) f(P_j^{n_k}) = \sum_j \lambda_j \operatorname{tr}(P_j) f(P_j).$$

This now follows from the  $\ell^1(\mathbb{N})$ -convergence of  $(\lambda_{i_j^{n_k}} \operatorname{tr}(P_j^{n_k}))$  to  $(\lambda_j \operatorname{tr}(P_j))$  and the norm convergence of each  $P_j^{n_k}$  to  $P_j$  (implying convergence in  $\mathbf{d}$ ) coupled with continuity and boundedness of  $f$ .  $\square$

We summarise the results in the next statement.

**Theorem 3.9.** *Let  $\varphi_n, \varphi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  be normal states and consider  $\mu_n \in \Lambda_{\varphi_n}^\perp, \mu \in \Lambda_\varphi^\perp$ . Then*

1. *If  $\operatorname{tr}(\cdot) \mu_n \rightarrow \operatorname{tr}(\cdot) \mu$  in duality with  $C_b(\mathbf{P}_c)$ , then  $\varphi_n \rightarrow \varphi$  in the  $w^*$ -sense.*
2. *If  $\varphi_n \rightarrow \varphi$  in the  $w^*$ -sense then there exist a subsequence  $(n_k)_{k \in \mathbb{N}}$  and  $\bar{\mu} \in \Lambda_\varphi^\perp$  such that  $\operatorname{tr}(\cdot) \mu_{n_k} \rightarrow \operatorname{tr}(\cdot) \bar{\mu}$  in duality with  $C_b(\mathbf{P}_c)$ .*

*Proof.* The first point is Lemma 3.4 while the second part of the claim is precisely Proposition 3.8.  $\square$

## 4 The Wasserstein distance between normal states

We will use the metric structure of  $(\mathbf{P}_c, \mathbf{d})$  reviewed in Section 2.1, together with the map  $\Psi$  (Definition 3.3) to define a *static* Wasserstein distance between normal states of  $\mathbf{B}(\mathbf{H})$ . The plan is to push the classical  $p$ -Wasserstein distance over  $(\mathbf{P}_c, \mathbf{d})$  (the fact that  $\mathbf{d}$  is an extended metric does not hurt the definition) to the normal states via  $\Psi$ .

We begin describing in details  $W_p^{\mathbf{P}_c}$ .

### 4.1 The Wasserstein distance over $\mathcal{P}(\mathbf{P}_c)$

In the classical setting optimal transportation is encoded in transport plans, i.e. probability measures over the product space with assigned marginals. As the metric  $\mathbf{d}$  is finite solely when restricted on each connected component of  $\mathbf{P}_c$ , we will consider a more stringent notion of transport plan (recall that  $P, Q \in \mathbf{P}_c$  belong to the same connected component if and only if  $\dim R(P) = \dim R(Q)$ ).

**Definition 4.1.** Given two probability measures  $\mu_0, \mu_1 \in \mathcal{P}(\mathbf{P}_c)$ , the set of  $\mathbf{d}$ -transport plans between  $\mu_0$  and  $\mu_1$  will be given by

$$\Pi_{\mathbf{d}}(\mu_0, \mu_1) := \left\{ \nu \in \Pi(\mu_0, \mu_1) : \dim(R(P)) = \dim(R(Q)), \nu - a.e. \right\}, \quad (4.1)$$

where  $\Pi(\mu_0, \mu_1) = \{ \nu \in \mathcal{P}(\mathbf{P}_c \times \mathbf{P}_c) : (\pi_1)_\# \nu = \mu_0, (\pi_2)_\# \nu = \mu_1 \}$  is the classical notation for transport plans and  $\pi_i : \mathbf{P}_c \times \mathbf{P}_c \rightarrow \mathbf{P}_c$  is the projection on the  $i$ -th component, for  $i = 1, 2$ .

The set of  $\mathbf{d}$ -transport plans  $\Pi_{\mathbf{d}}(\mu_0, \mu_1)$  is a, possibly empty, convex subset of  $\mathcal{P}(\mathbf{P}_c \times \mathbf{P}_c)$ . Then we will define the following optimal transport distance.

**Definition 4.2.** Given  $\mu_0, \mu_1 \in \mathcal{P}(\mathbf{P}_c)$ , for any  $p \geq 1$  we define their  $W_p^{\mathbf{P}_c}$  distance as follows:

$$W_p^{\mathbf{P}_c}(\mu_0, \mu_1) := \inf_{\nu \in \Pi_{\mathbf{d}}(\mu_0, \mu_1)} \left( \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \nu(dP dQ), \right)^{\frac{1}{p}} \quad (4.2)$$

where  $\mathbf{d}$  is the extended geodesic distance of  $\mathbf{P}_c$ . Whenever the set  $\Pi_{\mathbf{d}}(\mu_0, \mu_1)$  is empty we pose  $W_p^{\mathbf{P}_c}(\mu_0, \mu_1) := +\infty$ .

It is fairly easy (and almost identical to the classical case) to prove existence of optimal transport plans.

**Theorem 4.3** (Existence of optimal plans). *Given  $\mu_0, \mu_1 \in \mathcal{P}(\mathbf{P}_c)$ , there exists an optimal plan  $\nu \in \Pi_{\mathbf{d}}(\mu_0, \mu_1)$  such that*

$$W_p^{\mathbf{P}_c}(\mu_0, \mu_1)^p = \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \nu(dP dQ),$$

*provided the set of admissible plan  $\Pi_{\mathbf{d}}(\mu_0, \mu_1)$  is not empty.*

*Proof.* Since  $\Pi_d(\mu_0, \mu_1) \neq \emptyset$  and  $d \leq \frac{\pi}{2}$ , there exists a minimizing sequence  $\nu_n \in \Pi_d(\mu_0, \mu_1)$  such that

$$\lim_n \int_{\mathbb{P}_c \times \mathbb{P}_c} d(P, Q)^p \nu_n(dPdQ) = W_p^{Pc}(\mu_0, \mu_1)^p.$$

Thanks to the following Lemma 4.4, there exist  $\nu_{n_k}, \nu \in \Pi_d(\mu_0, \mu_1)$  such that  $\nu_{n_k} \rightharpoonup \nu$ , in duality with  $C_b(\mathbb{P}_c \times \mathbb{P}_c)$ . Being the distance continuous and bounded it follows that

$$W_p^{Pc}(\mu_0, \mu_1)^p = \lim_k \int_{\mathbb{P}_c \times \mathbb{P}_c} d(P, Q)^p \nu_{n_k}(dPdQ) = \int_{\mathbb{P}_c \times \mathbb{P}_c} d(P, Q)^p \nu(dPdQ)$$

proving the claim.  $\square$

**Lemma 4.4.** *Given  $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{P}_c)$ , for any sequence  $\nu_n \in \Pi_d(\mu_0, \mu_1)$  there exist a subsequence  $\nu_{n_k}$  and  $\nu \in \Pi_d(\mu_0, \mu_1)$  such that  $\nu_{n_k} \rightharpoonup \nu$  in duality with any  $f \in C_b(\mathbb{P}_c \times \mathbb{P}_c)$ .*

Although the argument is similar to the classical case, for readers' convenience we include the proof.

*Proof.* By the inner regularity of probability measures over Polish spaces, for any  $\varepsilon > 0$  there exist compact sets  $K_1 \subset \text{supp}(\mu_0)$  and  $K_2 \subset \text{supp}(\mu_1)$  such that  $\mu_0(K_1), \mu_1(K_2) \geq 1 - \varepsilon$  implying that  $\nu_n(K_1 \times K_2) \geq 1 - 2\varepsilon$ . Thus  $\nu_n$  is tight. Then the Prohorov Theorem ensures the existence of a subsequence  $\nu_{n_k}$  and of a plan  $\nu \in \mathcal{P}(\mathbb{P}_c \times \mathbb{P}_c)$  such that  $\nu_{n_k} \rightharpoonup \nu$  in duality with any  $f \in C_b(\mathbb{P}_c \times \mathbb{P}_c)$ . In particular this implies that  $(\pi_1)_\# \nu = \mu_0$  and  $(\pi_2)_\# \nu = \mu_1$ , proving that  $\nu \in \Pi(\mu_0, \mu_1)$ .

To conclude, consider the function  $f : \mathbb{P}_c \times \mathbb{P}_c \rightarrow \mathbb{R}$  defined by  $f(P, Q) := \dim(R(P)) - \dim(R(Q))$ . The function  $f$  is locally constant and therefore continuous. Hence the set

$$C := \{(P, Q) \in \mathbb{P}_c \times \mathbb{P}_c : f(P, Q) > 0\}$$

is open giving that  $0 = \liminf \nu_{n_k}(C) \geq \nu(C) \geq 0$ . This proves the claim.  $\square$

We conclude this short overview on the optimal transport in  $\mathcal{P}(\mathbb{P}_c)$  by recalling the simple relation between the Wasserstein topology and the weak topology. Here we refer to [57, Theorem 6.9]: if  $\mu_n, \mu \in \mathcal{P}(\mathbb{P}_k)$  for some  $k$  independent of  $n$ , then

$$\mu_n \rightharpoonup \mu \iff W_p^{Pc}(\mu_n, \mu) \rightarrow 0,$$

for any  $p \geq 1$ . Indeed  $(\mathbb{P}_k, d)$  is a complete and separable metric space with  $d \leq \pi/2$ .

## 4.2 The Optimal Transport Cost in $\mathcal{S}_n(\mathbb{B}(\mathbb{H}))$

We now consider the natural optimal transport problem between normal states.

The existence of multiple representations for the density matrices, i.e.  $\Lambda_\varphi^\perp$  and  $\Lambda_\psi^\perp$ , as well as the several connected components of  $(\mathbb{P}_c, d)$ , motivate the following definition.

**Definition 4.5.** For any  $\varphi, \psi \in \mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  and  $p \geq 1$  define their optimal transport *cost* by

$$\mathcal{C}_p(\varphi, \psi) := \inf_{\substack{\mu_0 \in \Lambda_\varphi^\perp \\ \mu_1 \in \Lambda_\psi^\perp}} W_p^{Pc}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1), \quad (4.3)$$

where  $\Lambda_\varphi^\perp, \Lambda_\psi^\perp \subset \mathcal{D}_1^\perp(\mathbb{P}_c)$  have been defined in (3.9).

**Remark 4.6.** Clearly an alternative way of writing  $\mathcal{C}_p$  is to interpret it as the distance between two disjoint compact sets: For any  $\varphi, \psi \in \mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  and  $p \geq 1$

$$\mathcal{C}_p(\varphi, \psi) = W_p^{Pc}(\text{tr}(\cdot)\Lambda_\varphi^\perp, \text{tr}(\cdot)\Lambda_\psi^\perp), \quad (4.4)$$

where as usual the distance between two compact sets is computed taking the infimum of all possible distances.

It is immediate to check that  $\mathcal{C}_p$  is bounded.

**Lemma 4.7.** *Given any  $\varphi, \psi \in \mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  we have  $\mathcal{C}_p(\varphi, \psi) \leq \pi/2$ .*

*Proof.* It is sufficient to observe that given any  $\varphi \in \mathcal{S}_n(\mathbb{B}(\mathbb{H}))$  there exists  $\mu_0 \in \Lambda_\varphi^\perp$  such that  $\text{supp}(\mu_0) \subset \mathbb{P}_1$  (recall (2.9)). Hence by definition

$$\mathcal{C}_p(\varphi, \psi) \leq W_p^{Pc}(\mu_0, \mu_1) \leq \pi/2,$$

where  $\mu_1 \in \Lambda_\psi^\perp$  and  $\text{supp}(\mu_0), \text{supp}(\mu_1) \subset \mathbb{P}_1$ . The second inequality follows from  $d(P, Q) \leq \pi/2$  whenever  $P, Q \in \mathbb{P}$  belong to the same connected component.  $\square$

Relying on Proposition 2.6, we deduce that looking among those spectral representations of states using projections with one-dimensional range does not change the cost functional.

**Proposition 4.8.** *For any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ ,*

$$\mathcal{C}_p(\varphi, \psi) = \inf_{\substack{\mu_0 \in \Lambda_\varphi^\perp \cap \mathcal{P}(\mathbf{P}_1) \\ \mu_1 \in \Lambda_\psi^\perp \cap \mathcal{P}(\mathbf{P}_1)}} W_p^{\text{Pc}}(\mu_0, \mu_1),$$

*Proof.* Consider any  $\mu_0 \in \Lambda_\varphi^\perp$ ,  $\mu_1 \in \Lambda_\psi^\perp$  and  $\nu \in \Pi_d(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1)$  (it is not restrictive to assume the existence of at least one transport plan).

We prove the claim showing the existence of  $\gamma \in \mathcal{P}(\mathbf{P}_1 \times \mathbf{P}_1)$  such that

$$\int \mathbf{d}(P, Q)^p \gamma(dPdQ) \leq \int \mathbf{d}(P, Q)^p \nu(dPdQ),$$

with  $(\pi_1)_\# \gamma \in \Lambda_\varphi^\perp$  and  $(\pi_2)_\# \gamma \in \Lambda_\psi^\perp$ .

We proceed by writing  $\nu$  as follows: if  $\text{tr}(\cdot)\mu_0 = \sum_i \alpha_{0,i} \delta_{P_{0,i}}$  (and analogous one for  $\text{tr}(\cdot)\mu_1$ ), then

$$\nu = \sum_{i,j} \beta_{i,j} \delta_{P_{0,i}} \otimes \delta_{P_{1,j}},$$

for some  $\beta_{i,j} \geq 0$  summing to 1. Whenever  $\beta_{i,j} > 0$  and  $\text{tr}(P_{0,i}) = r > 1$ , we consider any orthonormal frame of  $R(P_{0,i})$ , say  $e_1, \dots, e_r$  such that  $\sum_{k \leq r} P_{e_k} = P_{0,i}$ .

We also consider  $Z \in T_{P_{0,i}} \mathbf{P}$  such that  $P_{1,j} = e^{iZ} P_{0,i} e^{-iZ}$  and consequently define  $P_{1,e_k} := e^{iZ} P_{e_k} e^{-iZ}$ . Clearly

$$\sum_{k \leq r} P_{1,e_k} = P_{1,j},$$

and by Proposition 2.6  $\mathbf{d}(P_{e_k}, P_{1,e_k}) \leq \mathbf{d}(P_{0,i}, P_{1,j})$ . We therefore define a new transport plan  $\bar{\nu}$  replacing  $\delta_{P_{0,i}} \otimes \delta_{P_{0,j}}$  by

$$\frac{1}{r} \sum_k \delta_{P_{e_k}} \otimes \delta_{P_{1,e_k}}.$$

Then

$$\int \mathbf{d}(P, Q)^p \nu(dPdQ) - \int \mathbf{d}(P, Q)^p \bar{\nu}(dPdQ) = \beta_{i,j} \left( \mathbf{d}(P_{0,i}, P_{1,j})^p - \frac{1}{r} \mathbf{d}(P_{e_k}, P_{1,e_k})^p \right) \geq 0$$

It is clear from the construction that the marginal measures of  $\nu$  are still admissible for the states  $\varphi$  and  $\psi$ . Repeating the argument at most countably many times proves the claim.  $\square$

After Proposition 2.6, we therefore introduce the following additional notation:

$$\Lambda_\varphi^{\perp,1} := \Lambda_\varphi^\perp \cap \mathcal{P}(\mathbf{P}_1). \quad (4.5)$$

Notice that  $\Lambda_\varphi^{\perp,1}$  is closed, and therefore compact, as  $\Lambda_\varphi^\perp$ .

Next we prove that the infimum of (4.3) can be replaced by a minimum.

**Proposition 4.9.** *Given any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ , there exist  $\mu_0, \mu_1$  and  $\nu$ , elements of  $\Lambda_\varphi^{\perp,1}, \Lambda_\psi^{\perp,1}$  and of  $\Pi_d(\mu_0, \mu_1)$  respectively, such that*

$$\mathcal{C}_p(\varphi, \psi) = W_p^{\text{Pc}}(\mu_0, \mu_1) = \left( \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \nu(dPdQ) \right)^{1/p}.$$

*Proof.* The second identity is proved in Theorem 4.3. It is enough therefore to show the first one. By Proposition 2.6, there exists two sequences  $(\mu_{0,n})_{n \in \mathbb{N}} \subset \Lambda_\varphi^{\perp,1}, (\mu_{1,n})_{n \in \mathbb{N}} \subset \Lambda_\psi^{\perp,1}$  such that

$$\lim_{n \rightarrow \infty} W_p^{\text{Pc}}(\mu_{0,n}, \mu_{1,n}) = \mathcal{C}_p(\varphi, \psi).$$

By compactness of  $\text{tr}(\cdot)\Lambda_\varphi^{\perp,1}$  and  $\text{tr}(\cdot)\Lambda_\psi^{\perp,1}$  (Proposition 3.5), we assume, up to subsequences that we omit, that:  $\mu_{0,n} \rightharpoonup \mu_0$ ,  $\mu_{1,n} \rightharpoonup \mu_1$ , for some  $\mu_0 \in \Lambda_\varphi^{\perp,1}$  and  $\mu_1 \in \Lambda_\psi^{\perp,1}$ .

Now take  $\nu_n \in \Pi_d(\mu_{0,n}, \mu_{1,n})$  any optimal transport plan (Theorem 4.3). Since its marginals are converging, by tightness,  $\nu_n$  is weakly converging, up to subsequences, as well to a certain  $\nu \in \Pi(\mu_0, \mu_1)$ . Since  $\mathbf{d}$  is continuous and bounded on  $\mathbf{P}_1$ :

$$\mathcal{C}_p(\varphi, \psi) = \lim_{n \rightarrow \infty} W_p^{\text{Pc}}(\mu_{0,n}, \mu_{1,n})^p = \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \nu(dPdQ) \geq W_p^{\text{Pc}}(\mu_0, \mu_1)^p.$$

Continuing the previous chain of inequalities with  $\geq \mathcal{C}_p(\varphi, \psi)$  proves the claim.  $\square$

We have shown that the optimal transport problem defining the transport cost has always a solution. Moreover, by the symmetry of  $\mathbf{d}$ , it is trivial to check that  $\mathcal{C}_p(\varphi, \psi) = \mathcal{C}_p(\psi, \varphi)$ .

Finally  $\mathcal{C}_p(\varphi, \psi) = 0$  implies  $\varphi = \psi$ . Indeed by the previous Proposition 4.9, there exist  $\mu_0 \in \Lambda_\varphi^{\perp,1}$  and  $\mu_1 \in \Lambda_\psi^{\perp,1}$  such that  $W_p^{\text{Pc}}(\mu_0, \mu_1) = 0$ . Hence  $\mu_0 = \mu_1$  and by definition of  $\Lambda_\varphi^{\perp,1}$  and  $\Lambda_\psi^{\perp,1}$ , we have:

$$\varphi = \Psi(\mu_0) = \Psi(\mu_1) = \psi.$$

Therefore the map  $\mathcal{C}_p : \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \times \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \rightarrow [0, \infty)$  is a semi-distance. First references for semi-distances date back to the first half of 20th century, see for instance [58]. We refer however to the recent [23] for a general overview on the topic. Here the main point is the fact that this semi-distance induces a topology  $\tau_{\mathcal{C}_p}$  with the following properties.

- The topology is generated by the open balls:  $U \subset \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  is open if for every  $\varphi \in U$  there exists  $r > 0$  so that  $B_r(\varphi) := \{\psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H})) : \mathcal{C}_p(\varphi, \psi) < r\} \subset U$ .
- Sequences describe the topology: by [23, Theorem 4.2] a set  $V$  is closed in  $\tau_{\mathcal{C}_p}$  if contains limit points (w.r.t. to  $\mathcal{C}_p$ ) of all the converging sequences entirely contained inside  $V$ .

**Theorem 4.10.** *Let  $\varphi_n, \varphi$  be normal states of  $\mathbf{B}(\mathbf{H})$ . Then*

$$\mathcal{C}_p(\varphi_n, \varphi) \rightarrow 0 \iff \varphi_n \xrightarrow{w^*} \varphi.$$

*Proof.* Suppose first that  $\mathcal{C}_p(\varphi_n, \varphi) \rightarrow 0$  as  $n \rightarrow \infty$ . By Proposition 4.9 there exist  $\mu_{0,n} \in \Lambda_{\varphi_n}^{\perp,1}$ ,  $\mu_{1,n} \in \Lambda_\varphi^{\perp,1}$  and  $\nu_n \in \Pi_{\mathbf{d}}(\mu_{0,n}, \mu_{1,n})$  such that

$$\mathcal{C}_p(\varphi_n, \varphi) = W_p^{\text{Pc}}(\mu_{0,n}, \mu_{1,n}) = \left( \int_{\mathbb{P}_c \times \mathbb{P}_c} \mathbf{d}(P, Q)^p \nu_n(dP dQ) \right)^{1/p} \rightarrow 0$$

By the compactness of  $\Lambda_\varphi^{\perp,1}$  in weak topology,  $\mu_{1,n}$  has a converging subsequence to some  $\mu_1 \in \Lambda_\varphi^{\perp,1}$ . By Lemma 3.4,  $\Psi(\mu_{0,n}) \rightarrow \Psi(\mu_1)$  in  $w^*$ -convergence. By definition  $\Psi(\mu_{0,n}) = \varphi_n$  and  $\Psi(\mu_1) = \varphi$  giving the first claim.

Assume now that  $\varphi_n \rightarrow \varphi$ . By the Proposition 4.9 there exist  $\mu_{0,n} \in \Lambda_{\varphi_n}^{\perp,1}$ ,  $\mu_{1,n} \in \Lambda_\varphi^{\perp,1}$  and  $\nu_n \in \Pi_{\mathbf{d}}(\mu_{0,n}, \mu_{1,n})$  such that

$$\mathcal{C}_p(\varphi_n, \varphi) = W_p^{\text{Pc}}(\mu_{0,n}, \mu_{1,n}).$$

We now invoke Proposition 3.8: from  $\varphi_n \rightarrow \varphi$  we deduce the existence of a subsequence  $\mu_{0,n_k}$  and  $\mu_1 \in \Lambda_\varphi^{\perp,1}$  such that  $\mu_{0,n_k} \rightarrow \mu_1$ . Then  $\mu_{0,n_k} \rightarrow \mu_1$  also in the Wasserstein distance over  $\mathbb{P}_1$ . Hence

$$\mathcal{C}_p(\varphi_n, \varphi) = W_p^{\text{Pc}}(\mu_{0,n}, \mu_{1,n}) \leq W_p^{\text{Pc}}(\mu_{0,n_k}, \mu_1) \rightarrow 0.$$

□

Theorem 4.10 together with [23, Theorem 4.2] imply the following.

**Corollary 4.11.** *The topology  $\tau_{\mathcal{C}_p}$  over the set of normal states coincide with the  $w^*$ -topology.*

*Proof.* Since the topology is determined by the sequences, Theorem 4.10 together with the metrizability of the  $w^*$ -topology over the bounded sets proves the claim. □

**Remark 4.12.** Concerning the triangular inequality for the cost  $\mathcal{C}_p$ , using Remark 4.6, one can deduce the following property: given  $\varphi, \psi, \phi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ , then:

$$\mathcal{C}_p(\varphi, \psi) \leq \inf_{\mu_0, \mu_1, \mu_2} \left\{ W_p^{\text{Pc}}(\mu_0, \mu_1) + W_p^{\text{Pc}}(\mu_1, \mu_2) \right\};$$

the infimum being taken with respect to  $\mu_0 \in \Lambda_\varphi^{\perp,1}$ ,  $\mu_1 \in \Lambda_\phi^{\perp,1}$  and  $\mu_2 \in \Lambda_\psi^{\perp,1}$ .

We do not present a proof of the previous inequality because it follows from a classical argument (gluing) in optimal transport that will be also used in the proof of the following Lemma 4.15. Moreover, whenever the intermediate normal state, say  $\phi$ , has density matrix with only simple eigenvalues (so that there is only one element in  $\Lambda_\phi^{\perp,1}$ ), then again by gluing, one obtains the triangular inequality:

$$\mathcal{C}_p(\varphi, \psi) \leq \mathcal{C}_p(\varphi, \phi) + \mathcal{C}_p(\phi, \psi).$$

The proof of Lemma 4.15 will clarify this point.

**Corollary 4.13.** *The triangular inequality for  $\mathcal{C}_p$  holds for the states with only simple eigenvalues, an open dense set inside  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ .*

### 4.3 The Wasserstein distance in $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$

Even though the cost functional  $\mathcal{C}_p$  is fully satisfactory (see Theorem 4.10), for completeness we address the issue of the lack of the triangular inequality for  $\mathcal{C}_p$ . Using the spectral decomposition without repetitions of eigenvalues permits to obtain the triangular inequality. As a drawback this produces an extended distance (not finite).

**Definition 4.14.** For any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  and  $p \geq 1$  define their  $p$ -Wasserstein distance by

$$W_p(\varphi, \psi) := W_p^{P_c}(\text{tr}(\cdot)\Phi(\varphi), \text{tr}(\cdot)\Phi(\psi)), \quad (4.6)$$

with the map  $\Phi$  defined in (3.7). Recall that by Definition 4.2, if no admissible transport plans exist, we assign to  $W_p(\varphi, \psi)$  the value  $+\infty$ .

We will now prove indeed that the map

$$W_p : \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \times \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \longrightarrow [0, \infty]$$

defines an extended distance over  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ . As before, the symmetry of  $\mathbf{d}$  implies the symmetry of  $W_p$  and if  $W_p(\varphi, \psi) = 0$ , it is straightforward to check that  $\varphi = \psi$ . The triangular inequality is the content of the following

**Lemma 4.15.** (*Triangular inequality for  $W_p$* ). Let  $\varphi, \psi$  and  $\phi$  be three elements of  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ . Then

$$W_p(\phi, \varphi) \leq W_p(\phi, \psi) + W_p(\psi, \varphi).$$

*Proof.* Consider two optimal plans  $\nu_1 \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\phi, \text{tr}(\cdot)\mu_\psi)$  and  $\nu_2 \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\psi, \text{tr}(\cdot)\mu_\varphi)$  whose existence is assured by Theorem 4.3.

If

$$\text{tr}(\cdot)\mu_\phi = \sum_i \alpha_i \delta_{P_{V_i}}, \quad \text{tr}(\cdot)\mu_\psi = \sum_i \beta_i \delta_{P_{W_i}} \quad \text{and} \quad \text{tr}(\cdot)\mu_\varphi = \sum_i \gamma_i \delta_{P_{Z_i}},$$

then the transport plans  $\nu_1$  and  $\nu_2$  can be written as

$$\nu_1 = \sum_{i,j} \alpha_{i,j}^1 \delta_{P_{V_i}} \otimes \delta_{P_{W_j}}, \quad \nu_2 = \sum_{i,j} \alpha_{i,j}^2 \delta_{P_{W_i}} \otimes \delta_{P_{Z_j}},$$

with  $\alpha_{i,j}^1, \alpha_{i,j}^2 \geq 0$  and  $\sum_{i,j} \alpha_{i,j}^1 = \sum_{i,j} \alpha_{i,j}^2 = 1$ . Moreover the marginal constraints are in the form:

$$\alpha_i = \sum_j \alpha_{i,j}^1, \quad \sum_i \alpha_{i,j}^1 = \beta_j = \sum_i \alpha_{j,i}^2, \quad \gamma_j = \sum_i \alpha_{i,j}^2.$$

Following the classical gluing procedure of transport plans, we define

$$\Theta := \sum_{i,j,k} \frac{\alpha_{i,j}^1 \alpha_{j,k}^2}{\beta_j} \delta_{P_{V_i}} \otimes \delta_{P_{W_j}} \otimes \delta_{P_{Z_k}}.$$

One can check that  $\nu_1 = (\pi_{12})_{\#} \Theta$ ,  $\nu_2 = (\pi_{23})_{\#} \Theta$  and  $(\pi_{13})_{\#} \Theta \in \Pi(\text{tr}(\cdot)\mu_\phi, \text{tr}(\cdot)\mu_\varphi)$ .

We also need to check that  $(\pi_{13})_{\#} \Theta$  is admissible: for  $(\pi_{13})_{\#} \Theta$ -a.e.  $P, Q$  it holds  $\dim(R(P)) = \dim(R(Q))$ . This immediately follows by the gluing procedure because for  $\Theta$ -a.e. the projections  $P, Q$  and  $V$  belong to the same connected component of  $P_c$ . By the triangular inequality applied on these connected components, we have for every  $p \geq 1$ :

$$\begin{aligned} W_p(\phi, \varphi) &\leq \left( \int \mathbf{d}(P, Q)^p (\pi_{13})_{\#} \Theta(dPdQ) \right)^{1/p} \\ &= \left( \int \mathbf{d}(P, Q)^p \Theta(dPdVdQ) \right)^{1/p} \\ &\leq \left( \int (\mathbf{d}(P, V) + \mathbf{d}(V, Q))^p \Theta(dPdVdQ) \right)^{1/p} \\ &\leq \left( \int \mathbf{d}(P, V)^p \Theta(dPdVdQ) \right)^{1/p} + \left( \int \mathbf{d}(V, Q)^p \Theta(dPdVdQ) \right)^{1/p} \\ &= W_p(\phi, \psi) + W_p(\psi, \varphi), \end{aligned}$$

concluding the proof.  $\square$

**Corollary 4.16.** The couple  $(\mathcal{S}_n(\mathbf{B}(\mathbf{H})), W_p)$  is an extended metric space in the sense that

$$W_p : \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \times \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \rightarrow [0, \infty]$$

verifies for any  $\varphi$  and  $\psi$  the following properties:  $W_p(\varphi, \varphi) = 0$ , and if  $W_p(\varphi, \psi) = 0$  then  $\varphi = \psi$ ;  $W_p(\varphi, \psi) = W_p(\psi, \varphi)$  and the triangular inequality holds true.

By definition it is straightforward to check that

$$\mathcal{C}_p(\varphi, \psi) \leq W_p(\varphi, \psi).$$

In particular the  $W_p$ -convergence implies the  $\mathcal{C}_p$ -convergence and, by Theorem 4.10, the  $w^*$ -convergence. However, as expected the  $w^*$ -convergence does not implies the  $W_p$ -convergence. We have a simple counterexample.

**Example 2.** Consider the case of  $\mathbf{H} = \mathbb{C}^2$  and

$$\varphi_n := \left(\frac{1}{2} - \frac{1}{n}\right) |e_1\rangle\langle e_1| + \left(\frac{1}{2} + \frac{1}{n}\right) |e_2\rangle\langle e_2| \rightarrow \varphi := \frac{1}{2} (|e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|).$$

The corresponding measures over the space of projections of  $\mathbb{C}^2$  will be

$$\mu_{\varphi_n} = \left(\frac{1}{2} - \frac{1}{n}\right) \delta_{P_1} + \left(\frac{1}{2} + \frac{1}{n}\right) \delta_{P_2}, \quad \mu_\varphi = \delta_{\text{Id}},$$

where  $P_1$  and  $P_2$  project onto the span of  $e_1$  and  $e_2$ , respectively. Since  $P_1, P_2$  and  $\text{Id}$  belong to two different connected components of  $\mathbf{P}_c$ , it holds  $W_p(\text{tr}(\cdot)\mu_{\varphi_n}, \text{tr}(\cdot)\mu_\varphi) = \infty$ .

## 5 Kantorovich duality for $W_p$ and consequences

In this part we will go through the Kantorovich duality for the optimal transport problem over  $(\mathbf{P}_c, \mathbf{d})$ . In particular we will analyse cyclically monotone sets and solutions of the dual problem. The duality will always be referred to the Wasserstein distance  $W_p$ .

### 5.1 Kantorovich duality

As before, when dealing with optimal transport arguments, we will repeatedly restrict  $\mathbf{d}$  to each connected component  $\mathbf{P}_n$  of  $\mathbf{P}_c$  where we rely on the classical results. We start recalling the following classical definition from the theory of optimal transport: A subset  $\Gamma$  of  $\mathbf{P}_c \times \mathbf{P}_c$  is  $\mathbf{d}^p$ -cyclically monotone if and only if for any  $n \in \mathbb{N}$  and  $(P_1, Q_1), \dots, (P_n, Q_n) \in \Gamma$  the inequality

$$\sum_{i \leq n} \mathbf{d}(P_i, Q_i)^p \leq \sum_{i \leq n} \mathbf{d}(P_i, Q_{i+1})^p,$$

is valid with the convention  $Q_{n+1} = Q_1$ . It is also tacitly assumed that for each  $(P, Q) \in \Gamma$ ,  $\dim R(P) = \dim R(Q)$ . Accordingly, given  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ , a plan  $\nu \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$  will be called  $\mathbf{d}^p$ -cyclically monotone if there exists a  $\mathbf{d}^p$ -cyclically monotone set  $\Gamma$  such that  $\pi(\Gamma) = 1$ .

By the lower semicontinuity of  $\mathbf{d}$ , it is well-known that  $\mathbf{d}^p$ -cyclical monotonicity is a necessary condition for the optimality (see for instance [13, Proposition B.16]).

**Proposition 5.1.** *Let  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  be given and  $p \geq 1$ . Then any optimal transport plan  $\nu \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$  for the  $W_p$  distance is  $\mathbf{d}^p$ -cyclically monotone.*

Looking at the transport on each single connected component of  $\mathbf{P}_c$ , it is clear that cyclical monotonicity is indeed a sufficient condition for global optimality.

**Proposition 5.2.** *Let  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  be given and let  $\nu \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$  be any  $\mathbf{d}^p$ -cyclically monotone transport plan. Then  $\nu$  is  $W_p$ -optimal, i.e.*

$$\int \mathbf{d}(P, Q)^p \nu(dP dQ) = W_p(\varphi, \psi)^p.$$

*Proof.* Decompose both  $\mu_\varphi$  and  $\mu_\psi$  into a sum of measures supported on each connected component  $\mathbf{P}_n = \{P \in \mathbf{P}_c : \text{tr}(P) = n\} \subset \mathbf{P}_c$  (i.e. their restrictions). Then

$$\mu_\varphi = \sum_n \mu_{\varphi, n}, \quad \mu_\psi = \sum_n \mu_{\psi, n},$$

with  $\mu_{\varphi, n}$  and  $\mu_{\psi, n}$  supported inside  $\mathbf{P}_n$ . Then any plan  $\nu \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$  has to send  $\mu_{\varphi, n}$  to  $\mu_{\psi, n}$  and its optimality is equivalent to optimality between each  $\mu_{\varphi, n}$  and  $\mu_{\psi, n}$ .

Let us now consider  $\nu \in \Pi_{\mathbf{d}}(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$  and  $\Gamma$  a  $\mathbf{d}^p$ -cyclically monotone set with  $\nu(\Gamma) = 1$ . We decompose as above  $\nu = \sum \nu_n$  with  $\nu_n \perp \nu_m$  if  $n \neq m$  and  $\nu_n$  having marginals  $\mu_{\varphi, n}$  and  $\mu_{\psi, n}$ . Here  $\nu_n \perp \nu_m$  is in the sense of measure theory i.e. with disjoint supports. Then  $\mathbf{d}$  restricted to  $\mathbf{P}_n$  is finite and therefore, by the classical theory of optimal transport (see for instance [57]),  $\mathbf{d}^p$ -cyclical monotonicity is equivalent to optimality giving that each  $\nu_n$  is optimal. The optimality of  $\nu$  follows.  $\square$

From the classical theory [57, Theorem 5.10], the following dual formulation of the problem is valid: for any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$

$$\begin{aligned} & \min_{\nu \in \Pi_d(\text{tr}(\cdot) \mu_\varphi, \text{tr}(\cdot) \mu_\psi)} \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \pi(dPdQ) \\ &= \sup_{\substack{f, g \in C_b(\mathbf{P}_c) \\ g(Q) - f(P) \leq \mathbf{d}^p(P, Q)}} \left( \int g(Q) \text{tr}(Q) \mu_\psi(dQ) - \int f(P) \text{tr}(P) \mu_\varphi(dP) \right) \end{aligned}$$

The right hand side can actually be substituted with some special couples of functions.

**Definition 5.3** ( $\mathbf{d}^p$ -convex function). A function  $f : \text{supp}(\mu_\varphi) \rightarrow \mathbb{R} \cup \{+\infty\}$  is  $\mathbf{d}^p$ -convex if it is not identically  $+\infty$  and there exists  $h : \text{supp}(\mu_\psi) \rightarrow \mathbb{R} \cup \{\pm\infty\}$  such that for each  $P \in \text{supp}(\mu_\varphi)$

$$f(P) = \sup_{Q \in \text{supp}(\mu_\psi)} h(Q) - \mathbf{d}(P, Q)^p.$$

Then its  $\mathbf{d}^p$ -transform is a function  $f^{\mathbf{d}^p} : \text{supp}(\mu_\psi) \rightarrow \mathbb{R}$  defined for each  $Q \in \text{supp}(\mu_\psi)$  by:

$$f^{\mathbf{d}^p}(Q) := \inf_{P \in \text{supp}(\mu_\varphi)} f(P) + \mathbf{d}(P, Q)^p.$$

Theorem 5.10 of [57] permits to rewrite the previous duality as

$$\min_{\nu \in \Pi_d(\text{tr}(\cdot) \mu_\varphi, \text{tr}(\cdot) \mu_\psi)} \int_{\mathbf{P}_c \times \mathbf{P}_c} \mathbf{d}(P, Q)^p \pi(dPdQ) = \sup_{f \in L^1(\text{tr}(\cdot) \mu_\varphi)} \left( \int f^{\mathbf{d}^p} \text{tr}(\cdot) \mu_\psi - \int f \text{tr}(\cdot) \mu_\varphi \right).$$

In the above supremum one might as well impose  $f$  to be  $\mathbf{d}^p$ -convex. This supremum is actually achieved and the maximum will be called a Kantorovich potential.

**Theorem 5.4.** *Given any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  with  $W_p(\varphi, \psi) < \infty$ , there exists  $f \in L^1(\text{tr}(\cdot) \mu_\varphi)$  and  $\mathbf{d}^p$ -convex such that*

$$W_p(\varphi, \psi)^p = \int f^{\mathbf{d}^p} \text{tr}(\cdot) \mu_\psi - \int f \text{tr}(\cdot) \mu_\varphi.$$

In particular,  $\nu \in \Pi_d(\text{tr}(\cdot) \mu_\varphi, \text{tr}(\cdot) \mu_\psi)$  is  $W_p$ -optimal if and only if

$$\nu\left(\{(P, Q) \in \mathbf{P}_c \times \mathbf{P}_c : f^{\mathbf{d}^p}(Q) - f(P) = \mathbf{d}(P, Q)^p\}\right) = 1.$$

*Proof.* Reasoning like in the proof of Proposition 5.2, on each connected component  $\mathbf{P}_n$  of  $\mathbf{P}_c$ , the metric  $\mathbf{d}$  is continuous yielding (see [57, Theorem 5.10]) for each  $n \in \mathbb{N}$  the existence of  $\mathbf{d}^p$ -convex functions  $f_n : \text{supp}(\mu_{\varphi, n}) \rightarrow \mathbb{R}$  together with a sequence  $h_n : \text{supp}(\mu_{\psi, n}) \rightarrow \mathbb{R} \cup \{\pm\infty\}$  such that for each  $P \in \text{supp}(\mu_{\varphi, n})$  it holds:

$$f_n(P) = \sup_{Q \in \text{supp}(\mu_{\psi, n})} h_n(Q) - \mathbf{d}(P, Q)^p.$$

Moreover a transport plan between  $\mu_{\varphi, n}$  and  $\mu_{\psi, n}$  is optimal if and only if it is concentrated inside the following  $\mathbf{d}^p$ -cyclically monotone set:

$$\left\{ (P, Q) \in \text{supp}(\mu_{\varphi, n}) \times \text{supp}(\mu_{\psi, n}) : f_n^{\mathbf{d}^p}(Q) - f_n(P) = \mathbf{d}(P, Q)^p \right\},$$

where  $f_n^{\mathbf{d}^p}$  is defined considering the infimum only among those  $P \in \text{supp}(\mu_{\varphi, n})$ . In particular,

$$\int_{\text{supp}(\mu_{\varphi, n}) \times \text{supp}(\mu_{\psi, n})} \mathbf{d}(P, Q)^p \nu(dPdQ) = \int f_n^{\mathbf{d}^p} \text{tr}(\cdot) \mu_{\psi, n} - \int f_n \text{tr}(\cdot) \mu_{\varphi, n}.$$

Define then  $f(P) := f_n(P)$  and  $h(Q) := h_n(Q)$  for each  $P \in \text{supp}(\mu_{\varphi, n})$  and  $Q \in \text{supp}(\mu_{\psi, n})$  and notice that

$$f(P) = \sup_{Q \in \text{supp}(\mu_\psi)} h(Q) - \mathbf{d}(P, Q)^p.$$

This proves that  $f$  is  $\mathbf{d}^p$ -convex. Simply noticing that  $\mathbf{d}$  takes value  $+\infty$  if  $P$  and  $Q$  does not belong to the same connected component of  $\mathbf{P}_c$ , it follows that for  $Q \in \text{supp}(\mu_{\psi, n})$  we have  $f^{\mathbf{d}^p}(Q) = f_n^{\mathbf{d}^p}(Q)$ , where  $f_n^{\mathbf{d}^p}$  is its  $\mathbf{d}^p$ -transform, given by the formula

$$f_n^{\mathbf{d}^p}(Q) := \inf_{P \in \text{supp}(\mu_{\varphi, n})} f_n(P) + \mathbf{d}(P, Q)^p.$$

Hence  $W_p(\varphi, \psi)^p = \int f^{\mathbf{d}^p} \text{tr}(\cdot) \mu_\psi - \int f \text{tr}(\cdot) \mu_\varphi$ , and the second claim is straightforward.  $\square$

We now focus on representing Kantorovich potentials.

**Lemma 5.5.** *For any  $f \in L^1(\text{tr}(\cdot)\mu_\varphi)$ , there exists an unbounded linear and densely defined operator  $C$  such that  $C\rho_\varphi \in \mathcal{L}^1(\mathbf{H})$  (the composition extends from its dense domain to a bounded operator) and*

$$\text{tr}(P)f(P) = \text{tr}(CP), \quad P \in \text{supp}(\mu_\varphi).$$

*Proof.* Let  $\sum_i \lambda_i P_{V_i}$  be the spectral decomposition of  $\rho_\varphi$  with strictly decreasing eigenvalues. Then  $f$  defines a Borel function on the spectrum of  $\rho_\varphi$  with  $f(\lambda_i) := f(P_i)$  and possibly modify  $f$  (without changing notation) in a way that  $f(0) = 0$ . We simply define  $C := f(\rho_\varphi)$  by the functional calculus. In particular

$$\text{Dom}(C) = \left\{ x \in \mathbf{H} : \sum_i |f(P_{V_i})|^2 \|P_{V_i}x\|^2 < \infty \right\}$$

is of course dense. The rest is straightforward noticing that the condition  $f \in L^1(\text{tr}(\cdot)\mu_\varphi)$  implies that the sequence  $(\lambda_i f(P_i))_i$  is bounded.  $\square$

**Corollary 5.6.** *Given any  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  with  $W_p(\varphi, \psi) < \infty$ , there exist  $C$  and  $C_p^d$  unbounded linear and densely defined operators on  $\mathbf{H}$  such that the following points are verified.*

1. *The  $W_p$ -cost verifies  $W_p(\varphi, \psi) = \text{tr}(C_p^d \rho_\psi) - \text{tr}(C\rho_\varphi)$ .*
2. *Any  $\nu \in \Pi_d(\mu_\varphi, \mu_\psi)$  is  $W_p$ -optimal if and only if*

$$\nu \left( \left\{ (P, Q) \in \text{supp}(\mu_\varphi) \times \text{supp}(\mu_\psi) : \text{tr}(C_p^d - \text{tr}(CP)) = \frac{d(P, Q)^p}{\text{tr}(P)} \right\} \right) = 1;$$

where  $C, C_p^d$  are such that

$$\text{tr}(C_p^d Q) - \text{tr}(CP) \leq \frac{d(P, Q)^p}{\text{tr}(P)}, \quad \forall (P, Q) \in \text{supp}(\mu_\varphi) \times \text{supp}(\mu_\psi), \quad \text{tr}(P) = \text{tr}(Q).$$

*Proof.* To prove the first point we use Theorem 5.4 to deduce the existence of a solution  $f \in L^1(\text{tr}(\cdot)\mu_\varphi)$  of the dual problem with

$$W_p(\mu_\varphi, \mu_\psi) = \int f^{c_p} \text{tr}(\cdot) \mu_\psi - \int f \text{tr}(\cdot) \mu_\varphi.$$

Then apply Lemma 5.5 to  $f$  to obtain  $C$  such that  $f(P)\text{tr}(P) = \text{tr}(CP)$  for all  $P \in \text{supp}(\mu_\varphi)$ . This implies

$$\int f \text{tr}(\cdot) \mu_\varphi = \text{tr}(C\rho_\varphi).$$

Denoting with  $C_p^d$  any linear map representing  $f^{d_p}$ , the first point follows. The second point is then a reformulation of the second point of Theorem 5.4.  $\square$

## 5.2 Wasserstein geodesics

In this section and in the following one we will study how to match two other possible approaches in defining a Wasserstein type distance over normal states with the one we introduced in Section 4.

The geodesic structure of  $\mathbf{P}_c$  will permit to investigate the geodesic structure of  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ . We begin by recalling the classical definition of geodesic adapted to the setting of  $\mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ .

**Definition 5.7.** Given  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ , a curve

$$[0, 1] \ni t \mapsto \phi_t \in \mathcal{S}_n(\mathbf{B}(\mathbf{H})), \quad \phi_0 = \varphi, \quad \phi_1 = \psi,$$

is a  $C_p$ -geodesic (resp. a  $W_p$ -geodesic) if  $C_p(\phi_t, \phi_s) = |t - s|C_p(\varphi, \psi)$  (resp.  $W_p(\phi_t, \phi_s) = |t - s|W_p(\varphi, \psi)$ ), for any  $s, t \in [0, 1]$ .

We start looking for geodesic convexity of suitable subsets of  $\mathcal{P}(\mathbf{P}_c)$ .

**Proposition 5.8.** *The set  $\mathcal{D}_1(\mathbf{P}_c)$  of discrete, non-negative measures having integral of the trace equal to 1 as defined in (3.3) is weakly convex with respect to  $W_p^{P_c}$  in the following sense: for any  $\mu_0, \mu_1 \in \mathcal{D}_1(\mathbf{P}_c)$  such that  $W_p^{P_c}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1) < \infty$ , there exists a curve  $(\mu_t)_{t \in [0, 1]} \subset \mathcal{D}_1(\mathbf{P}_c)$  with initial point  $\mu_0$  and final point  $\mu_1$  such that  $t \mapsto \text{tr}(\cdot)\mu_t$  is a  $W_p^{P_c}$ -geodesic.*

*Proof.* Given  $\mu_0, \mu_1 \in \mathcal{D}_1(\mathbb{P}_c)$  such that  $W_p^{\mathbb{P}_c}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1) < \infty$ , Theorem 4.3 ensures the existence of an optimal transport plan  $\nu \in \Pi_d(\mu_0, \mu_1)$ . If  $\mu_0 = \sum_i \alpha_i \delta_{P_{V_i}}$  and  $\mu_1 = \sum_i \beta_i \delta_{P_{Z_i}}$ , there exist non-negative coefficients  $\theta_{i,j}$  such that

$$\nu = \sum_{i,j} \theta_{i,j} \delta_{P_{V_i}} \otimes \delta_{P_{Z_j}}, \quad \sum_j \theta_{i,j} = \alpha_i \text{tr}(P_{V_i}), \quad \sum_i \theta_{i,j} = \beta_j \text{tr}(P_{Z_j}).$$

Since  $\nu$  is admissible, whenever  $\theta_{i,j} > 0$  it follows that  $\text{tr}(P_{V_i}) = \text{tr}(P_{Z_j})$  hence we can select a family (any) of geodesics  $\gamma_{i,j}$  in  $(\mathbb{P}_c, \mathbf{d})$  connecting  $V_i$  to  $Z_j$ . Its existence is assured by the fact that  $V_i$  and  $Z_j$  belong to the same connected component of  $(\mathbb{P}_c, \mathbf{d})$ . In particular  $\text{tr}(\gamma_{i,j}(t))$  is constant for each  $t \in [0, 1]$  and depends only on  $i$ .

Now define the following non-negative measure

$$\gamma := \sum_{i,j} \frac{\theta_{i,j}}{\text{tr}(P_{V_i})} \delta_{\gamma_{i,j}}$$

over  $\text{Geo}(\mathbb{P}_c)$ . Denoting by  $e_t : \text{Geo}(\mathbb{P}_c) \rightarrow \mathbb{P}_c$  the evaluation map at time  $t$  we have a curve of measures  $[0, 1] \ni t \mapsto \mu_t := (e_t)_\#(\gamma)$ . First notice that  $\mu_t \in \mathcal{D}_1(\mathbb{P}_c)$ : indeed  $\gamma$  is a discrete measure; therefore the same is valid for  $\mu_t$  and

$$\int_{\mathbb{P}_c} \text{tr}(P) \mu_t(dP) = \int_{\mathbb{P}_c} \text{tr}(P) (e_t)_\#(\gamma)(dP) = \sum_{i,j} \theta_{i,j} = 1.$$

Hence  $\mu_t \in \mathcal{D}_1(\mathbb{P}_c)$  and finally

$$\begin{aligned} W_p^{\mathbb{P}_c}(\text{tr}(\cdot)\mu_s, \text{tr}(\cdot)\mu_t)^p &\leq \int \mathbf{d}(P, Q)^p ((e(s), e(t))_\#(\sum_{i,j} \theta_{i,j} \delta_{\gamma_{i,j}}))(dP dQ) \\ &= |s - t|^p \int \mathbf{d}(P, Q)^p ((e(0), e(1))_\#(\sum_{i,j} \theta_{i,j} \delta_{\gamma_{i,j}}))(dP dQ) \\ &= |s - t|^p \int \mathbf{d}(P, Q)^p \nu(dP dQ) \\ &= |s - t|^p W_p^{\mathbb{P}_c}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1)^p. \end{aligned}$$

This proves the claim.  $\square$

To obtain a Wasserstein geodesic between normal states, Proposition 5.8 must be reinforced with the additional assumption that  $\mu_t \in \mathcal{D}_1^\perp(\mathbb{P}_c)$ . The condition  $\mu_t \in \mathcal{D}_1^\perp(\mathbb{P}_c)$  is actually quite demanding and has strong and rigid consequences on the two measures  $\mu_\varphi, \mu_\psi$  at the extremes.

Recall the definition (3.4) of  $\mathcal{D}_1^\perp(\mathbb{P}_c)$  consisting of discrete measures supported on orthogonal families of projections and integrating the trace to one.

**Proposition 5.9.** *Given  $\mu_0, \mu_1 \in \mathcal{D}_1^\perp(\mathbb{P}_c)$  such that with  $W_p^{\mathbb{P}_c}(\mu_0, \mu_1) < \infty$ . Let  $\mu_t$  be any  $W_p^{\mathbb{P}_c}$ -geodesic provided from Proposition 5.8 and assume  $\mu_t \in \mathcal{D}_1^\perp(\mathbb{P}_c)$  for all  $t \in [0, 1]$ .*

*Then there exists a bijective map  $T : \text{supp}(\mu_0) \rightarrow \text{supp}(\mu_1)$  such that  $(Id, T)_\# \mu_0 \in \Pi_d(\mu_0, \mu_1)$  is an optimal plan. In particular, if  $\mu_0 = \sum_i \alpha_i P_{0,i}$ ,  $\mu_1 = \sum_i \beta_i P_{1,i}$  satisfy  $\alpha_i > \alpha_{i+1}$  and  $\beta_i > \beta_{i+1}$ , then*

$$\text{tr}(P_{0,i}) = \text{tr}(P_{1,i}), \quad \alpha_i = \beta_i,$$

and  $T(P_{0,i}) = P_{1,i}$ .

*Proof.* It follows from the classical theory of optimal transport applied to each connected component of  $\mathbb{P}_c$  that  $\text{tr}(\cdot)\mu_t = (e_t)_\#(\gamma)$  with  $\gamma \in \mathcal{P}(\text{Geo}(\mathbb{P}_c))$ . Hence  $(e_0, e_1)_\# \gamma \in \Pi_d(\text{tr}(\cdot)\mu_\varphi, \text{tr}(\cdot)\mu_\psi)$ . Put  $\nu = (e_0, e_1)_\# \gamma$ ; then necessarily

$$\nu = \sum_{i,j} \theta_{i,j} \delta_{P_{V_i}} \otimes \delta_{P_{Z_j}}, \quad \sum_{i,j} \theta_{i,j} = 1, \quad \theta_{i,j} \geq 0.$$

Assume now by contradiction that there exist  $i_1 \neq i_2$  and  $j \in \mathbb{N}$  such that both  $\theta_{i_1,j}, \theta_{i_2,j} > 0$ . Then there exist  $\gamma_{i_1,j}, \gamma_{i_2,j} \in \text{Geo}(\mathbb{P}_c)$  such that

$$\gamma_{i_1,j}(t), \gamma_{i_2,j}(t) \in \text{supp}(\mu_t), \quad \gamma_{i_1,j}(0) = P_{V_{i_1}}, \gamma_{i_2,j}(0) = P_{V_{i_2}}, \quad \gamma_{i_1,j}(1) = P_{Z_j}, \gamma_{i_2,j}(1) = P_{Z_j},$$

with  $P_{V_{i_1}} \perp P_{V_{i_2}}$ . Then  $\mu_t \in \mathcal{D}_1^\perp(\mathbb{P}_c)$  implies that either  $\gamma_{i_1,j}(t) = \gamma_{i_2,j}(t)$  or  $\text{tr}(\gamma_{i_1,j}(t)\gamma_{i_2,j}(t)) = 0$  with the former verified at  $t = 0$  and the latter at  $t = 1$ . Now the continuity of the map  $t \mapsto \gamma_{i_1,j}(t), \gamma_{i_2,j}(t)$  gives a contradiction.

The argument can be reverted and implies that for each  $i \in \mathbb{N}$  there is a unique  $j \in \mathbb{N}$  such that  $\theta_{i,j} > 0$  and for each  $j \in \mathbb{N}$  there is a unique  $i \in \mathbb{N}$  such that  $\theta_{i,j} > 0$ . This is equivalent to the existence of a bijection  $T : \text{supp}(\mu_\varphi) \rightarrow \text{supp}(\mu_\psi)$  such that

$$\nu = (\text{Id} \times T)_\# \mu_0.$$

The first part of the claim is proved. The remaining claims are straightforward consequences.  $\square$

**Corollary 5.10.** *Fix  $p \geq 1$ . Given  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  with  $W_p(\varphi, \psi) < \infty$ , consider the measures  $\mu_\varphi, \mu_\psi$  and any curve  $\mu_t$  from Proposition 5.8. If  $\mu_t \in \mathcal{D}_1^\perp(\mathbf{P}_c)$ , then defining*

$$\rho_t := \Psi(\mu_t) \in \mathcal{C}(\mathbf{H}),$$

*the corresponding curve of normal states  $[0, 1] \ni t \mapsto \varphi_{\rho_t}$  is a  $W_p$ -geodesic.*

*Proof.* To fix the notation, the  $\mu_t$  from Proposition 5.8 can be written as  $\mu_t = \sum_i \alpha_i \delta_{P_i(t)}$ . Then using the notations of Section 3,  $\Psi(\mu_t) = \sum_i \alpha_i P_i(t)$  is a well-defined element of  $\mathcal{C}(\mathbf{H})$ . From Proposition 5.9 it follows that

$$\Phi(\varphi_{\rho_t}) = \Phi(\Psi(\mu_t)) = \mu_t.$$

Notice indeed that  $\mu_t$  is an element of  $\mathcal{D}_1^\perp(\mathbf{P}_c)$  giving different weights on each element of its support. Hence, by definition of  $W_p$  (recall (4.6))

$$\begin{aligned} W_p(\varphi_{\rho_s}, \varphi_{\rho_t}) &= W_p^{\mathbf{P}^c}(\text{tr}(\cdot)\Phi(\varphi_{\rho_s}), \text{tr}(\cdot)\Phi(\varphi_{\rho_t})) \\ &= W_p^{\mathbf{P}^c}(\text{tr}(\cdot)\mu_s, \text{tr}(\cdot)\mu_t) \\ &= |t - s| W_p^{\mathbf{P}^c}(\text{tr}(\cdot)\mu_0, \text{tr}(\cdot)\mu_1) \\ &= |t - s| W_p(\varphi_{\rho_0}, \varphi_{\rho_1}), \end{aligned}$$

proving the claim.  $\square$

**Remark 5.11.** If the condition  $\mu_t \in \mathcal{D}_1^\perp(\mathbf{P}_c)$  is not known, then one can anyway define a curve of normal states because  $\rho_t := \Psi(\mu_t) = \sum_i \alpha_i P_i(t)$  is a well-defined element of  $\mathcal{C}(\mathbf{H})$  implying that  $\varphi_{\rho_t} \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  (see Lemma 2.8). However, the spectral decomposition of  $\rho_t$  will not be given by  $\sum_i \alpha_i P_i(t)$  and

$$\Phi(\varphi_{\rho_t}) = \Phi(\Psi(\mu_t)) \neq \mu_t,$$

and nothing can be deduced on  $W_p(\varphi_{\rho_t}, \varphi_{\rho_s})$ .

**Remark 5.12.** In the proof of Proposition 5.9 the fact that  $\text{tr}(\cdot)\mu_t$  is a  $W_p^{\mathbf{P}^c}$ -geodesic it was not directly used, rather the existence of a  $\gamma \in \mathcal{P}(\text{Geo}(\mathbf{P}_c))$  such that

$$\text{tr}(\cdot)\mu_t = (e_t)_\# \gamma, \quad \mu_t \in \mathcal{D}_1^\perp(\mathbf{P}_c).$$

This implies indeed that  $\gamma$  has to be a discrete measure as well and  $t \mapsto \text{tr}(\cdot)\mu_t$  is  $W_p^{\mathbf{P}^c}$ -continuous; this two facts being enough to close the argument.

Proposition 5.9 admits a partial converse.

**Proposition 5.13.** *Let  $\varphi, \psi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  be given states with  $W_p(\varphi, \psi) < \infty$ . If there exists a bijective map  $T : \text{supp}(\mu_\varphi) \rightarrow \text{supp}(\mu_\psi)$  such that  $(\text{Id} \times T)_\# \mu_\varphi \in \Pi_d(\mu_\varphi, \mu_\psi)$  then  $\varphi$  and  $\psi$  are in the same unitary orbit: there is a unitary  $u$  with  $\rho_\psi = u\rho_\varphi u^*$ .*

*Proof.* Let  $\rho_\varphi = \sum_i \lambda_i P_{V_i}$  be the spectral decomposition with distinguished positive eigenvalues  $\lambda_i$ . The condition  $(\text{Id} \times T)_\# \mu_\varphi \in \Pi_d(\mu_\varphi, \mu_\psi)$  implies that the spectral decomposition of  $\rho_\psi$  is:  $\rho_\psi = \sum_i \lambda_i T(P_{V_i})$  with  $P_{V_i} = R(P_{V_i}) \cong R(T(P_{V_i})) = T(P_{V_i})$  (because we have a  $d$ -transport plan). Identifying projections with subspaces we think  $T$  defined on the collection of the eigenspaces of  $\rho_\varphi$ . Every eigenspace is finite dimensional and this implies  $V_i^\perp \cong T(V_i)^\perp$ . Since these are mutually orthogonal we have  $R(\rho_\varphi) \cong R(\rho_\psi)$  and  $\text{Ker}(\rho_\varphi) \cong \text{Ker}(\rho_\psi)$ . By [20, Proposition 7] we find an invertible  $u \in \mathbf{B}(\mathbf{H})$  such that  $\rho_\psi = u\rho_\varphi u^*$ . In our case this  $u$  is unitary. Concretely let  $\{e_i, e'_j\}$  be an orthonormal system adapted to  $\mathbf{H} = R(\rho_\varphi) \oplus N(\rho_\varphi)$  and  $\{f_i, f'_j\}$  a corresponding one for  $\rho_\psi$ ; then

$$u = \sum_i |f_i\rangle\langle e_i| + \sum_j |f'_j\rangle\langle e'_j|.$$

We can do better finding a  $u$  such that  $uP_{V_i}u^* = T(P_{V_i})$  for every  $i$ . It suffices to choose the basis  $\{e_i\}$  adapted to the spectral decomposition  $\rho_\varphi = \sum_i \lambda_i P_{V_i}$ .  $\square$

## 6 Tensor product interpretation: a generalization

As specified in Section 2.2.2, the tensor product Hilbert space  $\mathbf{H} \otimes \mathbf{H}$  corresponds, in quantum mechanics, to a composite system and a natural way to match two normal states  $\varphi, \psi$  of  $\mathbf{B}(\mathbf{H})$  would be via a normal state  $\Xi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H}))$  satisfying the partial trace conditions  $J_b^1 \Xi = \varphi$  and  $J_b^2 \Xi = \psi$ .

In this section we shall reconcile this approach with the one we presented in Sections 3 and 4 based on transport plans between spectral-projections measures.

We begin with some preliminaries following [38], and the appendix B.1 for basics on antilinear operators. Let  $AHS(\mathbf{H})$  be the space of the antilinear Hilbert–Schmidt operators acting on  $\mathbf{H}$ . Firstly an antilinear operator is an additive operator  $T : \mathbf{H} \rightarrow \mathbf{H}$  such that  $T(\lambda x) = \bar{\lambda}x$ . To define the Hilbert–Schmidt ones we begin with the antilinear rank-one operators. They are in the form  $x \mapsto T_{\xi, \eta}(x) := \langle x, \xi \rangle \eta$  for fixed vectors  $\xi, \eta \in \mathbf{H}$ ; in particular we have  $T_{\xi, \eta}^* = T_{\eta, \xi}$ . On such operators we define the Hilbertian product (conjugate-linear in the first entry)  $\langle A, B \rangle := \text{tr}(BA^*)$  and we complete the linear span of all the  $T_{\xi, \eta}$  with respect to this Hilbert structure. We compute

$$\langle T_{\xi, \eta}, T_{x, y} \rangle = \text{tr}(T_{x, y} T_{\eta, \xi}) = \text{tr} \left( \zeta \mapsto \langle \xi, x \rangle \langle \eta, \zeta \rangle y \right) = \langle \xi, x \rangle \langle \eta, y \rangle.$$

The right quantity is exactly the inner product defined on  $\mathbf{H} \otimes \mathbf{H}$  i.e.  $\langle \xi \otimes \eta, x \otimes y \rangle = \langle \xi, x \rangle \langle \eta, y \rangle$ . Indeed we have a  $\mathbb{C}$ -linear isomorphism

$$\Theta : \mathbf{H} \otimes \mathbf{H} \longrightarrow AHS(\mathbf{H}) \quad (6.1)$$

uniquely determined by linearity and continuity on simple tensors by  $\Theta_{\xi \otimes \eta} := T_{\xi, \eta} \in AHS(\mathbf{H})$ . Some basic identities are immediate to prove

$$\Theta_{\xi \otimes \eta}^* = \Theta_{\eta \otimes \xi}, \quad |\Theta_{\zeta}|^2 = \text{Tr}_2 |\zeta\rangle\langle\zeta| \quad \text{and} \quad |\Theta_{\zeta}^*|^2 = \text{Tr}_1 |\zeta\rangle\langle\zeta|, \quad (6.2)$$

for  $\zeta \in \mathbf{H} \otimes \mathbf{H}$ . Let now  $W : \mathbf{H} \rightarrow \mathbf{H}$  be a linear (antilinear) partial isometry with initial space  $R(W^*W)$  and final space  $R(WW^*)$ , then  $W| : R(W^*W) \rightarrow R(WW^*)$  is a unitary (antiunitary) isomorphism. Let  $\mathbf{P}_c(R(W^*W)) \subset \mathbf{P}_c$  and  $\mathbf{P}_c(R(WW^*)) \subset \mathbf{P}_c$  be the corresponding Grassmannians (recall  $\mathbf{P}_c = \mathbf{P}_c(\mathbf{H})$ ). This means that we are identifying

$$\mathbf{P}_c(R(W^*W)) \cong \{P \in \mathbf{P}_c : P \leq W^*W\}$$

by taking the orthogonal complements. The corresponding identification is understood for  $WW^*$ . The adjoint action induces a diffeomorphism

$$\widetilde{W} : \mathbf{P}_c(W^*W) \longrightarrow \mathbf{P}_c(WW^*), \quad \widetilde{W}(P) := \text{Ad}_{W|} P = (W|)P(W|)^*$$

for  $P \in \mathbf{P}_c : P \leq R(W^*W)$ . We define

$$\mathcal{G}(\mathbf{H}) := \left\{ \begin{array}{l} (V, \phi, W) : V, W \subset \mathbf{H} \text{ closed subspaces, } \phi : \mathbf{P}_c(V) \rightarrow \mathbf{P}_c(W) \text{ is a smooth map} \\ \text{preserving the connected components} \end{array} \right\}$$

### 6.1 Pure States as transport maps

Let us consider a pure state  $\omega_{\zeta} \in \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) \subset \mathcal{S}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H}))$  represented by a vector  $\zeta \in \mathbf{H} \otimes \mathbf{H}$  with  $\|\zeta\| = 1$  and reduced density matrices

$$\rho_1 = \text{Tr}_2 |\zeta\rangle\langle\zeta|, \quad \text{and} \quad \rho_2 = \text{Tr}_1 |\zeta\rangle\langle\zeta|.$$

We will associate to  $\omega_{\zeta}$  a unique family of transport plans from the spectral-projection measures of  $\varphi_{\rho_1}$  to the one of  $\varphi_{\rho_2}$ .

We write the polar decomposition (see the appendix B.2) of the antilinear operator  $\Theta_{\zeta}$  associated via (6.1) to  $\zeta$ . Thus  $\Theta_{\zeta} = U_{\zeta} |\Theta_{\zeta}| = |\Theta_{\zeta}^*| U_{\zeta}$  and  $|\Theta_{\zeta}^*| = U_{\zeta} |\Theta_{\zeta}| U_{\zeta}^*$ . By (6.2) we see that  $|\Theta_{\zeta}| = \rho_1^{1/2}$  and  $|\Theta_{\zeta}^*| = \rho_2^{1/2}$ . It follows

$$\Theta_{\zeta} = U_{\zeta} \rho_1^{1/2} = \rho_2^{1/2} U_{\zeta} \quad \text{and} \quad \rho_2 = U_{\zeta} \rho_1 U_{\zeta}^*. \quad (6.3)$$

The antilinear partial isometry  $U_{\zeta} : \mathbf{H} \rightarrow \mathbf{H}$  is called the *correlation operator* and restricts to an antiunitary isomorphism  $\overline{R}(\rho_1) \xrightarrow{\cong} \overline{R}(\rho_2)$ . This operator is uniquely specified if we add one of the following equivalent conditions

$$N(U_{\zeta}) = N(\rho_1) \quad \text{and} \quad U_{\zeta}^* U_{\zeta} \mathbf{H} = N(\rho_1)^{\perp},$$

that we will always consider being satisfied.

We have associated to  $\zeta \in \mathbf{H} \otimes \mathbf{H}$  its marginals and an antilinear partial isometry intertwining them. In the following we won't use the map  $\Upsilon$  in the next proposition, rather some kind of its measure theory version.

**Proposition 6.1.** *The following map is well defined*

$$\Upsilon : \mathcal{PS}_n(\mathbf{H} \otimes \mathbf{H}) \longrightarrow \mathcal{G}(\mathbf{H}), \quad \omega_\zeta \longmapsto \left( R(U_\zeta^* U_\zeta), \widetilde{U}_\zeta, R(U_\zeta U_\zeta^*) \right).$$

It has the property  $\Upsilon(\omega_\zeta) = \Upsilon(\omega_\eta) \implies U_\zeta = U_\eta$  up to a phase i.e.  $U_\zeta = \lambda U_\eta$  for some  $\lambda \in \mathbf{U}(1)$ .

*Proof.* If  $\zeta$  is changed into  $\lambda\zeta$  for a phase  $\lambda \in \mathbf{U}(1)$  then  $U_{\lambda\zeta} = \lambda U_\zeta$  and  $\widetilde{U}_{\lambda\zeta} = \widetilde{U}_\zeta$ . The map is well defined at the states level. Assume now that  $\Upsilon(\omega_\zeta) = \Upsilon(\omega_\eta)$ ; then  $U_\zeta$  and  $U_\eta$  have the same initial and final space. Let  $x \in R(U_\zeta^* U_\zeta)$  be a unit vector. Evaluating on the rank-one projections

$$\widetilde{U}_\zeta(|x\rangle\langle x|) = |U_\zeta x\rangle\langle U_\zeta x| = \widetilde{U}_\eta(|x\rangle\langle x|) = |U_\eta x\rangle\langle U_\eta x|.$$

Evaluate again on the vector  $U_\zeta x$  to obtain  $U_\zeta x = \langle U_\eta | U_\zeta x \rangle U_\eta x$ . By computing the norm we find  $|\langle U_\eta x | U_\zeta x \rangle| = 1$ . Cauchy–Schwartz implies  $U_\zeta x = f(x) U_\eta x$  for every unit vector  $x$  (in the initial support of the involved isometries) where  $f$  is a map from the unit sphere of the initial support to  $\mathbf{U}(1)$ . But  $f$  has to be constant by the antilinearity of our isometries.  $\square$

Given two sets  $A, B$  we denote by  $\text{Bij}(A, B)$  the set of bijections from  $A$  to  $B$  and similarly to before we define a set of triples

$$\mathcal{M}(\mathbf{H}) := \left\{ (\varphi_1, F, \varphi_2) : \varphi_1, \varphi_2 \in \mathcal{S}_n(\mathbf{B}(\mathbf{H})), F \in \text{Bij}(\Lambda_{\varphi_1}^\perp, \Lambda_{\varphi_2}^\perp) \right\}. \quad (6.4)$$

then we have a map

$$\begin{array}{l} \Phi_\otimes : \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) \longrightarrow \mathcal{M}(\mathbf{H}) \\ \omega_\zeta \longmapsto \left( \varphi_1, (\widetilde{U}_\zeta)_\# , \varphi_2 \right) \end{array} \quad \left| \begin{array}{l} \text{marginals } \varphi_1, \varphi_2 \\ \varphi_1 = \varphi_{\rho_1}, \varphi_2 = \varphi_{\rho_2} \\ \rho_1 = \text{Tr}_2 |\zeta\rangle\langle \zeta| \\ \Theta_\zeta = U_\zeta \rho_1^{1/2}. \end{array} \right.$$

Recall that  $J_b^1$  is the map on  $\mathcal{S}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H}))$  that takes the first marginal. In the following we are going to omit the identification  $\mathcal{C}(\mathbf{H}) \cong \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ . In particular the integration map  $\Psi$  will be considered as a map  $\Psi : \mathcal{D}_1^\perp(\mathbf{P}_c) \longrightarrow \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$ .

**Definition 6.2.** We say that a map  $F : \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) \longrightarrow \mathcal{D}_1^\perp(\mathbf{P}_c)$  is compatible with the first marginal if  $\Psi(F(\omega)) = \rho(J_b^1(\omega))$  for any  $\omega \in \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H}))$ . This means that the following diagram

$$\begin{array}{ccc} \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) & \xrightarrow{F} & \mathcal{D}_1^\perp(\mathbf{P}_c) \\ J_b^1 \downarrow & \swarrow \Psi & \\ \mathcal{S}_n(\mathbf{B}(\mathbf{H})), & & \end{array}$$

commutes.

Of course as a particular example we can take the map  $\mathcal{F}$  obtained by the composition

$$\mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) \xrightarrow{J_b^1} \mathcal{S}_n(\mathbf{B}(\mathbf{H})) \xrightarrow{\Phi} \mathcal{D}_1^\perp(\mathbf{P}_c). \quad (6.5)$$

**Theorem 6.3.** *The map  $\Phi_\otimes$  is well defined and injective. For any fixed  $\varphi \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  and a measure  $\mu \in \Lambda_\varphi^\perp$  representing  $\varphi$  we have a map*

$$\Phi_\otimes^\mu : \left\{ \omega \in \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) : J_b^1 \omega = \varphi \right\} \longrightarrow \mathcal{D}(\mathbf{P}_c \times \mathbf{P}_c), \quad \Phi_\otimes^\mu(\omega_\zeta) = (\text{Id} \times \widetilde{U}_\zeta)_\# (\text{tr}(\cdot)\mu).$$

This map is valued in the set of admissible transport plans

$$\Phi_\otimes^\mu(\omega_\zeta) \in \Pi_d(\text{tr}(\cdot)\mu, (\widetilde{U}_\zeta)_\# (\text{tr}(\cdot)\mu)).$$

In a similar way, a map  $F$  which is compatible with the first marginal can be combined with  $\Phi_\otimes$  to the map

$$\Phi_\otimes^F : \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) \longrightarrow \mathcal{D}_1^\perp(\mathbf{P}_c \times \mathbf{P}_c), \quad \Phi_\otimes^F(\omega) := (\text{Id} \times \widetilde{U}_\zeta)_\# (F(\omega)).$$

We have a compatibility property expressed by the commutative diagram

$$\begin{array}{ccc} \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) & \xrightarrow{\Phi_{\otimes}^F} & \mathcal{D}_1^{\perp}(\mathbf{P}_c \times \mathbf{P}_c) \\ J_b^1 \downarrow & \searrow F & \downarrow \pi_{\#}^1 \\ \mathcal{S}_n(\mathbf{B}(\mathbf{H})) & \xleftarrow{\Psi} & \mathcal{D}_1^{\perp}(\mathbf{P}_c). \end{array}$$

When  $F = \mathcal{F}$  as before (eq. (6.5)) this becomes a compatibility with  $\Phi$  as the diagram

$$\begin{array}{ccc} \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H})) & \xrightarrow{\Phi_{\otimes}^F} & \mathcal{D}_1^{\perp}(\mathbf{P}_c \times \mathbf{P}_c) \\ J_b^1 \downarrow & & \downarrow \pi_{\#}^1 \\ \mathcal{S}_n(\mathbf{B}(\mathbf{H})) & \xrightarrow{\Phi} & \mathcal{D}_1^{\perp}(\mathbf{P}_c) \end{array}$$

commutes.

*Proof.* Among all the spectral measures associated to  $\varphi_1$  there are those with all the projections of rank-one. Then starting from the assumption  $\Phi_{\otimes}(\omega_{\zeta}) = \Phi_{\otimes}(\omega_{\eta})$  and testing the equality  $\widetilde{U}_{\zeta} = \widetilde{U}_{\eta}$  for an arbitrary choice of one of these rank-one presentations of spectral measures we get the existence of a orthonormal set of vectors  $(e_i)_i$  spanning the initial domain of  $U_{\zeta}$  and  $U_{\eta}$  where  $|U_{\zeta}e_i\rangle\langle U_{\zeta}e_i| = |U_{\eta}e_i\rangle\langle U_{\eta}e_i|$  for every  $i$ . As in the proof of Proposition 6.1  $U_{\zeta} = \lambda U_{\eta}$  for a phase  $\lambda$ . The marginals now coincide and this means  $\Theta_{\zeta} = \lambda \Theta_{\eta}$  which implies that the corresponding states are equal. The rest of the proof is straightforward. In particular notice we get admissible transport plans because at any instance the discrete measures are in the form  $\sum_i \lambda_i P_{V_i}$  for an orthogonal family of finite rank projections and the transport maps are induced by antilinear partial isometries  $U_{\zeta}$  with  $P_{V_i} \leq U_{\zeta}^* U_{\zeta}$ . This means that for every  $i$  the points  $P_{V_i}$  and  $\widetilde{U}_{\zeta}(P_{V_i})$  belong to the same connected component.  $\square$

**Remark 6.4.** Of course the role of the marginals is symmetric. The flip automorphism  $\mathbf{H} \otimes \mathbf{H} \rightarrow \mathbf{H} \otimes \mathbf{H}$  that on simple tensors is defined by  $x \otimes y := y \otimes x$  induces an homeomorphism of the space of the states that switches the marginals. One checks immediately  $\Phi_{\otimes}(\omega_{\zeta}) = (\varphi_2, (\widetilde{U}_{\zeta}^{-1})_{\#}, \varphi_1)$ .

**Remark 6.5** (Wasserstein Cost of pure states). After Theorem 6.3, we can define a Wasserstein cost, depending on  $p$ , for any pure normal state of  $\mathbf{B}(\mathbf{H} \otimes \mathbf{H})$ . In particular given  $\varphi_1, \varphi_2 \in \mathcal{S}_n(\mathbf{B}(\mathbf{H}))$  and  $\omega_{\zeta} \in \mathcal{PS}_n(\mathbf{B}(\mathbf{H} \otimes \mathbf{H}))$ , for each  $\mu \in \Lambda_{\varphi_1}^{\perp}$  we have the transport plan  $\Phi_{\otimes}^{\mu}(\omega_{\zeta})$  (induced by the map  $\widetilde{U}_{\zeta}$ ) between admissible representations of  $\varphi_1$  and  $\varphi_2$  whose  $p$ -cost will be

$$\int_{\mathbf{P}_c \times \mathbf{P}_c} d^p(P, Q) \Phi_{\otimes}^{\mu}(\omega_{\zeta})(dP dQ) = \int_{\mathbf{P}_c} d^p(P, \widetilde{U}_{\zeta}(P)) \text{tr}(P) \mu(dP).$$

Hence, the cost of  $\omega_{\zeta}$  will be given by taking the lowest possible cost among all the  $\Phi_{\otimes}^{\mu}(\omega_{\zeta})$ :

$$\mathcal{C}_p(\omega_{\zeta})^p := \inf_{\mu \in \Lambda_{\varphi_1}^{\perp}} \int_{\mathbf{P}_c} d^p(P, \widetilde{U}_{\zeta}(P)) \text{tr}(P) \mu(dP). \quad (6.6)$$

Following the Proposition 4.9, it is equivalent to restrict the minimisation only among those  $\mu$  concentrated inside  $\mathbf{P}_1$ . Moreover the inf is actually attained: there exists  $\mu \in \mathbf{P}_1$ , a priori not unique and depending on  $p \geq 1$ , such that

$$\mathcal{C}_p(\omega_{\zeta})^p = \int_{\mathbf{P}_c} d^p(P, \widetilde{U}_{\zeta}(P)) \mu(dP).$$

Notice however that by construction, it is immediate to see that

$$\mathcal{C}_p(J_b^1 \omega_{\zeta}, J_b^2 \omega_{\zeta}) \leq \mathcal{C}_p(\omega_{\zeta}).$$

## A Homogeneous spaces and principal bundles

### Homogeneous spaces and principal bundles

Let  $G$  be a group acting (say on the right) on a space  $M$ . We usually denote this action with  $x \cdot g$ . Sometimes also the symbol  $\mathcal{R}_g(x) = x \cdot g$  will be used. The action is *free* whenever  $x \cdot g = x$  for some  $x \in M$  implies  $g = e$ . Assume that  $G$  acts on two spaces  $M$  and  $N$ . A map  $\varphi : M \rightarrow N$  is equivariant if

$$\varphi(x \cdot g) = \varphi(x) \cdot g, \quad \forall x \in M, \text{ and } g \in G.$$

**Definition A.1.** Let  $G$  be a Lie group. An homogeneous space is a manifold  $M$  with a transitive left action of  $G$ .

Given a closed subgroup  $B \subset G$  we can prove that the space of the left cosets  $G/B$  is a manifold. The left action of  $G$  on itself commutes with the right  $B$ -action so that it descends to a left transitive action on  $G/B$ . Thus  $G/B$  is a basic example of an homogeneous space. On the other hand, let  $M$  be a homogeneous space and fix a point  $p \in M$ . The stabiliser  $I_p := \{g \in G : g \cdot p = p\}$  is a closed subgroup. It is easy to prove that  $M$  is equivariantly diffeomorphic to  $G/I_p$ . Therefore every homogeneous space is in the form  $G/B$  with  $B \subset G$  closed.

**Definition A.2.** (cfr [44]). Let  $M$  be a manifold and  $G$  a Lie group. A principal bundle over  $M$  with structure group  $G$  consists in a manifold  $E$  with a right action of  $G$  such that:

1. The action is free and  $M$  is the quotient space  $E/G$  with smooth projection  $\pi : E \rightarrow M$ .
2. The following local triviality of  $E$  is satisfied: any point  $x \in M$  has a neighborhood  $U$  such that  $\pi^{-1}(U)$  is isomorphic to  $U \times G$ . Here *isomorphic* means that we can find a diffeomorphism  $\psi : \pi^{-1}(U) \rightarrow U \times G$  in the form  $\psi(u) = (\pi(u), \varphi(u))$  for a smooth map  $\varphi : \pi^{-1}(U) \rightarrow G$  satisfying  $\varphi(u \cdot g) = \varphi(u)g$  for every  $g \in G$ .

To synthetise this definition we say that  $E \rightarrow M$  is a principal bundle.

**Example 3.** Every homogeneous space  $G/B$  is the base of a principal bundle. Indeed we can prove that  $G \rightarrow G/B$  is a principal bundle with structure group  $B$ . In particular the local triviality follows from the existence of local smooth sections of the projection. If we consider the left translation action of  $G$  on itself we also see that the projection is equivariant.

Let  $E \xrightarrow{\pi} M$  be a  $G$ -principal bundle. At every point  $p \in E$ , the vertical space  $\mathcal{V}_p := N(d\pi : T_p E \rightarrow T_\pi M) \subset T_p E$  is the tangent space of the fiber. Using the  $G$ -action it can be canonically identified with the Lie algebra  $\mathfrak{g}$  of  $G$  in the following way: every  $X \in \mathfrak{g}$  defines the *fundamental* vector field  $\tilde{X} \in \Gamma(T E)$  (sections of the tangent bundle) with  $\tilde{X}_p := \left. \frac{d}{dt} \right|_{t=0} p \cdot (\exp tX)$ . Fundamental vector fields are of course vertical and at every point the map  $\mathfrak{g} \rightarrow \mathcal{V}_p$  given by  $X \mapsto \tilde{X}_p$  is an isomorphism. However in general there is no preferred choice of *horizontal* subspaces of  $T E$ . This extra structure amounts to a *connection*.

**Definition A.3.** A connection on the principal bundle  $E \rightarrow M$  is a smooth distribution  $p \mapsto \mathcal{H}_p \subset T_p E$  of vector subspaces called *horizontal* with the properties:

1. For every  $p \in E$  we have  $T_p E = \mathcal{V}_p \oplus \mathcal{H}_p$ .
2. Invariance: for every  $g \in G$  and  $p \in E$  then  $d\mathcal{R}_g \mathcal{H}_p = \mathcal{H}_{p \cdot g}$ .

A connection on  $E$  provides us with a notion of horizontal curves and horizontal liftings of curves. Moreover given any representation  $G \rightarrow \text{End}(V)$  on a vector space, the classical *associated bundle construction* produces a vector bundle  $W \rightarrow M$  having  $V$  as typical fiber and the connection on  $E$  induces a covariant derivative (in the usual meaning) on  $W$ . In particular this gives a covariant derivative in the tangent bundle  $T M$  of the base.

## B Some basic facts in operator theory

### B.1 Antilinearity

Recall that our Hilbert spaces have inner products complex linear in the first entry. We follow [14] (there the inner product is linear in the first entry). An antilinear operator  $T : \mathbf{H} \rightarrow \mathbf{K}$  is an additive operator such that  $T\lambda\xi = \bar{\lambda}T\xi$ .

Let  $J : \mathbf{H} \rightarrow \mathbf{H}$  be antilinear and isometric:  $\|J\xi\| = \|\xi\|$  for every  $\xi$ . By polarization it follows  $\langle J\xi, J\eta \rangle = \langle \eta, \xi \rangle$  for every couple of vectors. If such  $J$  is invertible it is called *antiunitary*.

An antilinear and isometric  $J : \mathbf{H} \rightarrow \mathbf{H}$  is called an *involution* if  $J^2 = \text{Id}_{\mathbf{H}}$ . It follows that  $J$  is an antiunitary. Involutions always exist for every Hilbert space and are very useful: if  $T : \mathbf{H} \rightarrow \mathbf{K}$  is antilinear then  $JT$  is linear and we can safely talk about bounded antilinear operators by looking at  $JT$  (for just one  $J$ ; it does not depends on the choice).

Let  $T : \mathbf{H} \rightarrow \mathbf{K}$  be antilinear bounded, then the adjoint of  $T$  is the unique antilinear bounded operator  $T^* : \mathbf{K} \rightarrow \mathbf{H}$  such that

$$\langle T^*\xi, \eta \rangle = \langle T\eta, \xi \rangle, \quad \xi \in \mathbf{K}, \eta \in \mathbf{H}.$$

It satisfies:  $(\lambda T)^* = \lambda T^*$  in contrast with the behavior of the adjoint for linear operators. Using an involution on  $\mathbf{H}$  we can compute  $T^* := J(TJ)^*$  in terms of the adjoint of a linear operator.

## B.2 Polar decompositions

A bounded operator  $T : \mathbf{H} \rightarrow \mathbf{K}$  is a partial isometry if  $T^*T$  is a projection  $P$ . Therefore  $PH = N(T)^\perp$  and also  $Q := TT^*$  is the projection onto  $R(T)$ , the range of  $T$ . These are called respectively initial and final support of  $T$ . It also follows that  $T$  restricts to an isometry  $N(T)^\perp \rightarrow R(T)$ .

**Theorem B.1.** (*Left polar decomposition*) Any  $T \in \mathbf{B}(\mathbf{H}, \mathbf{K})$  (two Hilbert spaces) has the decomposition  $T = UP$  for a non negative operator  $P : \mathbf{H} \rightarrow \mathbf{H}$  and a partial isometry  $U : \mathbf{H} \rightarrow \mathbf{K}$ . This decomposition is unique if we require that  $N(U) = N(P)$ . Equivalently if we require that the initial support  $(U^*U)\mathbf{H}$  of  $U$  is  $N(P)^\perp$ . In this case we have the properties:  $P = |T| := \sqrt{T^*T}$  and the decomposition reads

$$T = U|T|,$$

with

$$U^*U = \text{Proj } \overline{N(T)^\perp} \quad \text{and} \quad UU^* = \text{Proj } \overline{R(T)}.$$

*Proof.* Let  $P := |T| = \sqrt{T^*T}$  then  $N(P) = N(T)$  and  $N(T)^\perp = \overline{R(T^*)} = \overline{R(|T|)}$ . It follows that on  $R(|T|)$  is well defined an isometric map  $U$  such that  $U(|T|x) = Tx$ . On the orthogonal, which is  $N(T)$  we declare it zero. Then  $U$  is defined everywhere (and remains isometric on the closure of  $R(|T|)$ ). Notice  $R(U) = \overline{R(T)}$ .

Assume we have decomposition  $T = UP$  with  $N(U) = N(P)$ . Then  $T^* = PU^*$  and  $T^*T = PU^*UP$  but  $U^*U = \text{Proj } R(P)$  i.e.  $T^*T = P^2$  which means  $P = |T|$ . We already know that  $U$  is uniquely determined on the range  $P = |T|$  and we are done.  $\square$

The left polar decomposition of  $T^*$  gives rise to the *right polar decomposition*

$$T = |T^*|U$$

of  $T$ . Begin with  $T = U|T|$ ; then  $T^* = |T|U^*$  and  $TT^* = U|T|^2U^*$ . This can be iterated; we get  $(TT^*)^n = U|T|^{2n}U^*$  for every power. It follows by the unicity of the functional calculus that  $|T^*| = U|T|U^*$  i.e.  $|T^*|U = U|T|$  (because  $|T|U^*U = |T|$ ).

Let us now consider an antilinear bounded operator  $T : \mathbf{H} \rightarrow \mathbf{K}$ . Using an involution as before we can construct polar decompositions

$$T = V|T| = |T^*|V,$$

for  $|T| = \sqrt{T^*T}$  which is a linear operator while  $V$  is an antilinear partial isometry with  $V^{**}V = \text{Proj}(N(T))^\perp$  and  $VV^* = \text{Proj } R(T)$ . In particular  $V$  reverts the order inside the inner product: on  $(N(T))^\perp$  we have  $\langle V\xi, V\eta \rangle = \langle \eta, \xi \rangle$ . We also have

$$|T^*| = V|T|V^*.$$

## References

- [1] J. AGREDO AND A. FAGNOLA, *On quantum versions of the classical Wasserstein distance*, Stochastics **89** 6–7 (2017), 910–922.
- [2] W.O. AMREIN AND K. SINHA, *On Pairs of Projections in a Hilbert Space*, Linear Algebra and its Applications 208-209 (1994), 425–435.
- [3] C.J. ATKIN, *The Finsler geometry of groups of isometries of Hilbert space*, J. Austral. Math. Soc. Ser. A **42** no. 2 (1987), 196–222.
- [4] S. ATTALL, *Tensor products and partial traces*, lecture notes from the *Lectures in quantum noise theory* book project: <http://math.univ-lyon1.fr/~attal/chapters.html>
- [5] E. ANDRUCHOW, *The Grassmann manifold of a Hilbert space*, In, Proceedings of the XI-Ith "Dr. Antonio A. R. Monteiro" Congress, Actas Congr. "Dr. Antonio A. R. Monteiro", Univ. Nac. del Sur, Bahía Blanca, 2014, 41–55.
- [6] E. ANDRUCHOW, *Operators which are the difference of two projections*, J. Math. Anal. Appl. **420** (2014), 1634–1653.
- [7] E. ANDRUCHOW, *Pairs of Projections: Geodesics, Fredholm and Compact Pairs*, Complex Anal. Oper. Theory **8** (2014), 1435–1453.
- [8] E. ANDRUCHOW AND A. VARELA, *Riemannian geometry of finite rank positive operators*, Differential Geometry and its Applications **23** Issue 3, 305–326.
- [9] E. ALFSEN, AND F. SHULTZ, *Unique decompositions, faces, and automorphisms of separable states*, J. Math. Phys. **51** (2010).

- [10] J. AVRON, R. SEILER, AND B. SIMON, *The index of a pair of projections*, J. Funct. Anal. **120**(1) (1994), 220–237.
- [11] J.-D. BENAMOU AND Y. BRENIER, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), 375–393.
- [12] P. BIANE AND D. VOICULESCU, *A free probability analogue of the Wasserstein metric on the trace-space space*, GAFA **11** (2009), 353–454.
- [13] S. BIANCHINI AND L. CARAVENNA, *On the extremality, uniqueness and optimality of transference plans*, Bull. Acad. Sinica **4** (2009), 353–454.
- [14] B. BLACKADAR, “Operator Algebras, Theory of  $C^*$ -algebras and von Neumann Algebras” Encyclopedia of Mathematical Sciences, Springer-Verlag Berlin Heidelberg, 2006.
- [15] V.I. BOGACHEV, “Measure Theory” Vol. I, Springer, Berlin 2007.
- [16] E. CAGLIOTI, F. GOLSE AND T. PAUL, *Quantum optimal transport is cheaper*, Jour. Stat. Phys. **181** (2020), 149–162.
- [17] E. A. CARLEN AND J. MAAS, *An analog of the 2-Wasserstein metric in non-commutative probability under which the fermionic Fokker-Planck equation is gradient flow for the entropy*, Comm. Math. Phys. **331** (2014), 887–926.
- [18] E. A. CARLEN AND J. MAAS, *Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance*, J. Funct. Anal. **273** (2017), 1810–1869.
- [19] E. A. CARLEN AND J. MAAS, *Non-commutative calculus, Optimal Transport and functional inequalities in dissipative quantum systems*, J. Statist. Phys. **178** (2020), 319–378.
- [20] F.M. CIAGLIA, A. IBORT, J. JOST AND G. MARMO, *Manifolds of classical probability distributions and quantum density operators in infinite dimensions*, Info. Geo. **2** (2019), 231–271.
- [21] F. CIPRIANI AND J.-L. SAUVAGEOT, *Derivations as square roots of Dirichlet forms*, J. Funct. Anal. **201** (2003), 78–120.
- [22] E. CHIUMENTO *Hopf-Rinow theorem in Grassmann manifolds of  $C^*$ -algebras*, In ”Geometric Methods in Physics, XXVII Workshop Trends in Mathematics”, Springer Nature, Switzerland, 2019, 145–152.
- [23] K. CHRZASZCZ, J. JACHYMSKI AND F. TUROBOS, *On characterizations and topology of regular semimetric spaces*, Publ. Math. Debrecen **93** (2018), 87–105.
- [24] A. CONNES: *Compact metric spaces, Fredholm modules and hyperfiniteness*, Ergodic Theory and Dynamical Systems **9** (1989), 207–220.
- [25] A. CONNES: “Noncommutative geometry”, Academic Press Inc., San Diego CA, 1994.
- [26] G. CORACH, H. PORTA AND L. RECHT, *The Geometry of Spaces of Projections in  $C^*$ -algebras*, Adv. Math. **101** (1993), 59–77.
- [27] G. DE PALMA AND D. TREVISAN *quantum optimal transport with quantum channels*, Ann. Henry Poincaré, **22** (2021), 3199–3234.
- [28] G. DE PALMA, M. MILAD, D. TREVISAN AND L. SETH, *The quantum Wasserstein distance of order 1*, IEEE Trans. Inf. Theory **67** No. 10 (2021), 6627–6643.
- [29] F. D’ANDREA, P. MARTINETTI, *A dual formula for the spectral distance in noncommutative geometry*, J. Geom. Phys. **159** (2021), 59–64.
- [30] F. D’ANDREA, P. MARTINETTI *A view on optimal transport from Noncommutative Geometry*, SIGMA, **6** (2010), Contribution to the Special Issue Noncommutative Spaces and Fields.
- [31] J.L. DANNY, *The support of extremal measures with given marginals*, Michigan Math. J. **27** (1980), 59–64.
- [32] I. DIMITRIĆ, *A note on equivariant embeddings of Grassmanians*, Publications de l’Institut math. Beograd (N.S.) **59** (1996), 131–137.
- [33] J. DIXMIER, “von Neumann algebras”, with a preface by E. C. Lance, translated from the second French edition by F. Jellet. North-Holland Mathematical Library, 27. North-Holland Publishing Co., Amsterdam-New York, 1981
- [34] A. EDELMAN, T. ARIAS AND S. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), 303–353.
- [35] J. EISERT, C. SIMON AND M.B. PLENIO, *On the quantification of entanglement in infinite-dimensional quantum systems*, Journal of Physics A: Mathematical and General, **35** Number 17.

- [36] F. GOLSE, C. MOUHOT, AND T. PAUL, *On the mean-field and classical limits of quantum mechanics*, Comm. Math. Phys. **343** (2016), 165–205.
- [37] P. HARMS AND A. C. G. MENNUCCI, *Geodesics in infinite dimensional Stiefel and Grassmann manifolds*, C. R. Acad. Sci. Paris, Ser. I **350** (2012), 773–776.
- [38] F. HERBUT, *On bipartite pure-state entanglement structure in terms of disentanglement*, Journal of Mathematical Physics **47** (2006).
- [39] D. HORNSHAW, *Quantum optimal transport for approximately finite-dimensional  $C^*$ -algebras preprint*, arXiv:1910.03312.
- [40] J. M. ISIDRO AND M. MACKEY, *The manifold of finite rank projections in the algebra  $\mathcal{L}(H)$  of bounded linear operators*, Expositiones Mathematicae, **20** 2 (2002), 97–116.
- [41] V. JURDJEVIC, I. MARKINA AND F. SILVA LEITE, *Extremal Curves on Stiefel and Grassmann Manifolds*, The Journal of Geometric Analysis **30** (2020), 3948–3978.
- [42] T. KATO, “Perturbation Theory for Linear Operators”, volume 132 of Grundlehren der Mathematischen Wissenschaften, Springer–Verlag, Berlin, 1980.
- [43] R. V. KADISON AND J. R. RINGROSE, “Fundamentals of the theory of operator algebras”, volume II Advanced theory, corrected reprint of the 1986 original, Graduate Studies in Mathematics 16, American Mathematical Society, Providence, 1997.
- [44] S. KOBAYASHI AND T. NOMITZU, “Foundations of differential geometry” vol. I, Interscience Tracts in Pure and Applied Mathematics 15, New York-London 1963.
- [45] A. KRIEGL AND P. MICHOR, “The convenient setting of global analysis”, Mathematical Surveys and Monographs, 53, American Mathematical Society, Providence, 1997.
- [46] K. LANDSMAN, “Foundations of Quantum Theory”, Fundamental Theories of Physics. Cham: Springer International Publishing, 2017.
- [47] J. LAWSON, *Ordered probability spaces*, J. Math. Anal. Appl, **455**, No. 1 (2017), 167–179.
- [48] S. LANG, “Differential and Riemannian manifolds”, Graduate Texts in Mathematics, Springer–Verlag, New York, 1995.
- [49] M. RIEFFEL, *Metric on state spaces*, Documenta Math. **4** (1999), 559–600.
- [50] W. SŁOMCZYŃSKI AND K. ZYCZKOWSKI, *The Monge distance between quantum states*, J. Phys. A, Math. Gen., **45** vol. 31 (1998).
- [51] W. SŁOMCZYŃSKI AND K. ZYCZKOWSKI, *The Monge metric on the sphere and geometry of quantum states*, J. Phys. A, Math. Gen., **34** (2001).
- [52] N. LEBEDEVA AND A. PETRUNIN, *Curvature bounded below: a definition a la Berg-Nikolaev*, Elec. res. ann. in math. sc. **17** (2020), 122–124.
- [53] E. MARCZEWSKI AND P. SIKORSKI, *Measures in Nonseparable Metric Spaces*, Colloq. Math. **1** (1948), 133–139.
- [54] A. PERES, *Separability Criterion for Density Matrices*, Physical Review Letters, (1996).
- [55] H. PORTA AND L. RECHT, *Minimality of geodesics in Grassmann manifolds*, Proc. Am. Math. Soc., **100** (1987), 464–466.
- [56] D. ROBINSON, *Normal and locally normal states*, Comm. Math. Phys. (1970), 219–234.
- [57] C. VILLANI, “Optimal transport - old and new”, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer–Verlag, Berlin, 2009.
- [58] W. A. WILSON, *On Semi-Metric Spaces*, Amer. J. Math., **53** (1931), 361–373.
- [59] M. WIRTH, *A Noncommutative Transport Metric and Symmetric Quantum Markov Semigroups as Gradient Flows of the Entropy*, preprint, arXiv:1808.05419, (2018).
- [60] N.E. WEGGE-OLSEN, “ $K$ -theory and  $C^*$ -algebras. A friendly approach.”, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1993.