Geometry of Grassmannians and optimal transport of quantum states

Paolo Antonini* and Fabio Cavalletti†

Abstract

Let H be a separable Hilbert space. We prove that the Grassmannian $\mathsf{P}_c(\mathsf{H})$ of the finite dimensional subspaces of H is an Alexandrov space of nonnegative curvature and we employ its metric geometry to develop the theory of optimal transport for the normal states of the von Neumann algebra of linear and bounded operators $\mathsf{B}(\mathsf{H})$. Seeing density matrices as discrete probability measures on $\mathsf{P}_c(\mathsf{H})$ (via the spectral theorem) we define an optimal transport cost and the Wasserstein distance for normal states. In particular we obtain a cost which induces the w^*-topology.

Our construction is compatible with the quantum mechanics approach of composite systems as tensor products $\mathsf{H} \otimes \mathsf{H}$. We provide indeed an interpretation of the pure normal states of $\mathsf{B}(\mathsf{H} \otimes \mathsf{H})$ as families of transport maps. This also defines a Wasserstein cost for the pure normal states of $\mathsf{B}(\mathsf{H} \otimes \mathsf{H})$, reconciling with our proposal.

Contents

1 Introduction **2**
1.1 Notations . . . 4

2 Preliminaries **4**
2.1 Geometry of the space of projections . . . 4
2.1.1 Metric aspects . . . 6
2.2 Normal States . . . 11
2.2.1 Topology on the space of states . . . 12
2.2.2 Partial traces and marginals . . . 13

3 Spectral-projections measures **14**
3.1 Weak* topology and convergence of projections . . . 16

4 Wasserstein distance between normal states **18**
4.1 Wasserstein distance over $\mathcal{P}(\mathsf{P}_c)$ . . . 18
4.2 The Optimal Transport Cost in $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ . . . 19
4.3 Wasserstein distance in $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ . . . 22

5 Kantorovich duality for W_p and consequences **23**
5.1 Kantorovich duality . . . 23
5.2 Wasserstein geodesics . . . 25

6 Tensor product interpretation: a generalization **28**
6.1 Pure States as transport maps . . . 28

A Homogeneous spaces and principal bundles **30**

B Some basic facts in operator theory **31**
B.1 Antilinearity . . . 31
B.2 Polar decompositions . . . 32

*Dipartimento di Matematica e Fisica "E. de Giorgi", Università del Salento, Lecce (Italy) email: paolo.antonini@unisalento.it

†Mathematics Area, SISSA, Trieste (Italy), email: cavallet@sissa.it.

1 Introduction

In Connes' noncommutative geometry program [23] many naturally singular spaces of great interest in geometry or quantum physics can be fruitfully addressed using noncommutative operator algebras. There is nowadays a huge literature about these noncommutative spaces for which we refer to the aforementioned book [23]. We limit ourselves here to list some of the most known and interesting examples. Among these we find: leaf spaces of foliations, the space of unitary representations of a discrete group and the phase space of quantum mechanics. This last one is related with the current paper where we address the problem of optimal transport for quantum states.

Due to the pervasiveness of noncommutative spaces, extensions of the classical tools as measure theory, topology, differential calculus and Riemannian geometry, have been pursued in the noncommutative setting and during the last few years, as it is naturally expected, also the search of an appropriate analogue of a Wasserstein distance received a great deal of attention. Some important progresses have been obtained.

In the noncommutative setting states take over the role of probability measures; for example, in the case of the algebra of matrices (as well in $\mathsf{B}(\mathsf{H})$ if we consider only normal states) using the matrix trace, states can be identified with positive definite matrices with unit trace which are indeed called density matrices.

In the context of spectral triples considered as noncommutative manifolds, where the noncommutative algebra $\mathcal{A}$ interacts with a Dirac operator, Connes [22] 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 Kantorovitch 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 [45, 25] 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 [11]. Their metric extends the classical Wasserstein metric.

A proposal for the finite dimensional case, which follows the principle to adapt the dynamical formulation of optimal transport à la Benamou and Brenier [10] has been given by Carlen and Maas [15, 16, 17]. Here one assigns a length to each path of probability measures connecting the marginals.

A key property of the resulting quantum distance in loc. cit. is the fact that it is induced by a Riemannian metric on the manifold of quantum states and the quantum generalisation of the heat semigroup is the gradient flow of the von Neumann entropy $\mathrm{Ent}(\rho) = \mathrm{Trace}(\rho \log \rho)$. This replaces the classical relative entropy of the commutative case. Also the relation of this approach to the rate of convergence of the quantum Ornstein-Uhlenbeck semigroup [17] have been established.

Subsequent developments worth mentioning include: the one of Wirth [53], based on the noncommutative Dirichlet forms of Cipriani and Sauvageot [19] and the work of Hornshow [35] where also the approximately finite dimensional case is considered estabilishing lower bounds on Ricci curvature. We refer to these papers for more details. Finally another proposal by Golse, Mouhot and Paul [31] arose in the context of the study of the semiclassical limit of quantum mechanics and it relies on the concept of *couplings* with applications to the study of the mean-field limit of quantum mechanics.

Our contribution goes in a new direction to study a static formulation of the optimal transport problem between quantum states. We base our constructions on the geometric structure of the Grassmann manifold of all the finite rank projections of the underlying separable Hilbert space H.

Let us describe more precisely the setting.

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

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

We introduce a distance between density matrices relying on the optimal transport problem between probability measures over the Grassmanian of H. latter is denoted by P and is defined as the collection of all orthogonal projections of 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 P is induced by the Spectral Theorem: by compactness and self-adjointness the following correspondence is rather natural:

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

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 the projection onto the kernel does not belong to the support of the associate measure μ_φ. Since $\mathrm{tr}(\rho_\varphi) = 1$, it follows that $\mathrm{tr}(\cdot)\mu_\varphi$ is a probability measure over the Polish space P_c, the submanifold of P of finite rank orthogonal projections.

The Polish structure of P_c, making it amenable to standard measure theory techniques, is the one inherited as a Finsler submanifold of $\mathsf{B}(\mathsf{H})$. However P_c admits a more convenient geometric structure induced infinitesimally by viewing 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 P_c is an Alexandrov space of non-negative curvature. A fact giving a very natural setting to explore geometric links between normal states and optimal transport. This will be thoroughly studied in Section 2.1.

Denoting by d the geodesic distance of P_c, the Wasserstein distance between the normal states φ, ψ can be then defined as the Wasserstein distance of the spectral measures μ_φ and μ_ψ after being normalized to be probability measures:

$$W_p(\varphi, \psi) := W_p^{\mathsf{P}_c}\big(\mathrm{tr}(\cdot)\mu_\varphi, \mathrm{tr}(\cdot)\mu_\psi\big). \tag{1.1}$$

In this formula $W_p^{\mathsf{P}_c}$ denotes the classical Wasserstein distance defined over the Polish space $(\mathsf{P}_c, \mathsf{d})$. 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 φ we consider the set $\Lambda_\varphi^\perp$ of discrete measures $\mu = \sum \lambda_i \delta_{P_i}$ 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 φ and ψ as

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

as the Wasserstein distance between the two (compact) sets of associated measures representing the states. The main properties we obtain for $\mathcal{C}_p$ are the following ones:

Existence of optimal configurations: for any couple of normal states φ and ψ, 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 μ_0, μ_1 can always be taken with support contained inside the connected component P_1, i.e. the space of projections with one dimensional rank (Proposition 2.6). This is $\mathbb{P}(\mathsf{H})$ the projective space of 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 $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ (Theorem 4.11).

We also obtain the 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 $(\mathsf{P}_c, \mathsf{d})$, we also study W_p-geodesics of $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ in Section 5.2.

In the last section we study tensor product Hilbert spaces $\mathsf{H} \otimes \mathsf{H}$ corresponding in quantum mechanics to composite systems. A natural way to match two normal states φ, ψ of $\mathsf{B}(\mathsf{H})$ would be via a normal state $\Xi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H}))$ satisfying the partial trace conditions $J_\flat^1 \Xi = \varphi$ and $J_\flat^2 \Xi = \psi$ (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 ω_ζ of $\mathcal{PS}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H}))$, i.e. any pure normal state of $\mathsf{B}(\mathsf{H} \otimes \mathsf{H})$ with partial traces φ and ψ, we associate a family of admissible transport plans between admissible representations of φ and ψ. In particular this permits to assign a well-defined optimal transport cost to any pure normal state of $\mathsf{B}(\mathsf{H} \otimes \mathsf{H})$ (Remark 6.5).

We conclude by mentioning that we tried to keep the paper as self-contained as possible. In particular in Section 2 we have collected, and in some cases re-proved, many of the known geometric properties of the Grassmanian P_c that are used in this paper and that were distributed through different references.

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 ie. $P = P^*$ and $P^2 = P$ when $P^2 = P$ we say idempotent. Also for two projections we write $Q \leq P$ if and only if $Q\mathsf{H} \subset P\mathsf{H}$. This is equivalent to $PQ = Q$ or $QP = Q$.

2 Preliminaries

2.1 Geometry of the space of projections

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

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

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

We describe its geometry mainly following [4, 5, 6, 24, 49]. Fundamental is the natural action of the unitary group $\mathsf{U}(\mathsf{H})$ by conjugation $g \cdot P = gPg^*$ for $g \in \mathsf{U}(\mathsf{H})$.

We recall here few but important facts about the group $\mathsf{U}(\mathsf{H})$. This is a Banach–Lie group, closed inside $\mathsf{B}(\mathsf{H})$ with Lie algebra identified with the skew adjoint operators $\mathfrak{u} := T_1\mathsf{U}(\mathsf{H}) = \mathsf{B}_{sa}(\mathsf{H})$ having the operators commutator as Lie bracket. The exponential map $\exp : \mathfrak{u} \longrightarrow \mathsf{U}(\mathsf{H})$ is the operators exponentiation. It is surjective because in $\mathsf{B}(\mathsf{H})$ we may form Borel functions of normal operators; this gives a logarithm for every skew-adjoint operator.

All the curves in the form

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

with $X \in \mathsf{B}_h(\mathsf{H})$ i.e. the translations of one parameter groups are called the *group geodesics* of $\mathsf{U}(\mathsf{H})$. The name is legitimated by the fact that we can find a natural class of linear connections on $\mathsf{U}(\mathsf{H})$ creating such geodesics. Moreover one can show that are minimal curves inside $\mathsf{U}(\mathsf{H})$ with respect to the natural Finsler structure inherited by the embedding $\mathsf{U}(\mathsf{H}) \subset \mathsf{B}(\mathsf{H})$ (see [2]). We are now ready to discuss the geometry of P.

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

$$T_P\mathsf{P} = \big\{Y \in \mathsf{B}_h(\mathsf{H})\colon PY + YP = Y\big\}; \tag{2.1}$$

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 $\mathsf{B}_h(\mathsf{H})$

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

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 $\mathscr{C}_P$.

In symbols

$$\mathscr{C}_P = \big\{Y \in \mathsf{B}_h(\mathsf{H})\colon PY + YP = Y\big\}. \tag{2.3}$$

Let us prove the (2.1). The first inclusion comes 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 $X \in \mathscr{C}_P$ satisfies $X = [[X, P], P]$. This also means (every commutator with P is codiagonal) that $\mathscr{C}_P = \{i[X, P] : X \in \mathsf{B}_s(\mathsf{H})\}$. 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 γ are exactly the geodesics through P with respect to a family of natural connections. Summing up:

$$T_P\mathsf{P} = \mathscr{C}_P = \big\{Y \in \mathsf{B}_h(\mathsf{H})\colon PY + YP = Y\big\} = \big\{i[X, P] : X \in \mathsf{B}_h(\mathsf{H})\big\}.$$

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

$$\mathsf{B}_h(\mathsf{H}) = \mathscr{C}_P \oplus \mathscr{D}_P. \tag{2.4}$$

2. **Homogeneous space structure of the connected components.**
The $\mathsf{U}(\mathsf{H})$ action on 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 [54, Corollary 5.2.9]). In other words the $\mathsf{U}(\mathsf{H})$-orbits, i.e. the conjugacy classes are the connected components in P :

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

These connected components are easily found; let $R(Q)$ denote the range of the operator Q and $N(Q)$ its kernel. Then 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 $P \in \mathsf{P}$ (for the rest of this section). The stabiliser $\mathsf{I}_P = \{g : g \cdot p = p\}$ coincides with the subgroup $\{g \in \mathsf{U}(\mathsf{H}) : [g, P] = 0\}$ and the quotient $\mathsf{U}(\mathsf{H})/\mathsf{I}_P$ is diffeomorphic to $\mathcal{O}_P$. More precisely, using the canonical projection

$$\mathsf{U}(\mathsf{H}) \longrightarrow \mathsf{U}(\mathsf{H})/\mathsf{I}_P \cong \mathcal{O}_P, \tag{2.5}$$

we get a principal bundle with equivariant projection. In other words $\mathcal{O}_P$ is an homogeneous space [4, 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*). The canonical connection induces in the customary way a notion of parallel translation, covariant derivative and geodesics for P. We don't construct them explicitly here because we will consider in a while, a second, more direct connection on $T\mathsf{P}$ sharing the same geodesics.

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

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

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

$$\frac{DX}{dt} = E_{\gamma(t)}\Big(\frac{d}{dt}X(\gamma(t))\Big). \tag{2.7}$$

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

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

All the geodesics starting at $P \in \mathsf{P}$ are in the form $\gamma(t) = e^{itZ}Pe^{-itZ}$ with $Z \in T_P(\mathsf{P})$ [5, 24]. As anticipated we can prove that these are also all the geodesics with respect to the connection induced by the natural connection in P as an homogeneous reductive space.
To check that the geodesic equation is satisfied for $\gamma(t) = e^{itZ}Pe^{-itZ}$ we take the opportunity to discuss the manifold of symmetries $\mathsf{S} := \big\{S \in \mathsf{B}_h(\mathsf{H}) : S^2 = 1\big\}$, diffeomorphic to P via the map

$$\mathcal{F} : \mathsf{P} \longrightarrow \mathsf{S}, \quad P \longmapsto 2P - 1. \tag{2.8}$$

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

$$T_S\mathsf{S} = \Big\{X \in \mathsf{B}_h(\mathsf{H}) \colon SX + XS = 0\Big\}.$$

We have a corresponding projection on the tangent space which has the form

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

also inducing a connection on S. This is given by the same formula as (2.6). On the other hand the map $\mathcal{F} : \mathsf{P} \longrightarrow \mathsf{S}$ 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 $\mathsf{B}_h(\mathsf{H})$ and its differential $d_P\mathcal{F}(X) = 2X$ intertwines the two projections onto P and S.

Now thanks to the inclusion $\mathsf{S} \subset \mathsf{U}(\mathsf{H})$ some formulas simplify when passing to S. Start with the curve $\gamma(t) = e^{itZ} P e^{-itZ}$ in P with $Z \in T_P\mathsf{P}$. 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 γ 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 S and by the properties of $\mathcal{F}$ we see that γ is a geodesic too. Moreover $\mathcal{F}(\gamma)$ is also a geodesic in $\mathsf{U}(\mathsf{H})$ (a traslation of a one parameter group). In other words S is totally geodesic inside $\mathsf{U}(\mathsf{H})$).
Put $Y := -iSZ/2 \in T_S\mathsf{S}$ then the geodesic in S can also be written as $t \mapsto e^{tXS/2} S e^{-tXS/2}$. Indeed the exponential map is the restriction of the family of analytic mappings

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

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

2.1.1 Metric aspects

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

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

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

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

generalising the familiar riemannian (Kähler) structure on the finite dimensional Grassmann manifold. We summarise some of the basic properties (see [7, 40]) :

- the topology on P_c induced by the embedding $\mathsf{P}_c \subset B_h(\mathsf{H})$ where $B_h(\mathsf{H})$ is given with the norm topology coincides the topology induced by the embedding $\mathsf{P}_c \subset \mathsf{HS}(\mathsf{H})$. 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 [27]. We have orthogonal projections on the tangent space and on the normal space to P_c and the theory of submanifolds presents no differences with the finite dimensional case. In fact the orthogonal projection is exactly the projection on the codiagonals that we have already used.
  Now let $P \in \mathsf{P}_c$ and X, Y vector fields tangent to P_c; if we denote with D_XY the covariant derivative in the flat space $\mathsf{HS}(\mathsf{H})$, we have at P:

  $$D_XY = \big(PD_XY(1-P) + (1-P)D_XYP\big) + (XY + YX)(1 - 2P).$$

  The first addendum is tangential to P_c while the second one is normal. Therefore

  $$\nabla_XY = PD_XY(1-P) + (1-P)D_XYP, \quad \text{the connection of } \mathsf{P}_c \text{ at } P,$$

  $$\sigma(X, Y) = (XY + YX)(1 - 2P) \quad \text{the second fundamental form at } P.$$
- The geodesics that we have already discussed are geodesics for the metric in P_c too. In particular $t \longmapsto e^{t[X,P]} P e^{-t[X,P]}$ is the unique geodesic starting from P with initial velocity X.
- The curvature tensor is

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

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

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

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

[(a)] since the operators are codiagonal the trace of XY is real valued

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

To describe in more details the geometry of P_c is useful to follow the techniques in [33] presented in the real case. The extension to our, complex case is straightforward as we will show in the following.

To start with, we present P_c as the base of a second principal bundle with fiber U_r. This is in contrast with the previous discussion. Firstly we introduce a notation for the connected components of P_c

$$\mathsf{P}_r := \big\{P \in \mathsf{P}_c : \dim R(P) = r\big\}. \tag{2.9}$$

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

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

with tangent space

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

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

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

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

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

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

at φ is obtained combining the decomposition

$$\mathsf{H} = R(\varphi) \oplus R(\varphi)^\perp \tag{2.11}$$

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

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

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

There are two commuting left and right action

$$\mathsf{U}(\mathsf{H}) \;\circlearrowright\; \mathsf{St}(r,\mathsf{H}) \;\circlearrowleft\; \mathsf{U}_r = \mathsf{U}(\mathbb{C}^r)$$

corresponding to post and pre composition

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

The $\mathsf{U}(\mathsf{H})$ action is transitive while the U_r one is free. Two points φ and ψ 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

$$\pi^{\mathsf{St}} : \mathsf{St}(r,\mathsf{H}) \longrightarrow \mathsf{St}(r,\mathsf{H})/\mathsf{U}_r \cong \mathsf{P}_r, \quad \varphi \longmapsto \varphi\mathsf{U}_r \longmapsto \varphi\varphi^*. \tag{2.12}$$

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

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

Therefore X is horizontal if and only if $\Re\,\mathrm{tr}(X^*\varphi Y) = 0$ for every $Y \in \mathsf{B}_{sa}(\mathbb{C}^r)$. 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 P_r. For horizontal vectors $X, Y \in T_\varphi \mathsf{St}(r, \mathsf{H})$ we have

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

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

Following [29] we derive the geodesic equation

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

Starting with the the condition $\gamma^*\gamma = \mathrm{Id}_r$ and differentiating two times we get $\ddot\gamma^*\gamma + 2\dot\gamma^*\dot\gamma + \gamma^*\ddot\gamma = 0$. If γ is a geodesic, the normal component of the second derivative is zero i.e. $\ddot\gamma = -\gamma S$ for some curve $S(t) = S(t)^* \in \mathsf{B}(r, \mathsf{H})$. Inserting this condition in the previous equation we get (2.13). On the other hand if a curve $t \mapsto \mathsf{St}(r, \mathsf{H})$ satisfies (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 $\varphi_0 \in \mathsf{St}(r, \mathsf{H})$. We continue to use the splitting (2.11) induced by φ_0 so that operators in H are 2×2 block-matrices. For any skew-adjoint operator

$$\mathcal{M} = \begin{pmatrix} A & B \\ -B^* & 0 \end{pmatrix} \quad \text{with skew-adjoint} \quad A : R(\varphi_0) \to R(\varphi_0),$$

put $\mathcal{Q} := \begin{pmatrix} A/2 & 0 \\ 0 & 0 \end{pmatrix}$. Then $\mathcal{Q}^* = -\mathcal{Q}$ and we have a curve

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

Proposition 2.2. *The curve γ is the geodesic in $\mathsf{St}(r, \mathsf{H})$ 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 $X \in T_{\varphi_0}\mathsf{St}(r, \mathsf{H})$ 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*

$$A = 2(\varphi_0\varphi_0^*)X\varphi_0^*\big|_{R(\varphi_0)} \quad \textit{and} \quad B = \varphi_0 X^*(\varphi_0\varphi_0^* - \mathrm{Id})\big|_{R(\varphi_0)^\perp}.$$

Proof. The proof that γ is a geodesic is the computation in [41, Section 3.4.1] that we write for definiteness. Since we already know that $\gamma(t) \in \mathsf{St}(r, \mathsf{H})$ at every time 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

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

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

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

It follows that γ 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 intended restricted to $R(\varphi_0)$ or $R(\varphi_0)^\perp$. ∎

Corollary 2.3. *For a geodesic $\gamma : [0,1] \longrightarrow \mathsf{St}(r,\mathsf{H})$, the image of the map $\gamma(t) : \mathbb{C}^r \longrightarrow \mathsf{H}$ (for every t) is contained in the subspace of 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* $\mathrm{span}(\gamma(0), \gamma(1))$ *for every $t \in [0,1]$. The geodesic* moves inside a finite dimensional subspace of 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:

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

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

An embedding $\iota : \mathsf{K} \hookrightarrow \mathsf{H}$ of Hilbert spaces induces embeddings $\iota_* : \mathsf{St}(r,\mathsf{K}) \hookrightarrow \mathsf{St}(r,\mathsf{H})$ and $\iota_* : \mathsf{P}_r(\mathsf{K}) \hookrightarrow \mathsf{P}_r(\mathsf{H})$ 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 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. [33]. *Let K be a Hilbert space; for every embedding $\iota : \mathsf{K} \longrightarrow \mathsf{H}$ the corresponding $\iota_* : \mathsf{St}(r,\mathsf{K}) \hookrightarrow \mathsf{St}(r,\mathsf{H})$ is an isometric embedding with totally geodesic image. Moreover:*

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

Proof. As already mentioned, the proof in [33] 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 $y \in \mathsf{St}(r,\mathsf{H})$ 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 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 $\iota : \mathsf{K} \hookrightarrow \mathsf{H}$ let $\mathsf{U}(\iota(\mathsf{K})^\perp)$ be the unitary group of the complement. it is included (diagonally) in $\mathsf{U}(\mathsf{H})$ and acts by isometries on $\mathsf{St}(r,\mathsf{H})$ with fixed points being exactly $\iota_*(\mathsf{St}(r,\mathsf{K}))$. Therefore $\iota_*(\mathsf{St}(r,\mathsf{K}))$ is totally geodesic because is the fixed point set of a set of isometries. 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 $\mathsf{d}_\mathsf{K}(x,y) \geq \mathsf{d}_\mathsf{H}(\iota_*(x), \iota_*(y))$ because $\iota_*(\mathsf{St}(r,\mathsf{K}))$ is totally geodesic. For the reversed inclusion, let $\gamma \subset \mathsf{St}(r,\mathsf{H})$ 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 K. In other words $\gamma = \iota_* \circ \widetilde\gamma$ for a geodesic $\widetilde\gamma \subset \mathsf{St}(r,\mathsf{K})$. Using $\widetilde\gamma$ the inequality $\mathsf{d}_\mathsf{H}(\iota_*(x), \iota_*(y)) \geq \mathsf{d}_\mathsf{K}(x,y)$ immediately follows. Point 2. is 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 $\mathsf{d}_{\mathsf{K}}(x,y) \geq \mathsf{d}_{\mathsf{H}}(\iota_*(x),\iota_*(y))$. To prove the converse, one takes any smooth path ζ connecting $\iota_*(x)$ and $\iota_*(y)$. We can divide ζ in subpaths $\zeta\big|_{[t_i,t_{i+1}]}$ $(i=1,...,n)$ such that each one is 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 manifacture a finite dimensional vector space $\widetilde{\mathsf{K}}$ which contains every image of the map $\eta(t)$ for every t. Of course we can enlarge it to ensure $\mathsf{K} \subset \widetilde{\mathsf{K}}$. Now we apply the finite dimensional case (in $\widetilde{\mathsf{K}}$) to estimate

$$\mathsf{d}_{\mathsf{K}}(x,y) = \mathsf{d}_{\widetilde{\mathsf{K}}}(\iota_*(x),\iota_*(y)) \leq \ell(\eta) \leq \ell(\xi)$$

and we are done.

Proof of step 3. We check just point 1. and 2. in the finite dimensional case because the infinite dimensional etension is similar to the one performed for the Stiefel case. First point: we have $\mathsf{d}_{\mathsf{P}_r(\mathsf{K})}(x,y) \geq \mathsf{d}_{\mathsf{P}_r(\mathsf{H})}(\iota_*(x),\iota_*(y))$ 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 γ be a minimal geodesic in $\mathsf{P}_r(\mathsf{H})$ joining $\iota_*(x)$ and $\iota_*(y)$. Lift this to a curve $\zeta(t)$ in the Stiefel manifold $\mathsf{St}(r,\mathsf{H})$. The images of the maps $\zeta(0)$ and $\zeta(1)$ are exactly x and y. This means that ζ belongs to the image of $\mathsf{St}(r,\mathsf{K})$ and in turn that γ belongs to the image of the embedding $\iota_* : \mathsf{P}_r(\mathsf{K}) \hookrightarrow \mathsf{P}_r(\mathsf{H})$. It follows $\mathsf{d}_{\mathsf{P}_r(\mathsf{K})}(x,y) \leq \mathsf{d}_{\mathsf{P}_r(\mathsf{H})}(\iota_*(x),\iota_*(y))$. As before this fact implies the point 2. □

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

Proof. According to [46] 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

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

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. □

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

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

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

In particular taking γ minimal $\mathsf{d}(Q,Q_1) \leq \mathsf{d}(P,P_1)$.

Proof. Let $Z \in T_P\mathsf{P}$; 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:

$$Q = \begin{pmatrix} QP & 0 \\ 0 & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}. \tag{2.14}$$

Now $Q(t) := e^{itZ}Qe^{-itZ}$ is a geodesic from Q to Q_1 with $\dot Q(0) = i[Q,Z]$ and $\|\dot Q(0)\|^2_{T\mathsf{P}_c} = \mathrm{tr}(\dot Q(0)^2)$. 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 $\|\dot Q(0)\|^2_{T_Q\mathsf{P}_c} = 2\,\mathrm{tr}(QX^*XQ)$. In the same way $\|\dot\gamma(0)\|^2_{T\mathsf{P}_c} = 2\,\mathrm{tr}(X^*X)$. The result is clear from the positivity of X^*X. □

2.2 Normal States

Let $\mathcal{A}$ be a C^*-algebra. A linear functional $\varphi : \mathcal{A} \longrightarrow \mathbb{C}$ is positive if $\varphi(a^*a) \geq 0$ for every $a \in \mathcal{A}$. Then φ is automatically bounded; if $\|\varphi\| = 1$ it is called a state. When the algebra is unital this normalisation is equivalent to the condition $\varphi(1) = 1$. Denoted with $\mathcal{S}(\mathcal{A})$, the space of the states of $\mathcal{A}$ included in the dual $\mathcal{A}^*$ and considered with the topology induced by the w^*-one. 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 $\varphi : \mathcal{A} \longrightarrow \mathbb{C}$ 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. □

We will denote by $\mathcal{PS}(\mathcal{A})$ the set of *pure states* that is the extreme 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 $\mathbb{K} = \mathbb{K}(\mathsf{H})$, the C^*-algebra of compact operators. We have an identification

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

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

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

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

$$\mathcal{C}(\mathsf{H}) := \{\rho \in \mathcal{L}^1 \colon \rho \geq 0,\ \operatorname{tr}(\rho) = 1\}, \tag{2.16}$$

with the identification denoted by

$$\mathcal{C}(\mathsf{H}) \ni \rho \longmapsto \varphi_\rho \in \mathcal{S}(\mathbb{K}). \tag{2.17}$$

Viceversa we may, sometimes, use the notation ρ_φ or $\rho(\varphi)$ for the density matrix of φ.

Example 1. Every unit vector $\xi \in \mathsf{H}$ defines a state $\omega_\xi : \mathsf{B}(\mathsf{H}) \longrightarrow \mathbb{R}$ by $\omega_\xi(B) = \langle \xi, B\xi \rangle$. The density matrix of ω_ξ is the rank one operator $\rho : \mathsf{H} \to \mathsf{H}$ with $\rho(\eta) = \langle \xi, \eta \rangle \xi$. This follows from: $\operatorname{tr}(\rho B) = \operatorname{tr}(B\rho) = \langle \xi, B\xi \rangle$. 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 $\mathsf{B}(\mathsf{H})$; on the other hand there are many states on $\mathsf{B}(\mathsf{H})$ which are not in this form. Precisely a state $\varphi \in \mathcal{S}(\mathsf{B}(\mathsf{H}))$ comes from a density matrix if and only if it satisfies one of the following equivalent properties (see [42, Theorem 4.12], [37, Theorem 7.1.8] and [28, Theorem 1, Part I, Chapter 4]):

1. it is *normal*: $\varphi(T) = \sup_{\mathcal{F}} \varphi(F)$ for every directed family $\mathcal{F} \subset \mathsf{B}(\mathsf{H})^+$ 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
$$\varphi(\sum_j p_j) = \sum_j \varphi(p_j).$$
The sum $\sum_j p_j$ is defined as the projection on the closure of the smallest subspace in H containing all the $p_j\mathsf{H}$. This is is exactly the operation of forming $\sup_j p_j$ in the partially ordered set of all the projections in $\mathsf{B}(\mathsf{H})$ with the order given by the inclusion $e \leq f$ iff $e\mathsf{H} \subset f\mathsf{H}$ (see [42]). 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

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

in the sense of norm convergence. The vectors ξ_n can be taken pairwise orthogonal [37, Theorem 7.1.9].

By the spectral theorem we see that that a pure state φ of $\mathbb{K}$ is necessarily a vector state i.e. in the form $\varphi = \omega_\xi$ for a unit vector ξ. 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 $\mathcal{PS}(\mathbb{K}) \cong \mathbb{P}(\mathsf{H})$. On the right we have the projective space of H, the quotient of the unit sphere by the $\mathsf{U}(1)$-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 generated by the vector). Then the density matrix of φ is:*

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

Proof. This fact is more general. A proof can be found in [37, Theorem 7.1.9] (see also the following remark therein). In our case the proof is simpler. The series $\sum_{n=1}^{\infty} \|\xi_n\|^2 P_n$ converges in the operator norm to a non negative operator T. Of course T is compact and can be diagonalised $T = \sum_{n=1}^{\infty} \lambda_n^2 \, |\eta_n\rangle\langle\eta_n|$ with the complete orthonormal system $\{\eta_n\}_n$. Now for every $A \in \mathsf{B}(\mathsf{H})$ we compute

$$\varphi(A) = \sum_{n=1}^{\infty} \langle A\xi_n, \xi_n\rangle = \sum_{n=1}^{\infty} \langle A \sum_{m=1}^{\infty} \langle \xi_n, \eta_m\rangle \eta_m, \xi_n\rangle = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \langle A\eta_m, \xi_n\rangle\langle \xi_n, \eta_m\rangle.$$

We can interchange the sums because the series converges absolutely and using the identity $P_n = \dfrac{|\xi_n\rangle\langle\xi_n|}{\|\xi_n\|^2}$ we get

$$\varphi(A) = \sum_{m=1}^{\infty} \langle A\eta_m, \sum_{n=1}^{\infty} \overline{\langle \xi_n, \eta_m\rangle}\xi_n\rangle = \sum_{m=1}^{\infty} \langle A\eta_m, (\sum_{n=1}^{\infty} \|\xi_n\|^2 P_n)\eta_m\rangle = \sum_{m=1}^{\infty} \langle A\eta_m, T\eta_m\rangle.$$

This is $\varphi(A) = \operatorname{tr}(TA)$. □

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

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

where the symbol $\simeq$ denotes an isomorphism between the two convex sets. This isomorphism maps extremals to extremals: any pure state ω on $\mathbb{K}$ has a unique extension to a normal state ω' on $\mathsf{B}(\mathsf{H})$ given by the same density operator which is extremal for $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$. We refer to [42] for more details. Based on this we make the following

Definition 2.9. We denote with $\mathcal{PS}_n(\mathsf{B}(\mathsf{H}))$ the set of the *pure normal states of* $\mathsf{B}(\mathsf{H})$. *These are precisely the extremals of $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ identifiable with $\mathbb{P}(\mathsf{H})$ the projective space of* 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 two 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 $\mathsf{B}(\mathsf{H})$ 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 $\varphi_{\rho_n}(B) = \operatorname{tr}(\rho_n B) \longrightarrow \varphi_\rho(B) = \operatorname{tr}(\rho B)$ 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 $B \in \mathsf{B}(\mathsf{H})$. This defines $\sigma(\mathcal{S}(\mathbb{K}), \mathsf{B}(\mathsf{H}))$ called the weak topology in virtue of the identification $\mathsf{B}(\mathsf{H}) = (\mathcal{L}^1)'$.

Importantly Robinson proved that all the above topologies coincide [50, 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*

$$\rho_n \xrightarrow{\mathcal{L}^1} \rho \qquad \textit{iff} \qquad \varphi_{\rho_n} \xrightarrow{w^*} \varphi_\rho\,.$$

2.2.2 Partial traces and marginals

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

$$J_\flat^{\mathsf{H}} : \mathcal{S}(\mathsf{B}(\mathsf{H} \otimes \mathsf{K})) \longrightarrow \mathcal{S}(\mathsf{B}(\mathsf{H})),$$

and the corresponding map $J_\flat^{\mathsf{K}}$. For definiteness we give the formula of the first one by dualising the inclusion $J^{\mathsf{H}} : \mathsf{B}(\mathsf{H}) \longrightarrow \mathsf{B}(\mathsf{H}) \otimes \mathsf{B}(\mathsf{K})$, $T \longmapsto T \otimes \mathrm{Id}_{\mathsf{K}}$:

$$J_\flat^{\mathsf{H}} \varphi(T) = \varphi(T \otimes \mathrm{Id}_{\mathsf{K}}).$$

Let us describe now partial traces. We follow closely the lecture notes [3] where all the proofs can be found.
Assume that H and K are separable. Every vector $\xi \in \mathsf{K}$ defines linear bounded operators

$$R_\xi : \mathsf{H} \longrightarrow \mathsf{H} \otimes \mathsf{K}, \qquad R_\xi^* : \mathsf{H} \otimes \mathsf{K} \longrightarrow \mathsf{H},$$

uniquely specified on simple tensors by

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

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

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

By definition: $\langle \zeta, {}_\xi T_\xi\, \eta \rangle = \langle \zeta \otimes \xi, T\eta \otimes \xi \rangle$, for every $\zeta, \eta \in \mathsf{H}$ and one proves

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

Theorem 2.11. *Let $T \in \mathcal{L}^1(\mathsf{H} \otimes \mathsf{K})$ be a trace class operator; there is a unique trace class operator $\mathrm{Tr}_{\mathsf{K}}(T) \in \mathcal{L}^1(\mathsf{H})$ such that*

$$\mathrm{tr}(\mathrm{Tr}_{\mathsf{K}}(T)\, B) = \mathrm{tr}(T(B \otimes \mathrm{Id}_{\mathsf{K}})) \tag{2.18}$$

for every $B \in \mathsf{B}(\mathsf{H})$. Concretely $\mathrm{Tr}_{\mathsf{K}}(T)$, that we call the partial trace with respect to K, can be constructed taking any ortonormal basis $(\xi_n)_n$ of K:

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

We have the following properties

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

Exchanging the role of H and K we define in the same way the partial trace Tr_{H}. If $\mathsf{K} = \mathsf{H}$, the unique case we shall treat we denote with Tr_1 and Tr_2 the two partial traces. For instance for $\xi \otimes \eta \in \mathsf{H} \otimes \mathsf{H}$:

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

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

$$\rho(J_\flat^{\mathsf{H}} \varphi) = \mathrm{Tr}_{\mathsf{K}}(\rho_\varphi).$$

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

Given normal states $\varphi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ and $\psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{K}))$ the tensor product $\varphi \otimes \psi$ is a normal state on $\mathsf{B}(\mathsf{H} \otimes \mathsf{K})$. We say that $\varphi \otimes \psi$ is *separable*. More generally we agree with [30] on the following.

[(b)] in [3] is denoted by ${}_{\mathsf{K}}\langle \xi | T | \xi \rangle_{\mathsf{K}}$

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

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

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 $\mathsf{B}(\mathsf{H}\otimes\mathsf{K})$ is *entangled* if it is not separable.

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

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

3 Spectral-projections measures

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

$$\rho=\sum_i \lambda_i P_{V_i}, \qquad V_i\subset\mathsf{H}, \qquad \mathrm{tr}(\rho)=\sum_i \lambda_i \dim(V_i)=1, \tag{3.1}$$

where $\lambda_i>0$ are the eigenvalues of ρ and $P_{V_i}\in\mathsf{P}_c$ 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:

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

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

$$\mathcal{D}(\mathsf{P}_c):=\Big\{\mu=\sum_i\lambda_i\delta_{P_i}:\ P_i\in\mathsf{P}_c,\ \lambda_i\geq 0\Big\}, \tag{3.2}$$

with $\mathcal{D}(\mathsf{P}_c)$ mnemonic for "discrete" measures. Then we consider the following subsets

$$\mathcal{D}_1(\mathsf{P}_c):=\Big\{\mu\in\mathcal{D}(\mathsf{P}_c)\colon \mathrm{tr}(\cdot)\,\mu\in\mathcal{P}(\mathsf{P}_c)\Big\}, \tag{3.3}$$

playing the role of probability measures and

$$\mathcal{D}_1^{\perp}(\mathsf{P}_c):=\Big\{\mu\in\mathcal{D}_1(\mathsf{P}_c)\colon \mathrm{tr}(PQ)=0,\ \text{ for all } P\neq Q\in\mathrm{supp}(\mu)\Big\}, \tag{3.4}$$

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

We are then ready to define the following injection:

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

For consistency, we will also denote $\Phi(\rho)$ by μ_ρ. Notice that $\mathrm{tr}(\cdot)\mu_\rho(\mathsf{P}_c)=1$ follows from $\mathrm{tr}(\rho)=1$. The spectral Theorem implies that $\Phi(\rho_\varphi)\subset\mathcal{D}_1^{\perp}(\mathsf{P}_c)$ for every ρ_φ (pairwise orthogonal projections). Moreover as no repetition of eigenvalues is present in $\Phi(\varphi)$:

$$\Phi(\varphi)(P)\neq\Phi(\varphi)(Q), \qquad \text{for all } P\neq Q\in\mathrm{supp}\big(\Phi(\varphi)\big).$$

This property actually characterizes the image $\Phi(\mathcal{C}(\mathsf{H}))$.

Definition 3.1. Using the isomorphism between $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ and $\mathcal{C}(\mathsf{H})$, the map Φ given by

$$\Phi:\mathcal{S}_n(\mathsf{B}(\mathsf{H}))\longrightarrow\mathcal{D}_1(\mathsf{P}_c), \qquad \Phi(\varphi):=\sum_i\lambda_i\delta_{P_{V_i}}, \tag{3.6}$$

is well defined (with a slight abuse of notation). The notation μ_φ in place of $\Phi(\varphi)$ will sometimes be preferred.

Remark 3.2. The support of μ_φ is $\{P_{V_i} : i \in \mathbb{N}\} \subset \mathsf{P}_c$, a totally disconnected set; notice indeed that by orthogonality of the eigenspaces, $\|P_{V_i} - P_{V_j}\|_{\mathsf{B(H)}} = 1$ 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 $\mathrm{supp}(\mu_\varphi)$.

We define now the converse correspondence.

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

$$\Psi : \mathcal{D}_1(\mathsf{P}_c) \longrightarrow \mathcal{C}(\mathsf{H}), \quad \Psi(\mu) = \Psi\left(\sum_i \lambda_i \delta_{P_i}\right) := \sum_i \lambda_i P_i. \tag{3.7}$$

Notice indeed $\rho = \sum_i \lambda_i P_i$ converges in the trace norm to a well defined symmetric operator having $\mathrm{tr}(\rho) = \int \mathrm{tr}(P)\, \mu(dP) = 1$; hence $\Psi(\mathcal{D}_1(\mathsf{P}_c)) \subset \mathcal{C}(\mathsf{H})$.

By the spectral Theorem again one notices that

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

Hence Ψ is the left-inverse of Φ while, in general, it fails to satisfy $\Phi(\Psi(\mu)) = \mu$. Particularly relevant for us will be the sets

$$\Lambda_\varphi^\perp := \Psi^{-1}(\varphi) \cap \mathcal{D}_1^\perp(\mathsf{P}_c) = \Big\{\mu \in \mathcal{D}_1^\perp(\mathsf{P}_c) \colon \Psi(\mu) = \rho_\varphi\Big\}, \tag{3.8}$$

the set of the measures concentrated on pairwise orthogonal projections whose corresponding symmetric operator is the density matrix of φ. In particular any element of $\Lambda_\varphi^\perp$ represents a spectral decomposition of ρ_φ admitting repeated eigenvalues.

Coming to the topological properties of these sets, we recall that a sequence of probability measures $\mu_n \in \mathcal{P}(\mathsf{P}_c)$ is said to weakly converge to $\mu \in \mathcal{P}(\mathsf{P}_c)$ if by definition

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

It is well-known that, for Polish spaces, the Lévy-Prokhorov metric gives a metrization of weak convergence; in particular it makes $\mathcal{P}(\mathsf{P}_c)$ complete and separable. It will be therefore enough to describe topologically the subsets of $\mathcal{P}(\mathsf{P}_c)$ only using weakly converging sequences.

Moreover we recall the following classical fact about compact subsets of probability measures: if $(\mathcal{X}, \mathsf{d})$ is a metric space (considered with its Borel σ–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 $\Psi : \mathcal{D}_1(\mathsf{P}_c) \longrightarrow \mathcal{S}_n(\mathsf{B(H)})$ is continuous in the following sense: if $\mathrm{tr}(\cdot)\mu_n \rightharpoonup \mathrm{tr}(\cdot)\mu$, then $\Psi(\mu_n) \rightharpoonup \Psi(\mu)$.*

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

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

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

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

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

Then we analyse 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*

$$\mathrm{tr}(\cdot)\mathcal{D}_1^\perp(\mathsf{P}_c) := \Big\{\mathrm{tr}(\cdot)\mu \colon \mu \in \mathcal{D}_1^\perp(\mathsf{P}_c)\Big\}$$

is closed. Moreover for any $\varphi \in \mathcal{S}_n(\mathsf{B(H)})$, $\mathrm{tr}(\cdot)\Lambda_\varphi^\perp$ is compact.

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

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

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

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 $(Q_1, ..., Q_N) \in \mathsf{P}_c^N$ such that the Q_j that are different from zero are mutually orthogonal and satisfy $\sum_{j=1}^r P_{Q_j} = P_{V_i}$. If we call $\mathcal{Q}_i$ the set of all such decompositions we have N projections $q_j^{(i)} : \mathcal{Q}_i \longrightarrow \mathsf{P}_c$. 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 can embed $\mathcal{Q}_i$ into the union of all products $\mathbb{G}\big(V_i, \mathrm{tr}(Q_1)\big) \times \cdots \times \mathbb{G}\big(V_i, \mathrm{tr}(Q_N)\big)$ with the union running over the finite set of all the possible ways of writing $N = s_1 + \cdots + s_N$ with $s_j \in \mathbb{N}$ (including zero). We adopt the convention that $\mathbb{G}\big(V_i, 0\big) = \bullet$, the space with a point. Now 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 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 eigenvalues of $\Phi(\varphi)$ but here now they may be repeated. It holds true,

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

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

3.1 Weak* topology and convergence of projections

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

Lemma 3.6. *Let $(P_n)_n$ be a sequence of projections inside $\mathcal{L}^1$; assume moreover $P_n \rightharpoonup P$ in the w-topology to some $P \in \mathcal{L}^1$, i.e. in duality with $\mathsf{B}(\mathsf{H})$. 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 $P_n \rightharpoonup P$ in the weak topology we deduce that

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

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

$$\mathrm{tr}(P_n K P_n) = \mathrm{tr}(P_n^2 K) = \mathrm{tr}(P_n K) \longrightarrow \mathrm{tr}(PK);$$

on the other hand, since $KP_n \to KP$ in $\mathcal{L}^1$, it follows that $\mathrm{tr}(P_n K P_n) \to \mathrm{tr}(PKP)$. Therefore $P^2 = P$. □

Lemma 3.7. *Let $\varphi_n, \varphi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ such that $\varphi_n \rightharpoonup \varphi$. Consider the corresponding density matrices $\rho_n = \rho_{\varphi_n}$, $\rho = \rho_\varphi$ for which we consider any spectral decompositions (in the sense of (3.8))*

$$\rho_n = \sum_i \lambda_i^n P_i^n, \qquad \rho = \sum_i \lambda_i P_i,$$

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

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

$$\widetilde{P}_i \leq P_j, \qquad \widetilde{\lambda}_i = \lambda_j. \tag{3.9}$$

Proof. We start noticing the following: for any $B \in \mathsf{B}(\mathsf{H})$ it holds true $\mathrm{tr}\big(\rho_{n_k} P_i^{n_k} B\big) \longrightarrow \mathrm{tr}\big(\rho \widetilde{P}_i B\big)$. Indeed

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

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

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

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

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

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

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

Proof. **Step 1.** Consider the sequences $(\lambda_i^n)_{i\in\mathbb{N}}$ and $(\lambda_i)_{i\in\mathbb{N}}$ of eigenvalues of ρ_n and ρ, respectively, arranged in decreasing order and repeated according to the multiplicity; in particular both sequences have norm 1 in ℓ^1. Then [50, Theorem 2] proves that

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

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

$$\sum_{i\geq M} \lambda_i^n \leq \varepsilon, \quad \forall\, n \geq n_\varepsilon, \qquad \sum_{i\geq M} \lambda_i \leq \varepsilon. \tag{3.10}$$

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 (3.10); then the ℓ^1-convergence will imply (3.10) 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

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

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 pose $m = \infty$.

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

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

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 $i_j < m$

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

as $k \to \infty$. We define then $\mu := \sum_j \lambda_j \delta_{P_j}$. 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 ε was arbitrarily chosen and did not play any role in the construction of μ, it follows that $\sum_j \lambda_j \mathrm{tr}(P_j) = 1$, giving, $\mu \in \Lambda_\varphi^\perp$. As byproduct we have also shown that the sequence $\big(\lambda_{i_j^{n_k}} \mathrm{tr}(P_j^{n_k})\big)$ converges to $\big(\lambda_j \mathrm{tr}(P_j)\big)$ in $\ell^1(\mathbb{N})$.

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

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

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

We summarise the results in the next statement whose proof will be an easy consequence of previous convergence results.

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

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

Proof. The first point is Lemma 3.4 while the second part of the claim is precisely Proposition 3.8. □

4 Wasserstein distance between normal states

We will use the metric structure of $(\mathsf{P}_c, \mathsf{d})$ reviewed in Section 2.1, together with the map Ψ (Definition 3.3) to define a *static* Wasserstein distance between normal states of $\mathsf{B}(\mathsf{H})$. The classical definition of p-Wasserstein distance over $(\mathsf{P}_c, \mathsf{d})$ (being d an extended metric does not hurt the definition), the plan is to push it to normal states via Ψ.
With this motivation in mind, we begin describing in details $W_p^{\mathsf{P}_c}$.

4.1 Wasserstein distance over $\mathcal{P}(\mathsf{P}_c)$

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 d is finite solely when restricted on each connected component of P_c, we will consider a more stringent notion of transport plan (recall that $P, Q \in \mathsf{P}_c$ belong to the same connected component if and only if $\dim R(P) = \dim R(Q)$).

Definition 4.1. Given two probability measures $\mu_0, \mu_1 \in \mathcal{P}(\mathsf{P}_c)$, the set of d-transport plans between μ_0 and μ_1 will be given by

$$\Pi_{\mathsf{d}}(\mu_0, \mu_1) := \Big\{\nu \in \Pi(\mu_0, \mu_1)\colon\ \dim(R(P)) = \dim(R(Q)),\ \nu - a.e.\Big\}, \tag{4.1}$$

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

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

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

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

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

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 $\mu_0, \mu_1 \in \mathcal{P}(\mathsf{P}_c)$, there exists an optimal plan $\nu \in \Pi_{\mathsf{d}}(\mu_0, \mu_1)$ such that*

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

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

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

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

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

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

proving the claim. □

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

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

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

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

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

is open giving that $0=\liminf\nu_{n_k}(C)\geq\nu(C)\geq 0$, proving the claim. □

We conclude this short overview on optimal transport in $\mathcal{P}(\mathsf{P}_c)$ by recalling the simple relation between Wasserstein topology and weak topology. Here we refer to [51, Theorem 6.9]: if $\mu_n,\mu\in\mathcal{P}(\mathsf{P}_k)$ for some k independent of n, then

$$\mu_n\rightharpoonup\mu\iff W_p^{\mathsf{P}_c}(\mu_n,\mu)\to 0,$$

for any $p\geq 1$. Recall indeed that $(\mathsf{P}_k,\mathsf{d})$ is a complete and separable metric spaces with $\mathsf{d}\leq\pi/2$.

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

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

The use of multiple representations for the density matrices, i.e. $\Lambda_\varphi^\perp$ and $\Lambda_\psi^\perp$, together with the many connected components $(\mathsf{P}_c,\mathsf{d})$, motivate the following defintion.

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

$$\mathcal{C}_p(\varphi,\psi):=\inf_{\substack{\mu_0\in\Lambda_\varphi^\perp\\ \mu_1\in\Lambda_\psi^\perp}}W_p^{\mathsf{P}_c}\Big(\mathrm{tr}(\cdot)\,\mu_0,\mathrm{tr}(\cdot)\,\mu_1\Big),\tag{4.3}$$

where $\Lambda_\varphi^\perp,\Lambda_\psi^\perp\subset\mathcal{D}_1^\perp(\mathsf{P}_c)$ have been defined in (3.8).

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 $\varphi,\psi\in\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ and $p\geq 1$

$$\mathcal{C}_p(\varphi,\psi)=W_p^{\mathsf{P}_c}(\mathrm{tr}(\cdot)\Lambda_\varphi^\perp,\mathrm{tr}(\cdot)\Lambda_\psi^\perp),\tag{4.4}$$

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

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

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

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

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 $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$,*

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

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

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

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

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

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

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

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

We also consider $Z \in T_{P_{0,i}}\mathsf{P}$ 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

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

and by Proposition 2.6 $\mathsf{d}(P_{e_k}, P_{1,e_k}) \le \mathsf{d}(P_{0,i}, P_{1,j})$. 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

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

It is clear from the construction that the marginal measures of ν are still admissible measures for the states φ and ψ. Repeating the argument at most countably many times proves the claim. □

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

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

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

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

Proposition 4.9. *Given any $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$, there exist μ_0, μ_1 and ν, elements of $\Lambda_\varphi^{\perp,1}, \Lambda_\psi^{\perp,1}$ and of $\Pi_{\mathsf{d}}\big(\mu_0, \mu_1\big)$ respectively, such that*

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

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

$$\lim_{n\to\infty} W_p^{\mathsf{P}_c}\big(\mu_{0,n}, \mu_{1,n}\big) = \mathcal{C}_p(\varphi,\psi).$$

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

Now take $\nu_n \in \Pi_{\mathsf{d}}(\mu_{0,n}, \mu_{1,n})$ any optimal transport plan (Theorem 4.3). Since its marginal are converging, by tightness, ν_n is weakly converging, up to subsequences, as well to a certain $\nu \in \Pi(\mu_0, \mu_1)$. Since d is continuous and bounded on P_1:

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

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

Hence we have shown that the optimal transport problem defining the transport cost has always a solution. Moreover, by the symmetry of 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 $W_p^{\mathsf{P}_c}\big(\mu_0,\mu_1\big)=0$. Hence $\mu_0=\mu_1$ and by definition of $\Lambda_\varphi^\perp$ and $\Lambda_\psi^\perp$

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

We have therefore that $\mathcal{C}_p:\mathcal{S}_n(\mathsf{B}(\mathsf{H}))\times\mathcal{S}_n(\mathsf{B}(\mathsf{H}))\longrightarrow[0,\infty)$ is a semi-distance. First references for semi-distances date back to the first half of 20th century, see for instance [52]. We refer however to the recent [21] for a general overview on the topic.

Remark 4.10. Concerning triangular inequality for the cost $\mathcal{C}_p$, using Remark 4.6, one can deduce the following property: given $\varphi,\psi,\phi\in\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$

$$\mathcal{C}_p(\varphi,\psi)\le\inf_{\mu_0,\,\mu_1,\mu_2}\Big\{W_p^{\mathsf{P}_c}\big(\mu_0,\mu_1\big)+W_p^{\mathsf{P}_c}\big(\mu_1,\mu_2\big)\Big\},$$

infimum 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 a classical argument (gluing) in optimal transport that will be also used in the proof of the following Lemma 4.14. Moreover, whenever the intermediate normal state, say ϕ, 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:

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

The proof of Lemma 4.14 will clarify this point.

We now investigate the topology induced by $\mathcal{C}_p$, starting by its converging sequences. Notice indeed that semi-distances induce a topology whose open sets are in the form $U\subset\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ for which for every $\varphi\in U$ there exists $r>0$ so that $B_r(\varphi):=\big\{\psi\in\mathcal{S}_n(\mathsf{B}(\mathsf{H}))\colon\mathcal{C}_p(\varphi,\psi)<r\big\}\subset U$.

Theorem 4.11. *Let φ_n,φ be normal states of $\mathsf{B}(\mathsf{H})$. Then*

$$\mathcal{C}_p(\varphi_n,\varphi)\longrightarrow 0\qquad\Longleftrightarrow\qquad\varphi_n\xrightarrow{w^*}\varphi\,.$$

Proof. Suppose first that $\mathcal{C}_p(\varphi_n,\varphi)\to 0$ as $n\to\infty$. By Proposition 4.9 there exist $\mu_{0,n}\in\Lambda_{\varphi_n}^{\perp,1}$, $\mu_{1,n}\in\Lambda_\varphi^{\perp,1}$ and $\nu_n\in\Pi_{\mathsf{d}}(\mu_{0,n},\mu_1)$ such that

$$\mathcal{C}_p(\varphi_n,\varphi)=W_p^{\mathsf{P}_c}(\mu_{0,n},\mu_{1,n})=\left(\int_{\mathsf{P}_c\times\mathsf{P}_c}\mathsf{d}(P,Q)^p\,\nu_n(dPdQ)\right)^{1/p}\to 0$$

By 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, $\Psi(\mu_{0,n})\to\Psi(\mu_1)$ in w^*-convergence. By definition $\Psi(\mu_{0,n})=\varphi_n$ and $\Psi(\mu_1)=\varphi$ giving the first claim.

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

$$\mathcal{C}_p\big(\varphi_n,\varphi\big)=W_p^{\mathsf{P}_c}(\mu_{0,n},\mu_{1,n}\big).$$

We now invoke Proposition 3.8: from $\varphi_n\rightharpoonup\varphi$ we deduce the existence of a subsequence μ_{0,n_k} and $\mu_1\in\Lambda_\varphi^{\perp,1}$ such that $\mu_{0,n_k}\rightharpoonup\mu_1$. Then $\mu_{0,n_k}\to\mu_1$ also in Wasserstein distance over P_1. Hence

$$\mathcal{C}_p(\varphi_n,\varphi)=W_p^{\mathsf{P}_c}(\mu_{0,n},\mu_{1,n})\le W_p^{\mathsf{P}_c}(\mu_{0,n_k},\mu_1)\to 0,$$

giving the claim. □

Theorem 4.11 together with [21, Theorem 4.2] imply following

Corollary 4.12. *The topology $\tau_{\mathcal{C}_p}$ over the set of normal states coincide with the w^*-topology.*

Proof. [21, Theorem 4.2] states that $\tau_{\mathcal{C}_p}$ coincide with the topology induced by sequences in the usual sense: U is closed if contains limit points (w.r.t. to $\mathcal{C}_p$) of all converging sequences all contained inside U. Then Theorem 4.11 and metrizability of w^*-topology over bounded set proves the claim. □

4.3 Wasserstein distance in $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$

Even though the cost functional $\mathcal{C}_p$ is fully satisfactory (see Theorem 4.11), for completeness we address the issue of the lack of 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 4.13. For any $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ and $p \geq 1$ define their p-Wasserstein distance by

$$W_p(\varphi, \psi) := W_p^{\mathsf{P}_c}(\mathrm{tr}(\cdot)\Phi(\varphi), \mathrm{tr}(\cdot)\Phi(\psi)), \tag{4.6}$$

with the map Φ defined in (3.6). 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

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

defines an extended distance over $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$. As before, the symmetry of 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 4.14. *(Triangular inequality for W_p). Let φ, ψ and ϕ be three elements of $\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$. Then*

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

Proof. Consider $\nu_1 \in \Pi_{\mathsf{d}}\big(\mathrm{tr}(\cdot)\,\mu_\phi, \mathrm{tr}(\cdot)\,\mu_\psi\big)$ and $\nu_2 \in \Pi_{\mathsf{d}}\big(\mathrm{tr}(\cdot)\,\mu_\psi, \mathrm{tr}(\cdot)\,\mu_\varphi\big)$ optimal plan whose existence is assured by Theorem 4.3.

If $\mathrm{tr}(\cdot)\,\mu_\phi = \sum_i \alpha_i \delta_{P_{V_i}}$, $\mathrm{tr}(\cdot)\,\mu_\psi = \sum_i \beta_i \delta_{P_{W_i}}$ and $\mathrm{tr}(\cdot)\,\mu_\varphi = \sum_i \gamma_i \delta_{P_{Z_i}}$ then the transport plans ν_1 and ν_2 can be written as

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

with $\alpha^1_{i,j}, \alpha^2_{i,j} \geq 0$ and $\sum_{i,j} \alpha^1_{i,j} = \sum_{i,j} \alpha^2_{i,j} = 1$; moreover marginal constraint are given in the following form

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

Following the classical gluing procedure of transport plans, we define

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

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

We also need to check that $(\pi_{13})_\sharp \Theta$ is admmissible: for $(\pi_{13})_\sharp \Theta$-a.e. P, Q it holds $i(P, Q) = 0$ (or $\dim(R(P)) = \dim(R(Q))$). Moreover from [9] if (P, Q) and (Q, V) are Fredholm pairs, and either $Q - V$ or $P - Q$ is compact, then (P, V) is a Fredholm pair and

$$i(P, Q) = i(P, V) + i(V, Q).$$

For Θ-a.e. $(P, V, Q) \in \mathsf{P}_c \times \mathsf{P}_c \times \mathsf{P}_c$, we have that

$$i(P, Q) = i(P, V) + i(V, Q) = 0, \qquad \Theta - a.e.$$

showing that $(\pi_{13})_\sharp \Theta \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\,\mu_\phi, \mathrm{tr}(\cdot)\,\mu_\varphi)$. For the same reason, Θ-a.e. the projections P, Q and V belong to the same connected component of P_c where triangular inequalities can be used. Hence for any $p \geq 1$:

$$\begin{aligned}
W_p(\phi, \varphi) &\leq \left(\int \mathsf{d}(P, Q)^p \, (\pi_{13})_\sharp \Theta(dP dQ) \right)^{1/p} \\
&= \left(\int \mathsf{d}(P, Q)^p \, \Theta(dP dV dQ) \right)^{1/p} \\
&\leq \left(\int (\mathsf{d}(P, V) + \mathsf{d}(V, Q))^p \, \Theta(dP dV dQ) \right)^{1/p} \\
&\leq \left(\int \mathsf{d}(P, V)^p \Theta(dP dV dQ) \right)^{1/p} + \left(\int \mathsf{d}(V, Q)^p \, \Theta(dP dV dQ) \right)^{1/p} \\
&= W_p(\phi, \psi) + W_p(\psi, \varphi),
\end{aligned}$$

concluding the proof. ∎

Corollary 4.15. *The couple $(\mathcal{S}_n(\mathsf{B}(\mathsf{H})), W_p)$ is an extended metric space in the sense that*

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

verifies for any φ and ψ 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 is straightforward to check that

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

In particular W_p-convergence implies $\mathcal{C}_p$-convergence and, by Theorem 4.11, w^*-convergence. However, as expected w^*-convergence does not imply W_p-convergence. We have a simple counterexample.

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

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

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

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

where P_1 and P_2 are the projections over the span of e_1 and e_2, respectively. Since P_1, P_2 and Id belong to two different connected components of P_c, $W_p\big(\mathrm{tr}(\cdot)\mu_{\varphi_n}, \mathrm{tr}(\cdot)\mu_\varphi\big) = \infty$.

5 Kantorovich duality for W_p and consequences

In this part we will go through the Kantorovich duality for the optimal transport problem over $(\mathsf{P}_c, \mathsf{d})$. In particular we will analyse cyclically montone 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 d to each connected component P_n of P_c and invoke the classical results. We start recalling the following classical definition from the theory of optimal transport: A subset Γ of $\mathsf{P}_c \times \mathsf{P}_c$ is 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 following inequality is valid

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

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 $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$, $\nu \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\, \mu_\varphi, \mathrm{tr}(\cdot)\, \mu_\psi)$ will be called d^p-cyclically monotone if there exists a d^p-cyclically monotone set Γ such that $\pi(\Gamma) = 1$.

By lower semicontinuity of d, it is well-known that d^p-cyclical monotonicity is a necessary condition for being optimal (see for instance [12, Proposition B.16]).

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

Looking at the transport on each single connected component of P_c, it is clear that cyclical monotonicity is indeed a sufficient condition for global optimality.

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

$$\int \mathsf{d}(P, Q)^p \, \nu(dPdQ) = W_p(\varphi, \psi)^p.$$

Proof. Decompose both μ_φ and μ_ψ into as sum of their restriction to each connected component of P_c, $\mathsf{P}_n = \{P \in \mathsf{P}_c \colon \mathrm{tr}(P) = n\}$. Then

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

with $\mu_{\varphi,n}$ and $\mu_{\psi,n}$ having supports in P_n. Then any plan $\nu \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\, \mu_\varphi, \mathrm{tr}(\cdot)\, \mu_\psi)$ 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 $\nu \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\,\mu_\varphi, \mathrm{tr}(\cdot)\,\mu_\psi)$ and Γ a 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 ν_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 d restricted to P_n is finite and therefore, by classical theory of optimal transport (see for instance [51]), d^p-cyclical monotonicity is equivalent to optimality giving that each ν_n is optimal and therefore optimality of ν follows. □

From the classical theory [51, Theorem 5.10], the following dual formulation of the problem is valid: for any $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$

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

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

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

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

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

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

Theorem 5.10 of [51] gives that the previous duality can be rewritten as follows

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

and in the above supremum one might as well impose that f be d^p-convex. The previous supremum is actually achieved and the maximum will be called a Kantorovich potential.

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

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

In particular, $\nu \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\,\mu_\varphi, \mathrm{tr}(\cdot)\,\mu_\psi)$ is W_p-optimal if and only if

$$\nu\Big(\big\{(P,Q) \in \mathsf{P}_c \times \mathsf{P}_c \colon f^{\mathsf{d}^p}(Q) - f(P) = \mathsf{d}(P,Q)^p\big\}\Big) = 1.$$

Proof. Reasoning like in the proof of Proposition 5.2, on each connected component P_n of P_c, the metric d is continuous yielding (see [51, Theorem 5.10]) for each $n \in \mathbb{N}$ the existence of d^p-convex functions $f_n : \mathrm{supp}(\mu_{\varphi,n}) \to \mathbb{R}$, meaning that it is not identically $+\infty$ and there exists $h : \mathrm{supp}(\mu_{\psi,n}) \to \mathbb{R} \cup \{\pm\infty\}$ such that for each $P \in \mathrm{supp}(\mu_{\varphi,n})$

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

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

$$\Big\{(P,Q) \in \mathrm{supp}(\mu_{\varphi,n}) \times \mathrm{supp}(\mu_{\psi,n}) \colon f_n^{\mathsf{d}^p}(Q) - f_n(P) = \mathsf{d}(P,Q)^p\Big\},$$

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

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

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

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

giving that f is d^p-convex. Simply noticing that d takes value $+\infty$ if P and Q does not belong to the same connected component of P_c, it follows that for $Q \in \mathrm{supp}(\mu_{\psi,n})$ satisfies $f^{\mathsf{d}^p}(Q) = f_n^{\mathsf{d}^p}(Q)$, where f^{d^p} is its d^p-transform, given by

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

Hence $W_p(\varphi,\psi)^p = \int f^{\mathsf{d}^p} \mathrm{tr}(\cdot)\, \mu_\psi - \int f \mathrm{tr}(\cdot)\, \mu_\varphi$, and the second claim follows straightforwardly. □

We now focus on representing Kantorovich potentials.

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

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

Proof. Let $\sum_i \lambda_i P_{V_i}$ be the spectral decomposition of ρ_φ with strictly decreasing eigenvalues. Then f defines a Borel function on the spectrum of ρ_φ with $f(\lambda_i) := f(P_i)$ and possibly $f(0) = 0$. We simply define $C := f(\rho_\varphi)$ by the functional calculus. In particular

$$\mathrm{Dom}(C) = \Big\{ x \in \mathsf{H} : \sum_i |f(P_{V_i})|^2 \|P_{V_i} x\|^2 < \infty \Big\}$$

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

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

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

$$\nu\left(\Big\{ (P,Q) \in \mathrm{supp}(\mu_\varphi) \times \mathrm{supp}(\mu_\psi) \colon \mathrm{tr}(C^{\mathsf{d}^p} Q) - \mathrm{tr}(CP) = \frac{\mathsf{d}(P,Q)^p}{\mathrm{tr}(P)} \Big\} \right) = 1;$$

with C, C^{d^p} are such that

$$\mathrm{tr}(C^{\mathsf{d}^p} Q) - \mathrm{tr}(CP) \le \frac{\mathsf{d}(P,Q)^p}{\mathrm{tr}(P)}, \qquad \forall\, (P,Q) \in \mathrm{supp}(\mu_\varphi) \times \mathrm{supp}(\mu_\psi),\ \mathrm{tr}(P) = \mathrm{tr}(Q).$$

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

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

Then apply Lemma 5.5 to such f to obtain C such that $f(P)\mathrm{tr}(P) = \mathrm{tr}(CP)$ for all $P \in \mathrm{supp}(\mu_\varphi)$, implying

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

Denoting with C^{d^p} 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. □

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

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

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

is a $\mathcal{C}_p$-geodesic (resp. a W_p-geodesic) if $\mathcal{C}_p(\phi_t, \phi_s) = |t-s|\mathcal{C}_p(\varphi,\psi)$ (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 $\mathcal{P}(\mathsf{P}_c)$.

Proposition 5.8. *The set $\mathcal{D}_1(\mathsf{P}_c)$ of discrete, non-negative measures having integral of the trace equal to 1 as defined in (3.3) is weakly convex with respect to $W_p^{\mathsf{P}_c}$ in the following sense: for any $\mu_0, \mu_1 \in \mathcal{D}_1(\mathsf{P}_c)$ such that $W_p^{\mathsf{P}_c}\big(\mathrm{tr}(\cdot)\mu_0, \mathrm{tr}(\cdot)\mu_1\big) < \infty$, there exists a curve $(\mu_t)_{t\in[0,1]} \subset \mathcal{D}_1(\mathsf{P}_c)$ with initial point μ_0 and final point μ_1 such that $t \longmapsto \mathrm{tr}(\cdot)\mu_t$ is a $W_p^{\mathsf{P}_c}$-geodesic.*

Proof. Given $\mu_0, \mu_1 \in \mathcal{D}_1(\mathsf{P}_c)$ such that $W_p^{\mathsf{P}_c}\big(\mathrm{tr}(\cdot)\mu_0, \mathrm{tr}(\cdot)\mu_1\big) < \infty$, Theorem 4.3 ensures the existence of an optimal transport plan $\nu \in \Pi_{\mathsf{d}}(\mu_0, \mu_1)$. 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

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

Since ν is admissible, whenever $\theta_{i,j} > 0$ it follows that $\mathrm{tr}(P_{V_i}) = \mathrm{tr}(P_{Z_j})$ hence we can consider $\gamma_{i,j}$ any geodesic of $(\mathsf{P}_c, \mathsf{d})$ connecting V_i to Z_j. Its existence is assured by the fact that V_i and Z_j belong same connected component of $(\mathsf{P}_c, \mathsf{d})$. In particular $\mathrm{tr}(\gamma_{i,j}(t))$ is constant for each $t \in [0,1]$ and depends only on i.
Now define the following non-negative measure over $\mathrm{Geo}(\mathsf{P}_c)$

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

and, denoting by $\mathrm{e}_t : \mathrm{Geo}(\mathsf{P}_c) \to \mathsf{P}_c$ the evaluation map at time t we have a curve of measures $[0,1] \ni t \mapsto \mu_t := (\mathrm{e}_t)_\sharp(\boldsymbol{\gamma})$. First notice that $\mu_t \in \mathcal{D}_1(\mathsf{P}_c)$: indeed $\boldsymbol{\gamma}$ is a discrete measure therefore the same is valid for μ_t and

$$\int_{\mathsf{P}_c} \mathrm{tr}(P)\, \mu_t(dP) = \int_{\mathsf{P}_c} \mathrm{tr}(P)\, (\mathrm{e}_t)_\sharp(\boldsymbol{\gamma})(dP) \ = \sum_{i,j} \theta_{i,j} = 1.$$

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

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

This proves the claim. □

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

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

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

Then there exists a bijective map $T : \mathrm{supp}(\mu_0) \to \mathrm{supp}(\mu_1)$ such that $(Id, T)_\sharp \mu_0 \in \Pi_{\mathsf{d}}(\mu_0, \mu_1)$ is an optimal plan. In particular, if $\mu_0 = \sum_i \alpha_i P_{0,i}$, $\mu_1 = \sum_i \beta_i P_{1,i}$ with $\alpha_i > \alpha_{i+1}$ and $\beta_i > \beta_{i+1}$, then

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

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

Proof. From the classical theory of optimal transport applied to each connected component of P_c, $\mathrm{tr}(\cdot)\mu_t = (\mathrm{e}_t)_\sharp(\boldsymbol{\gamma})$ with $\boldsymbol{\gamma} \in \mathcal{P}(\mathrm{Geo}(\mathsf{P}_c))$. Hence $(\mathrm{e}_0, \mathrm{e}_1)_\sharp\boldsymbol{\gamma} \in \Pi_{\mathsf{d}}(\mathrm{tr}(\cdot)\mu_\varphi, \mathrm{tr}(\cdot)\mu_\psi)$. Posing $\nu = (\mathrm{e}_0, \mathrm{e}_1)_\sharp\boldsymbol{\gamma}$, necessarily

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

Assume now by contradiction 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 exists $\gamma_{i_1,j}, \gamma_{i_2,j} \in \mathrm{Geo}(\mathsf{P}_c)$ such that

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

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

The argument can be reverted and implies that for each $i \in \mathbb{N}$ there is only one $j \in \mathbb{N}$ such that $\theta_{i,j} > 0$ and for each $j \in \mathbb{N}$ there is only one $i \in \mathbb{N}$ such that $\theta_{i,j} > 0$: this is equivalent to the existence of a bijective map $T : \mathrm{supp}(\mu_\varphi) \to \mathrm{supp}(\mu_\psi)$ such that

$$\nu = (\mathrm{Id} \times T)_\sharp \mu_0,$$

proving the first part of the claim. The remaining claims are straightforward consequences. □

Corollary 5.10. *Fix $p \geq 1$. Given $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ with $W_p(\varphi, \psi) < \infty$, consider μ_φ, μ_ψ and any μ_t from Proposition 5.8. If $\mu_t \in \mathcal{D}_1^\perp(\mathsf{P}_c)$, then posing*

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

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

Proof. To fix notation, μ_t from Proposition 5.8 can be then 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 $\mathcal{C}(\mathsf{H})$. From Proposition 5.9 it follows that

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

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

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

proving the claim. □

Remark 5.11. If the condition $\mu_t \in \mathcal{D}_1^\perp(\mathsf{P}_c)$ is not known, then one can anyway define a curve of normal states because $\Psi(\mu_t) = \sum_i \alpha_i P_i(t) =: \rho_t$ is a well-defined element of $\mathcal{C}(\mathsf{H})$ implying that $\varphi_{\rho_t} \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ (see Lemma 2.8). However, the spectral decomposition of ρ_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 it was not directly used the fact that $\mathrm{tr}(\cdot)\mu_t$ is a $W_p^{\mathsf{P}_c}$-geodesic, rather that there exists $\boldsymbol{\gamma} \in \mathcal{P}(\mathrm{Geo}(\mathsf{P}_c))$ such that

$$\mathrm{tr}(\cdot)\mu_t = (\mathrm{e}_t)_\sharp \boldsymbol{\gamma}, \quad \mu_t \in \mathcal{D}_1^\perp(\mathsf{P}_c).$$

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

Proposition 5.9 admits a partial converse.

Proposition 5.13. *Let $\varphi, \psi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ be given states with $W_p(\varphi, \psi) < \infty$. If there exists a bijective map $T : \mathrm{supp}(\mu_\varphi) \to \mathrm{supp}(\mu_\psi)$ such that $(\mathrm{Id} \times T)_\sharp \mu_\varphi \in \Pi_{\mathsf{d}}(\mu_\varphi, \mu_\psi)$ then φ and ψ 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 λ_i. The condition $(\mathrm{Id} \times T)_\sharp \mu_\varphi \in \Pi_{\mathsf{d}}(\mu_\varphi, \mu_\psi)$ implies that the spectral decomposition of ρ_ψ 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 ρ_φ. 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 $\mathrm{Ker}(\rho_\varphi) \cong \mathrm{Ker}(\rho_\psi)$. By [18, Proposition 7] we find an invertible $u \in \mathsf{B}(\mathsf{H})$ such that $\rho_\psi = u\rho_\varphi u^*$. In our case this u is unitary. Concretely let $\{e_i, e'_j\}$ an orthonormal system adapted to $\mathsf{H} = R(\rho_\varphi) \oplus N(\rho_\varphi)$ and $\{f_i, f'_j\}$ a corresponding one for ρ_ψ; 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 $\rho_\varphi = \sum_i \lambda_i P_{V_i}$. □

6 Tensor product interpretation: a generalization

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

To fix notations we will use

$$\mathfrak{J}(\varphi,\psi):=\Big\{\Xi\in\mathcal{S}_n(\mathsf{B}(\mathsf{H}\otimes\mathsf{H}))\colon J_\flat^1\Xi=\varphi,\ J_\flat^2\Xi=\psi\Big\}. \tag{6.1}$$

In this section we will 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. We will follow [34], and the appendix B.1 for basics on antilinear operators. Let $AHS(\mathsf{H})$ be the space of the antilinear Hilbert–Schmidt operators acting on H. Firstly an antilinear operator is an additive operator $T:\mathsf{H}\to\mathsf{H}$ such that $T(\lambda x)=\overline{\lambda}x$. 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 $\xi,\eta\in\mathsf{H}$. On such operators we define the Hilbertian product (conjugate-linear in the first entry) $\langle A,B\rangle:=\mathrm{tr}(A^*B)$ and we complete the linear span of all the operators in the form $T_{\xi,\eta}$ with respect to this Hilbert structure. If we compute

$$\langle T_{\xi,\eta},T_{x,y}\rangle=\mathrm{tr}(T_{\eta,\xi}T_{x,y})=\mathrm{tr}\Big(\zeta\mapsto\langle x,\zeta\rangle\langle y,\eta\rangle\xi\Big)=\langle\xi,x\rangle\langle y,\eta\rangle.$$

On the right we have the inner product defined on $\mathsf{H}\otimes\mathsf{H}$ 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

$$\Theta:\mathsf{H}\otimes\mathsf{H}\longrightarrow AHS(\mathsf{H}) \tag{6.2}$$

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

$$\Theta^*_{\xi\otimes\eta}=\Theta_{\eta\otimes\xi},\quad |\Theta_\zeta|^2=\mathrm{Tr}_2\,|\zeta\rangle\langle\zeta|\quad\text{and}\quad|\Theta^*_\zeta|^2=\mathrm{Tr}_1\,|\zeta\rangle\langle\zeta|, \tag{6.3}$$

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

$$\mathsf{P}_c(R(W^*W))\cong\big\{P\in\mathsf{P}_c:P\le W^*W\big\}$$

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

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

for $P\in\mathsf{P}_c\colon P\le R(W^*W)$. We define

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

6.1 Pure States as transport maps

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

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

We will associate to ω_ζ a unique family of transport plans from the spectral-projection measures of φ_{ρ_1} to the one of φ_{ρ_2}.

We write the polar decomposition (see the appendix B.2) of the antilinear operator Θ_ζ associated via (6.2) to ζ. Thus $\Theta_\zeta=U_\zeta|\Theta_\zeta|=|\Theta^*_\zeta|U_\zeta$ and $|\Theta^*_\zeta|=U_\zeta|\Theta_\zeta|U^*_\zeta$. By (6.3) we see that $|\Theta_\zeta|=\rho_1^{1/2}$ and $|\Theta^*_\zeta|=\rho_2^{1/2}$. It follows

$$\Theta_\zeta=U_\zeta\,\rho_1^{1/2},\quad\Theta_\zeta=\rho_2^{1/2}\,U_\zeta\quad\text{and}\quad\rho_2=U_\zeta\,\rho_1\,U^*_\zeta. \tag{6.4}$$

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

$$N(U_\zeta)=N(\rho_1)\quad\text{and}\quad U^*_\zeta U_\zeta\mathsf{H}=N(\rho_1)^\perp$$

that we will always consider being satisfied.

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

Proposition 6.1. *The following map is well defined*

$$\Upsilon : \mathcal{PS}_n(\mathsf{H} \otimes \mathsf{H}) \longrightarrow \mathcal{G}(\mathsf{H}), \quad \omega_\zeta \longmapsto \Big(R(U_\zeta^* U_\zeta), \widetilde{U_\zeta}, R(U_\zeta U_\zeta^*)\Big).$$

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* $\lambda \in \mathsf{U}(1)$.

Proof. If ζ is changed into $\lambda\zeta$ for a phase $\lambda \in \mathsf{U}(1)$ 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_ζ and U_η 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

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

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 $\mathsf{U}(1)$. But f has to be constant by the antilinearity of our isometries. □

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

$$\mathcal{M}(\mathsf{H}) := \Big\{(\varphi_1, F, \varphi_2) \colon \varphi_1, \varphi_2 \in \mathcal{S}_n(\mathsf{B}(\mathsf{H})),\ F \in \mathrm{Bij}(\Lambda_{\varphi_1}^\perp, \Lambda_{\varphi_2}^\perp)\Big\}. \tag{6.5}$$

then we have a map

$$\begin{aligned} &\Phi_\otimes : \mathcal{PS}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H})) \longrightarrow \mathcal{M}(\mathsf{H}) \\ &\omega_\zeta \longmapsto \Big(\varphi_1, (\widetilde{U_\zeta})_\sharp, \varphi_2\Big) \end{aligned}$$

$$\begin{aligned} &\text{marginals}\, \varphi_1, \varphi_2 \\ &\varphi_1 = \varphi_{\rho_1},\ \varphi_2 = \varphi_{\rho_2} \\ &\rho_1 = \mathrm{Tr}_2\, |\zeta\rangle\langle\zeta|,\ \rho_2 = \mathrm{Tr}_1\, |\zeta\rangle\langle\zeta| \\ &\Theta_\zeta = U_\zeta \rho_1^{1/2} \end{aligned}$$

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

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

$$\begin{tikzcd} \mathcal{PS}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H})) \arrow[r, "F"] \arrow[d, "J_\flat^1"'] & \mathcal{D}_1^\perp(\mathsf{P}_c) \arrow[dl, "\Psi"] \\ \mathcal{S}_n(\mathsf{B}(\mathsf{H})) & \end{tikzcd}$$

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

$$\mathcal{PS}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H})) \xrightarrow{J_\flat^1} \mathcal{S}_n(\mathsf{B}(\mathsf{H})) \xrightarrow{\Phi} \mathcal{D}_1^\perp(\mathsf{P}_c). \tag{6.6}$$

Theorem 6.3. *The map* $\Phi_\otimes$ *is well defined and injective. Fix any* $\varphi \in \mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ *and a measure* $\mu \in \Lambda_\varphi^\perp$ *representing* φ*; then the* μ*-"component" of* $\Phi_\otimes$ *provides a map*

$$\Phi_\otimes^\mu : \Big\{\omega \in \mathcal{PS}_n(\mathsf{B}(\mathsf{H} \otimes \mathsf{H})) \colon J_\flat^1 \omega = \varphi\Big\} \longrightarrow \mathcal{D}(\mathsf{P}_c \times \mathsf{P}_c), \quad \Phi_\otimes^\mu(\omega_\zeta) = (\mathrm{Id} \times \widetilde{U_\zeta})_\sharp(\mathrm{tr}(\cdot)\mu).$$

This map is valued in the set of admissible transport plans

$$\Phi_\otimes^\mu(\omega_\zeta) \in \Pi_d\big(\mathrm{tr}(\cdot)\mu, (\widetilde{U_\zeta})_\sharp(\mathrm{tr}(\cdot)\mu)\big).$$

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

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

We have a compatibility property expressed by the commutative diagram

$$\begin{tikzcd}
\mathcal{PS}_n(\mathsf{B}(\mathsf{H}\otimes\mathsf{H})) \arrow[r, "\Phi^F_\otimes"] \arrow[d, "J^1_\flat"'] \arrow[dr, "F"] & \mathcal{D}^\perp_1(\mathsf{P}_c\times\mathsf{P}_c) \arrow[d, "\pi^1_\sharp"] \\
\mathcal{S}_n(\mathsf{B}(\mathsf{H})) & \mathcal{D}^\perp_1(\mathsf{P}_c). \arrow[l, "\Psi"]
\end{tikzcd}$$

When $F=\mathcal{F}$ as before (eq. (6.6)) this becomes a compatibility with Φ as the diagram

$$\begin{tikzcd}
\mathcal{PS}_n(\mathsf{B}(\mathsf{H}\otimes\mathsf{H})) \arrow[r, "\Phi^F_\otimes"] \arrow[d, "J^1_\flat"'] & \mathcal{D}^\perp_1(\mathsf{P}_c\times\mathsf{P}_c) \arrow[d, "\pi^1_\sharp"] \\
\mathcal{S}_n(\mathsf{B}(\mathsf{H})) \arrow[r, "\Phi"] & \mathcal{D}^\perp_1(\mathsf{P}_c)
\end{tikzcd}$$

commutes.

Proof. Among all the spectral measures associated to φ_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 ortonormal set of vectors $(e_i)_i$ spanning the initial domain of U_ζ and U_η 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 λ. 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_ζ with $P_{V_i}\le U_\zeta^*U_\zeta$. This means that for every i the points P_{V_i} and $\widetilde{U_\zeta}(P_{V_i})$ belong to the same connected component. □

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

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 $\mathsf{B}(\mathsf{H}\otimes\mathsf{H})$. In particular given $\varphi_1,\varphi_2\in\mathcal{S}_n(\mathsf{B}(\mathsf{H}))$ and $\omega_\zeta\in\mathcal{PS}_n(\mathsf{B}(\mathsf{H}\otimes\mathsf{H}))$, for each $\mu\in\Lambda^\perp_{\varphi_1}$ we have the transport plan $\Phi^\mu_\otimes(\omega_\zeta)$ (induced by the map $\widetilde{U}_\zeta$) between admissible representations of φ_1 and φ_2 whose p-cost will be

$$\int_{\mathsf{P}_c\times\mathsf{P}_c}\mathsf{d}^p(P,Q)\,\Phi^\mu_\otimes(\omega_\zeta)(dPdQ)=\int_{\mathsf{P}_c}\mathsf{d}^p(P,\widetilde{U}_\zeta(P))\mathrm{tr}(P)\,\mu(dP).$$

Hence, the cost of ω_ζ will be given by taking the lowest possible cost among all $\Phi^\mu_\otimes(\omega_\zeta)$:

$$\mathcal{C}_p(\omega_\zeta)^p:=\inf_{\mu\in\Lambda^\perp_{\varphi_1}}\int_{\mathsf{P}_c}\mathsf{d}^p(P,\widetilde{U}_\zeta(P))\mathrm{tr}(P)\,\mu(dP). \tag{6.7}$$

Following Proposition 4.9, it is equivalent to restrict the minimisation only among those μ concentrated inside P_1. Moreover the inf is actually attained giving that there exists $\mu\in\mathsf{P}_1$, a priori not unique and depending on $p\ge1$, such that

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

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

$$\mathcal{C}_p(J^1_\flat\omega_\zeta,J^2_\flat\omega_\zeta)\le\mathcal{C}_p(\omega_\zeta).$$

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 $\varphi:M\longrightarrow N$ 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. A 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 a 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 [38]). 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 canonical projection $\pi : E \longrightarrow M$.
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 $U \times G$. In the present context *isomorphic* means that we can find a diffeomorphism $\psi : \pi^{-1}(U) \to U \times G$ in the form $\psi(u) = (\pi(u), \varphi(u))$ for a smooth map $\varphi : \pi^{-1}(U) \to G$ satisfying $\varphi(u \cdot g) = \varphi(u)g$ for every $g \in G$.

To synthetize this definition we say that $E \longrightarrow M$ is a principal bundle.

Example 3. Every homogeneous space G/B is the base of a principal bundle. Indeed we can prove that $G \longrightarrow G/B$ 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 $\mathcal{V}_p := \mathrm{N}(d\pi : T_pE \to T_\pi M) \subset T_pE$ 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 $\widetilde{X} \in \Gamma(TE)$ (sections of the tangent bundle) with $\widetilde{X}_p := \left.\frac{d}{dt}\right|_{t=0} p\cdot(\exp tX)$. Fundamental vector fields are of course vertical and at every point the map $\mathfrak{g} \longrightarrow \mathcal{V}_p$ given by $X \longmapsto \widetilde{X}_p$ is an isomorphism. However in general there is no preferred choice of *horizontal* subspaces of TE. This is extra structure amounts to a *connection*.

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

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

A connection on E provides us with a notion of horizontal curves and horizontal liftings of curves. Moreover given any representation $G \longrightarrow \mathrm{End}(V)$ on a vector space, a classical construction going under the name of *associated bundle construction* produces a vector bundle $W \to M$ 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 TM 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 [13] (there the inner product is linear in the first entry). An antilinear operator $T : \mathsf{H} \longrightarrow \mathsf{K}$ is an additive operator such that $T\lambda\xi = \overline{\lambda}T\xi$.

Let $J : \mathsf{H} \longrightarrow \mathsf{H}$ be antilinear and isometric: $\|J\xi\| = \|\xi\|$ for every ξ. 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 $J : \mathsf{H} \longrightarrow \mathsf{H}$ is called an *involution* if $J^2 = \mathrm{Id}_{\mathsf{H}}$. It follows that J is an antiunitary. Involutions always exist for every Hilbert space and are very useful: if $T : \mathsf{H} \longrightarrow \mathsf{K}$ 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 $T : \mathsf{H} \longrightarrow \mathsf{K}$ be antilinear bounded, then the adjoint of T is the unique antilinear bounded operator $T^* : \mathsf{K} \longrightarrow \mathsf{H}$ such that

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

It satisfies: $(\lambda T)^* = \lambda T^*$ as opposite to the behaviour of the adjoint for linear operators. Using an involution on H we can compute $T^* := J(TJ)^*$ in terms of the adjoint of a linear operator.

B.2 Polar decompositions

A bounded operator $T : \mathsf{H} \longrightarrow \mathsf{K}$ is a partial isometry if T^*T is a projection P. Therefore $P\mathsf{H} = N(T)^\perp$ 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 $N(T)^\perp \longrightarrow R(T)$.

Theorem B.1. *(Left polar decomposition) Any $T \in \mathsf{B}(\mathsf{H}, \mathsf{K})$ (two Hilbert spaces) has the decomposition $T = UP$ for a non negative operator $P : \mathsf{H} \to \mathsf{H}$ and a partial isometry $U : \mathsf{H} \to \mathsf{K}$. This decomposition is unique if we require that $N(U) = N(P)$. Equivalently if we require that the initial support $(U^*U)\mathsf{H}$ 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

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

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 $U^*U = \operatorname{Proj} \overline{R(P)}$ 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. □

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 $TT^* = U|T|^2U^*$ which we can iterate getting for every power: $(TT^*)^n = U|T|^nU^*$. 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 $T : \mathsf{H} \longrightarrow \mathsf{K}$. Using an involution as before we can construct polar decompositions

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

for $|T| = \sqrt{T^*T}$ a linear operator while V is an antilinear partial isometry with $V^{**}V = \operatorname{Proj}(N(T))^\perp$ and $VV^* = \operatorname{Proj} \overline{R(T)}$. 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] W.O. Amrein and K. Sinha *On Pairs of Projections in a Hilbert Space* Linear Algebra and its Applications, 208-209, 425–435, (1994).

[2] C.J. Atkin *The Finsler geometry of groups of isometries of Hilbert space.* J. Austral. Math. Soc. Ser. A 42, no. 2, 196–222, (1987).

[3] 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`

[4] E. Andruchow: *The Grassmann manifold of a Hilbert space* , Proceedings of the XIIth "Dr. Antonio A. R. Monteiro" Congress, 41–55, Actas Congr. "Dr. Antonio A. R. Monteiro", Univ. Nac. del Sur, Bahía Blanca, 2014.

[5] E. Andruchow: *Operators which are the difference of two projections*, J. Math. Anal. Appl. 420, 1634–1653, (2014).

[6] E. Andruchow: *Pairs of Projections: Geodesics, Fredholm and Compact Pairs*, Complex Anal. Oper. Theory 8, 1435–1453, (2014).

[7] E. Andruchow and A. Varela: *Riemannian geometry of finite rank positive operators*, Differential Geometry and its Applications Volume 23, Issue 3, Pages 305–326.

[8] E. Alfsen, and F. Shultz: *Unique decompositions, faces, and automorphisms of separable states*, J. Math. Phys. 51, 052201 (2010).

[9] J. Avron, R. Seiler, and B. Simon: *The index of a pair of projections*, J. Funct. Anal. 120(1), 220–237 (1994).

[10] J.-D. Benamou and Y. Brenier: *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., **84**, (2000), 375–393.

[11] P. Biane and D. Voiculescu: *A free probability analogue of the Wasserstein metric on the trace-space space*, GAFA, **11** (2009), 353–454.

[12] S. Bianchini and L. Caravenna: *On the extremality, uniqueness and optimality of transference plans*, Bull. Acad. Sinica, **4** (2009), 353–454.

[13] B. Blackadar *Operator Algebras, Theory of C^*–algebras and von Neumann Algebras* Encyclopedia of Mathematical Sciences, Springer–Verlag Berlin Heidelberg, 2006.

[14] Bogachev: *Measure Theory*,

[15] 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.

[16] 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.

[17] E. A. Carlen and J. Maas: *Non-commutative calculus, Optimal Transport and functional inequalities in dissipative quantum systems*, J. Statist. Phys. , to appear 10.1007/s10955-019-02434-w.

[18] F.M. Ciaglia, A. Ibort, J. Jost, G. Marmo: *Manifolds of classical probability distributions and quantum density operators in infinite dimensions*, Info. Geo. **2** (2019), 231–271.

[19] F. Cipriani and J.-L. Sauvageot: *Derivations as square roots of Dirichlet forms*, J. Funct. Anal., **201**, (2003), 78–120.

[20] E. Chiumiento *Hopf-Rinow theorem in Grassmann manifolds of C^*-algebras*, Geometric Methods in Physics. XXVII Workshop Trends in Mathematics, Springer Nature Switzerland 145–152 (2019).

[21] K. Chrzaszcz, J. Jachymski and F. Turobos: *On characterizations and topology of regular semimetric spaces* Publ. Math. Debrecen **93/1-2** (2018), 87–105 DOI:10.5486/PMD.2018.8049.

[22] A. Connes: *Compact metric spaces, Fredholm modules and hyperfiniteness*, Ergodic Theory and Dynamical Systems **9**, 207–220, (1989).

[23] A. Connes: *Noncommutative geometry*, Academic Press Inc., San Diego, CA, 1994.

[24] G. Corach, H. Porta and L. Recht: *The Geometry of Spaces of Projections in C^*-algebras*, Adv. Math., 101, 59–77, (1993).

[25] F. D'andrea, P. Martinetti *A dual formula for the spectral distance in noncommutative geometry*, J. Geom. Phys., **159**, 59–64, (2021).

[26] J.L. Danny *The support of extremal measures with given marginals*, Michigan Math. J., **27** (1980), 59–64.

[27] I. Dimitrić *A note on equivariant embeddings of Grassmanians*, Publications de l'insitut math. (Beograd) (N.S.) **59** (73), (1996) 131–137.

[28] J. Dixmier: *von Neumann algebras*, with a preface by E. C. Lance. Translated from the second French edition by F. Jellett. North-Holland Mathematical Library, 27. North-Holland Publishing Co., Amsterdam-New York, 1981

[29] A. Edelman, T. Arias and S. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. 20, 303–353, (1998)

[30] 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, Volume 35, Number 17

[31] F. Golse, C. Mouhot, and T. Paul: *On the mean-field and classical limits of quantum mechanics*, Comm. Math. Phy. **343**, (2016), 165–205.

[32] KC. Hannabuss *Notes on Quantum Computing*, Oxford, HT, (2008), `http://users.ox.ac.uk/~kch/demos/quantc7/`

[33] P. Harms and A. C. G Mennucci *Geodesics in infinite dimensional Stiefel and Grassmann manifolds*, C. R. Acad. Sci. Paris, Ser. I, 350, 773–776, (2012)

[34] F. Herbut *On bipartite pure–state entanglement structure in terms of disentanglement*, Journal of Mathematical Physics 47, 122103 (2006).

[35] D. Hornshaw: *Quantum optimal transport for approximately finite-dimensional C^*-algebras preprint*, arXiv:1910.03312.

[36] T. Kato: *Perturbation Theory for Linear Operators*, volume 132 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1980.

[37] 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, RI, 1997

[38] S. Kobayashi and T. Nomitzu *Foundations of differential geometry*

[39] A. Kriegl and P. Michor *The convenient setting of global analysis,* Mathematical Surveys and Monographs, 53. American Mathematical Society, Providence, 1997.

[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, Vol. 20, 2, 97–116, 2002.

[41] V. Jurdjevic, I. Markina and F. Silva Leite *Extremal Curves on Stiefel and Grassmann Manifolds*, The Journal of Geometric Analysis 30, 3948–3978, (2020).

[42] K. Landsman: *Foundations of Quantum Theory*, Vol. 188. Fundamental Theories of Physics. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-51777-3.

[43] J. Lawson: *Ordered probability spaces*, arXiv:1612.03213.

[44] S. Lang: *Differential and Riemannian manifolds*, Graduate Texts in Mathematics, **160**. Springer-Verlag, New York, 1995.

[45] M. Rieffel: *Metric on state spaces*, Documenta Math. **4**, 559–600, (1999).

[46] N Lebedeva and A. Petrunin *Curvature bounded below: a definition a la Berg–Nikolaev* Elec. res. ann. in math. sc. 17, (2020), 122–124.

[47] E. Marczewski and P. Sikorski: *Measures in Nonseparable Metric Spaces*, Colloq. Math. 1 (1948), 133–139.

[48] A. Peres: *Separability Criterion for Density Matrices*, Physical Review Letters, 1996 - APS.

[49] H. Porta and L. Recht: *Minimality of geodesics in Grassmann manifolds*, Proc. Am. Math. Soc. 100, 3, 464–466, (1987).

[50] D. Robinson *Normal and locally normal states*, Comm. Math. Phys. 219–234, (1070).

[51] C. Villani: *Optimal transport - old and new*, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009.

[52] W. A. Wilson: *On Semi-Metric Spaces*, Amer. J. Math., **53** (1931), 361–373

[53] M. Wirth: *A Noncommutative Transport Metric and Symmetric Quantum Markov Semigroups as Gradient Flows of the Entropy*, preprint, arXiv:1808.05419.

[54] N.E Wegge-Olsen *K-theory and C^*-algebras. A friendly approach.* Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1993