Eötvös Loránd University

Faculty of Science

Institute of Mathematics

GEOMETRY OF QUANTUM WASSERSTEIN SPACES

Eszter Szabó

BSc in Mathematics

Supervisors: Dániel Virosztek, Balázs Maga

Budapest

2025

Contents

1 Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dániel Virosztek, who introduced me to this fascinating interface between mathematics and physics and suggested the topic of this thesis. His patience, availability and enthusiasm made the process of writing this thesis a truly enjoyable experience. Our regular consultations throughout the year greatly helped me acquire a solid understanding of optimal transport.

I am also thankful to Balázs Maga for suggesting some possible thesis topics and for assisting me in finding a supervisor.

I am deeply grateful to my family for sparking my interest in mathematics during my childhood and for their constant support throughout my studies.

Finally, I would like to thank my friends who motivated me during my BSc studies and offered help whenever I needed it.

2 Introduction

2.1 Motivation and history

The problem of optimal mass transportation was introduced by Monge at the end of the 18th century and then further improved by Kantorovich in the 1940s. Around the 1980s, analysts started to work in the field of optimal transport (abbreviated OT) more significantly, and since then it has evolved into an ever-growing area of research in analysis. Among the diverse application areas of optimal transport we should highlight the solutions of partial differential equations, geometry, stochastic analysis and various topics in mathematical physics, especially in fluid mechanics. Furthermore, the optimal transportation of information plays an important role in the mathematical theory of artificial intelligence, machine learning and image processing, so the recent rapid development of these areas attracts even more attention to OT [16]. Obviously the nature of the problem also implies application opportunities in Economics, since the transport cost minimization problem can be efficiently adapted to real life transportation problems.

Quantum optimal transport problems (abbreviated QOT), i.e., the classical optimal transport problems formulated using the framework of quantum mechanics are also of growing interest in the last decades. The non-commutativity of the quantities used in the quantum formulation usually makes the questions more difficult, thereby the solutions require a different approach. QOT is a widely applicable area as well, particularly in quantum information theory, but it can be used for quantum walks, quantum automata and quantum games, too [10].

The purpose of my thesis is to provide a brief introduction to both the theory of classical transport and the mathematical framework used in quantum mechanics, then, by fusing these two areas, to briefly describe different approaches to quantum optimal transport problems. Among these approaches, we are going to deal with the quantum channel formulation by De Palma and Trevisan [6] in more detail.

2.2 The classical optimal transport problem

Assume that we have some bakeries located in a country which produce a given quantity of bread every day. We also know how much bread a town needs per day to feed its population and we have a function which describes the cost of transporting a unit mass of bread from bakery x to town y. Our aim is to create a plan for transporting the bread from bakeries to towns so that the total cost of the transportation is minimal.

The classical optimal transport problem was first formulated mathematically by the French mathematician Gaspard Monge in 1781 in the following way [13]. We define the problem using the notations from Topics in Optimal Transportation by Cédric Villani [20].

We are given two measure spaces $(X, \mathscr{A}_X, \mu)$ and $(Y, \mathscr{A}_Y, \nu)$ on the Polish spaces (separable, complete metric spaces) X and Y such that $\mu(X) = \nu(Y) < \infty$. Without loss of generality we can assume that $\mu(X) = \nu(Y) = 1$, therefore μ and ν are both probability measures. The measure μ describes the initial distribution of products on the set X and ν describes the required distribution of products after the transportation. Moreover the cost function $c(x,y) : X \times Y \to \mathbb{R}_+ \cup \{+\infty\}$ is a measurable function which describes the cost of transporting one unit of the product from location x to location y.

Monge formulated the problem in the way that we cannot split the mass of products, i.e. we have to transport all the products from location x to the same location. In this case a transportation plan can be given by a measurable map $T : X \to Y$, where we transport the mass from location x to location $T(x)$.

Definition 2.1 (Pushforward of a measure)**.** *If $T : X \to X$ is a map and $(X, \mathscr{A}, \mu)$ is a measure space, for $A \subset X$ we denote $(T\#\mu)[A] := \mu[T^{-1}(A)] = \mu[x \in X : T(x) \in A]$ and call it the pushforward of μ by T.*

It is obvious that among the measurable maps $T : X \to Y$ the ones which describe a transportation from μ to ν are exactly those for which $\nu = \mu \circ T^{-1}$.

The total cost of the transportation given by the transport map T is the following integral:

$$I[T] := \int_X c(x, T(x)) d\mu(x).$$

Optimal transport plans can be found by minimizing on the set of all maps T which meet the above described conditions.

The problem remained unsolved for a long time and some further modification of the conditions was needed to create a more tangible formulation of the problem.

The most conspicuous issue with Monge's formulation is that such a map T which meets the conditions does not always exist. Indeed, let us consider $X = Y = [-1,1]$ with Dirac-masses: $\mu = \delta_0, \nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. All the mass from $x = 0$ must be transported into the same point, so no transfer map will have image ν.

It took more than 150 years until a Russian mathematician Leonid Vitalyevich Kantorovich suggested a relaxed version of the optimal transport problem [11]. The starting point is similar, we are again given two probability distributions: μ and ν on the sets X and Y. The set of probability measures on the set X is denoted as $\mathrm{Prob}(X)$. However, now instead of transfer maps we consider transference plans, which are probability measures on the product space $X \times Y$.

Definition 2.2 (Transference plan)**.** *A transference plan between the probability measures $\mu \in Prob(X)$ and $\nu \in Prob(Y)$ is a probability measure $\pi \in Prob(X \times Y)$ such that*

$$\int_Y d\pi(x,y) = d\mu(x) \text{ and } \int_X d\pi(x,y) = d\nu(y),$$

or in other words $\pi[A \times Y] = \mu[A]$ and $\pi[X \times B] = \nu[B]$ for all measurable subsets $A \subset X, B \subset Y$.

This definition describes a π with marginals μ and ν, so it meets our expected condition that the total mass transported from location x is $d\mu(x)$ and the total mass transported to location y is $d\nu(y)$.

In this case the transport cost corresponding to a given probability measure π is

$$I[\pi] := \int_{X \times Y} c(x,y) d\pi(x,y).$$

Let $\Pi(\mu,\nu)$ denote the set of transference plans from μ to ν. Using these notations Kantorovich's optimal transport problem is formulated as finding

$$\inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{X \times Y} c(x,y) d\pi(x,y) \right).$$

One of the most important aspects in which Kantorovich's problem differs from Monge's is that now mass from a given location can be split. The set $\Pi(\mu,\nu)$ is always nonempty, since $\mu\otimes\nu\in\Pi(\mu,\nu)$.

The Kantorovich problem with a specific class of cost functions admits a dual formulation.

Definition 2.3 (Lower semi-continuous function)**.** *A function $f: X\to\mathbb{R}\cup\{-\infty,+\infty\}$ is lower semi-continuous at $x_0\in X$ if* $\liminf_{x\to x_0} f(x)\geq f(x_0)$.

Theorem 2.4 (Kantorovich duality, Theorem 1.3. in[20])**.** *Let X,Y be Polish spaces, $\mu\in Prob(X),\nu\in Prob(Y)$ and let $c: X\times Y\to\mathbb{R}_+\cup\{+\infty\}$ be a lower semi-continuous function.*

If $\Phi_c=\{(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu): \varphi(x)+\psi(y)\leq c(x,y)$ for $(\mu\otimes\nu)$-every $(x,y)\}$ and $J(\varphi,\psi)=\int_X\varphi d\mu+\int_Y\psi d\nu$, then

(i)

$$\inf_{\Pi(\mu,\nu)} I[\pi]=\sup_{\Phi_c} J(\varphi,\psi).$$

(ii) $\inf_{\Pi(\mu,\nu)} I[\pi]=\min_{\Pi(\mu,\nu)} I[\pi]$, *i.e. the infimum is attained.*

In the case of $X=Y$ and the cost function being a distance (proper metric), that is $c(x,y)=d(x,y)$, the cost of the optimal transport

$$W_1(\mu,\nu):=\inf_{\pi\in\Pi(\mu,\nu)}\left(\int_{X\times X} d(x,y)d\pi(x,y)\right)$$

is a distance between probability distributions over X called the Wasserstein distance of order 1. Similarly we can define the Wasserstein distance of order p for all $p\geq 1$ values as

$$W_p(\mu,\nu):=\inf_{\pi\in\Pi(\mu,\nu)}\left(\int_{X\times X} d(x,y)^p d\pi(x,y)\right)^{1/p}.$$

2.3 Prerequisites from linear algebra

The formulation of quantum mechanics is based on linear algebra, but physicists tend to use different notations than linear algebra textbooks for mathematicians, so let us start this chapter by

reviewing the most important algebraic concepts and notations. Most of our notations and definitons are the same as in [14], which is considered a fundamental book in quantum information theory.

In quantum mechanics we usually consider a finite- or infinite-dimensional complex vector space V as the starting point of our calculations. In 1939, Paul Dirac proposed the use of the so-called bra-ket notation, which has since become a standard notation used in quantum mechanics. In this system vectors are denoted as $|\varphi\rangle$ and we call this object a *ket*.

On the vector space V we define an inner product $(,)$, which is a mapping $V \times V \to \mathbb{C}$ satisfying the following properties:

(i) conjugate symmetry: $(|\varphi\rangle, |\psi\rangle) = \overline{(|\psi\rangle, |\varphi\rangle)}$,

(ii) linearity in the second argument: $(|\xi\rangle, z_1 \cdot |\varphi\rangle + z_2 \cdot |\psi\rangle) = z_1 \cdot (|\xi\rangle, |\varphi\rangle) + z_2 \cdot (|\xi\rangle, |\psi\rangle)$,

(iii) positive definiteness: $(|\varphi\rangle, |\varphi\rangle) \geq 0$ for all $|\varphi\rangle$ and $(|\varphi\rangle, |\varphi\rangle) > 0$ for all $|\varphi\rangle \neq 0$.

From the first and second properties it is easy to show that the inner product is conjugate linear in its first argument. The standard inner product on $\mathbb{C}^n$ is defined as $(\mathbf{v}, \mathbf{w}) := \sum_{i=1}^{n} \overline{v_i} \cdot w_i$.

The standard bra-ket notation for the inner product is $\langle\varphi|\psi\rangle := (|\varphi\rangle, |\psi\rangle)$.

Definition 2.5 (Hilbert space)**.** *A Hilbert space is a linear space equipped with an inner product, which is complete with respect to the norm induced by the inner product.*

It can be shown that every finite-dimensional inner product space is a Hilbert space. From now on in this section we consider the case $\dim V < \infty$.

The dual of the vector $|\varphi\rangle$ is denoted as $\langle\varphi|$, and we call it a *bra*. This dual is a linear functional $\langle\varphi| : V \to \mathbb{C}$ which maps each $|\psi\rangle \in V$ to $\langle\varphi|\psi\rangle \in \mathbb{C}$.

A^T is the transpose of the matrix A, A^* is the entry-wise complex conjugate and $A^\dagger = (A^T)^*$ is the adjoint or Hermitian conjugate.

Besides vectors, the other most important objects are linear operators. The map $A : V \to W$ is a linear operator from the vector space V to another vector space W if

$$A\left(\sum_i z_i |\varphi_i\rangle\right) = \sum_i z_i A\left(|\varphi_i\rangle\right) \text{ for all } z_1, \ldots, z_m \in \mathbb{C} \text{ and } |\varphi_1\rangle, \ldots, |\varphi_m\rangle \in V.$$

Instead of $A(|\varphi\rangle)$ we often use the shorter notation $A|\varphi\rangle$. Some frequently used linear operators are the identity operator I, which maps all vectors to themselves, and the zero operator 0, which maps all vectors to the zero vector.

For the composition of two linear operators we use the notation $(BA)(|\varphi\rangle) := B(A|\varphi\rangle)$, i.e. the operators act on the vectors from right to left.

If the dimensions of the vector spaces V and W are $\dim V = n, \dim W = m$, then the linear operators $A : V \to W$ can be represented by complex $m \times n$ matrices if we fix a basis $|v_1\rangle, \ldots, |v_n\rangle$ in V, a basis $|w_1\rangle, \ldots, |w_m\rangle$ in W and we define the matrix $A \in \mathbb{C}^{m\times n}$ so that

$$A|v_j\rangle = \sum_{i=1}^{m} A_{ij}|w_i\rangle .$$

The Pauli matrices are four 2×2 complex matrices, which frequently occur in quantum mechanics, especially in quantum computing, where they represent the Pauli gates in quantum circuits. These four matrices are:

$$I = \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, X = \sigma_1 = \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, Y = \sigma_2 = \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, Z = \sigma_3 = \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The Pauli matrices form a basis of the 2×2 Hermitian ($A = A^\dagger$) matrices over the real numbers, they also form a basis of all 2×2 complex matrices over the complex numbers, they are unitary ($U^* = U^{-1}$) and have trace 0 (except for σ_0).

Definition 2.6 (Orthogonality)**.** *The vectors $|\varphi\rangle$ and $|\psi\rangle$ are orthogonal if $\langle\varphi|\psi\rangle = 0$.*

The norm of a vector $|\varphi\rangle$ is $||\,|\varphi\rangle\,|| := \sqrt{\langle\varphi|\varphi\rangle}$. Unit vectors are vectors with norm 1. A set of vectors $\{\varphi_i\}_{i\in I}$ is orthonormal if

$$\langle\varphi_i|\varphi_j\rangle = \delta_{ij} = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ otherwise .} \end{cases}$$

With the Gram–Schmidt procedure the existence of an orthonormal basis in every finite-dimensional vector space can be shown. In our calculations we always choose orthonormal bases for the matrix representation of the linear operators and if the operator maps the vector space to itself, we use

the same orthonormal basis for the domain and the codomain. This convention makes defining the inner product of two vectors more convenient. Indeed, if $|\varphi_i\rangle_{i\in I}$ is an orthonormal basis in the vector space V and for the vectors $|\psi\rangle, |\xi\rangle \in V$ we have $|\psi\rangle = \sum_{i\in I} \psi_i |\varphi_i\rangle$ and $|\xi\rangle = \sum_{j\in I} \xi_j |\varphi_j\rangle$, then

$$\langle\psi|\xi\rangle = \left(\sum_{i\in I} \psi_i |\varphi_i\rangle, \sum_{j\in I} \xi_j |\varphi_j\rangle\right) = \sum_{i,j\in I} \psi_i^* \xi_j \langle\varphi_i|\varphi_j\rangle = \sum_{i,j\in I} \psi_i^* \xi_j \delta_{ij} = \sum_{i\in I} \psi_i^* \xi_i.$$

So the vectors $|\varphi\rangle$ of the vector space V can be represented as row vectors and the duals $\langle\varphi|$ of the vectors as row vectors with coordinates conjugate to the vector $|\varphi\rangle$.

The bra-ket notation also makes it possible to define the outer product of vectors in an insightful way.

Definition 2.7 (Outer product of vectors)**.** *If $|\varphi\rangle \in V, |\psi\rangle \in W$, then let $|\psi\rangle\langle\varphi| : V \to W$ be the linear operator defined as*

$$(|\psi\rangle\langle\varphi|)(|\xi\rangle) := |\psi\rangle\langle\varphi|\xi\rangle = \langle\varphi|\xi\rangle \cdot |\psi\rangle.$$

The linear combination of such operators is defined similarly, i.e. $\sum_i z_i |\psi_i\rangle\langle\varphi_i|$ is a linear operator, which maps the vector $|\xi\rangle$ to $\sum_i z_i |\psi_i\rangle\langle\varphi_i|\xi\rangle$.

The outer product notation can be used to write the identity operator in a different form:

Theorem 2.8 (Completeness relation for orthonormal vectors)**.** *If the vector space V has an orthonormal basis $(|i\rangle)_{i\in I}$, then $\sum_i |i\rangle\langle i| = I_V$.*

Proof. $(|i\rangle)_{i\in I}$ is an orthonormal basis, hence every vector $|\varphi\rangle \in V$ can be written as $|\varphi\rangle = \sum_i \varphi_i |i\rangle$. We have seen before, that for orthonormal bases $\langle\varphi|\psi\rangle = \sum_i \varphi_i^* \psi_i$, hence $\varphi_i = \langle i|\varphi\rangle$, so

$$\left(\sum_i |i\rangle\langle i|\right)|\varphi\rangle = \sum_i |i\rangle\langle i|\varphi\rangle = \sum_i \varphi_i |i\rangle = |\varphi\rangle$$

□

We can represent an arbitrary linear operator $A : V \to W$ using the completeness relation. Supposing $(|\varphi_i\rangle)_{i\in I}$ is an orthonormal basis in V and $(|\psi_j\rangle)_{j\in J}$ is an orthonormal basis in W, we have $I_V = \sum_i |\varphi_i\rangle\langle\varphi_i|$ and $I_W = \sum_j |\psi_j\rangle\langle\psi_j|$.

$$A = I_W A I_V = \sum_{i\in I, j\in J} |\psi_j\rangle\langle\psi_j|A|\varphi_i\rangle\langle\varphi_i| = \sum_{i\in I, j\in J} \langle\psi_j|A|\varphi_i\rangle |\psi_j\rangle\langle\varphi_i|.$$

Definition 2.9 (Eigenvector, eigenvalue)**.** *The vector $|v\rangle$ is an eigenvector of the linear operator $A : V \to V$ with eigenvalue $\lambda \in \mathbb{C}$ if $A|v\rangle = \lambda|v\rangle$.*

Definition 2.10 (Diagonal representation)**.** *A linear operator A acting on V is diagonalizable if there exists an orthonormal set $(|i\rangle)_{i\in I}$ of its eigenvectors such that $A = \sum_i \lambda_i |i\rangle\langle i|$.*

If A is a bounded linear operator on a Hilbert space $\mathscr{H}$ (in this finite-dimensional case every linear operator is bounded), the Riesz representation theorem states that there exists a unique bounded linear operator denoted as $A^\dagger$ such that $(|\varphi\rangle, A|\psi\rangle) = (A^\dagger|\varphi\rangle, |\psi\rangle)$ for all $|\varphi\rangle, |\psi\rangle \in V$. We call the operator $A^\dagger$ the adjoint or the Hermitian conjugate of A.

Definition 2.11 (Hermitian operator)**.** *An operator A is Hermitian or self-adjoint if $A^\dagger = A$.*

Theorem 2.12. *All eigenvalues of a Hermitian operator are real.*

Proof. If $A = A^\dagger$ is a Hermitian operator with an eigenvector $|v\rangle$ and a corresponding eigenvalue λ, then by the properties of the complex inner product we have

$\lambda\langle v|v\rangle = \langle v|\lambda v\rangle = \langle v|Av\rangle = \langle A^\dagger v|v\rangle = \langle Av|v\rangle = \langle \lambda v|v\rangle = \overline{\lambda}\langle v|v\rangle$, and the eigenvector $|v\rangle$ is nonzero, hence by the positive definiteness of the inner product we have $\langle v|v\rangle > 0$, therefore $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$. □

Definition 2.13 (Projector)**.** *Let V be a vector space with a subspace W, $\dim V = n$, with an orthonormal basis $|1\rangle, \ldots, |n\rangle$ and $\dim W = k$ with the basis $|1\rangle, \ldots, |k\rangle$, which is obviously orthonormal as well.*

$P := \sum_{i=1}^{k} |i\rangle\langle i|$ is called the projector operator into the subspace W.

It is important to emphasize that this definition is independent of the choice of the orthonormal basis. It can be shown that P is Hermitian and its orthogonal complement $Q := I - P$ is a projector onto the subspace of V spanned by $|k+1\rangle, \ldots, |n\rangle$. Now we define some operator classes and look through their most important properties.

Definition 2.14 (Normal operator)**.** *An operator A is normal if it commutes with its Hermitian conjugate, i.e. $AA^\dagger = A^\dagger A$.*

One of the most frequently used results in operator theory is the following.

Theorem 2.15 (Spectral decomposition theorem)**.** *An operator $A : V \to V$ is diagonal with respect to an orthonormal basis of V if and only if it is normal.*

It can also be shown that a normal matrix is Hermitian if and only if all of its eigenvalues are real. Using the spectral decomposition theorem we can create operator functions from a given function $f : \mathbb{C} \to \mathbb{C}$, for example we can take the n-th root or the exponential of a normal operator. If A is a normal operator with $A = \sum_{i \in I} \lambda_i |\varphi_i\rangle \langle\varphi_i|$ spectral decomposition, we define $f(A)$ as $f(A) := \sum_{i \in I} f(\lambda_i) |\varphi_i\rangle \langle\varphi_i|$.

Theorem 2.16 (Pauli matrix exponentials)**.** *If $\vec{v} \in \mathbb{R}^3$, $||\vec{v}|| = 1$ and $t \in \mathbb{R}$, then $exp(it\vec{v} \cdot \vec{\sigma}) = \cos(t)I + i\sin(t)\vec{v} \cdot \vec{\sigma}$, where $\vec{v} \cdot \vec{\sigma} = \sum_{j=1}^{3} v_j \sigma_j$.*

Definition 2.17 (Unitary operator)**.** *An operator U acting on a finite dimensional vector space V is unitary if $UU^\dagger = I_V$ or equivalently $U^\dagger U = I_V$.*

Theorem 2.18 (Properties of unitary operators [19])**.** *Let $\mathscr{H}$ be a finite dimensional Hilbert space and $U : \mathscr{H} \to \mathscr{H}$ be a bounded linear operator. The following properties are equivalent:*

(i) U is an isometry: $||Ux|| = ||x||$ for all $x \in \mathscr{H}$,

(ii) U preserves the inner product: $\langle Ux|Uy\rangle = \langle x|y\rangle$ for all $x, y \in \mathscr{H}$,

(iii) U is unitary.

This result implies that if $(|\varphi_i\rangle)_{i \in I}$ is an orthonormal basis and U is unitary, then the set of the vectors$|\psi_i\rangle := U|\varphi_i\rangle$ is also orthonormal and $U = \sum_{i \in I} |\psi_i\rangle \langle\varphi_i|$. It can also be shown that the operator defined with the two orthonormal bases $(|\varphi_i\rangle)_{i \in I}$ and $(|\psi_i\rangle)_{i \in I}$ as $A := \sum_{i \in I} |\psi_i\rangle \langle\varphi_i|$ is unitary.

Definition 2.19 (Positive operator)**.** *An operator $A : V \to V$ is positive if $(|\varphi\rangle, A|\varphi\rangle) \geq 0$ for all $|\varphi\rangle \in V$. If the stricter condition $(|\varphi\rangle, A|\varphi\rangle) > 0$ holds for all $|\varphi\rangle \neq \mathbf{0}$, then the operator is positive definite.*

Theorem 2.20 (Square root of positive operator [19]). *Let $\mathscr{H}$ be a complex Hilbert space and let $A \in \mathscr{B}(\mathscr{H})$ be a positive operator. There exists a unique positive operator $B \in \mathscr{B}(\mathscr{H})$ such that $B^2 = A$.*

Definition 2.21. *We call the positive operator B the positive square root of the positive operator A if $B^2 = A$. The positive square root is denoted as $A^{1/2}$ or $\sqrt{A}$.*

We continue by defining the tensor product of Hilbert spaces, which is an essential tool used in the description of composite quantum systems. If V and W are Hilbert spaces with $\dim V = n, \dim W = m$, then their tensor product space is an mn-dimensional Hilbert space denoted as $V \otimes W$. The elements of the tensor product space are linear combinations of elements of the form $|\varphi\rangle \otimes |\psi\rangle$, where $|\varphi\rangle \in V, |\psi\rangle \in W$. If the space V has an orthonormal basis $(|i\rangle)_{i=1}^n$ and the space W has an orthonormal basis$(|j\rangle)_{j=1}^m$, then by definition the space $V \otimes W$ has an orthonormal basis $(|i\rangle)_{i=1}^n \otimes |j\rangle)_{j=1}^m$.

The tensor product has to satisfy the following properties:

- $z(|\psi\rangle \otimes |\xi\rangle) = (z|\psi\rangle) \otimes |\xi\rangle = |\psi\rangle \otimes (z|\xi\rangle)$ for $z \in \mathbb{C}, |\psi\rangle \in V, |\xi\rangle \in W$,
- $(|\psi_1\rangle + |\psi_2\rangle) \otimes |\xi\rangle = |\psi_1\rangle \otimes |\xi\rangle + |\psi_2\rangle \otimes |\xi\rangle$ for $|\psi_1\rangle, |\psi_2\rangle \in V, |\xi\rangle \in W$,
- $|\psi\rangle \otimes (|\xi_1\rangle + |\xi_2\rangle) = |\psi\rangle \otimes |\xi_1\rangle + |\psi\rangle \otimes |\xi_2\rangle$ for $|\psi\rangle \in V, |\xi_1\rangle, |\xi_2\rangle \in W$.

Linear operators on the tensor product space can be defined in the following way. Let $A : V \to V$ and $B : W \to W$ be linear operators, then let $A \otimes B : V \otimes W \to V \otimes W$ be an operator defined as $(A \otimes B)(|\psi\rangle \otimes |\xi\rangle) := A|\psi\rangle \otimes B|\xi\rangle$ for $|\psi\rangle \in V, |\xi\rangle \in W$. If we extend this definition for all elements of $V \otimes W$ linearly, we get the linear operator $A \otimes B$.

With the help of the inner product on V and W, we can also define an inner product on $V \otimes W$ as:

$$\left(\sum_i z_i |\varphi_i\rangle \otimes |\xi_i\rangle, \sum_j w_j |\psi_j\rangle \otimes |\zeta_j\rangle \right) := \sum_{i,j} z_i^* w_j \langle \varphi_i | \psi_j \rangle \langle \xi_i | \zeta_j \rangle .$$

We represent the tensor product in a fixed basis by the Kronecker product.

The Kronecker product of the $A \in \mathbb{C}^{i\times j}$ and $B \in \mathbb{C}^{k\times l}$ matrices is a $\mathbb{C}^{ik\times jl}$ matrix

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \dots & a_{1,j}B \\ \vdots & \ddots & \vdots \\ a_{i,1}B & \dots & a_{i,j}B \end{pmatrix}.$$

For example

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \otimes \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1 \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} & 2 \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \\ 3 \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} & 4 \cdot \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 5 & 6 & 10 & 12 \\ 7 & 8 & 14 & 16 \\ 15 & 18 & 20 & 24 \\ 21 & 24 & 28 & 32 \end{pmatrix}.$$

The Kronecker product is bilinear and associative, however it is not commutative.

Claim 1 (Mixed product property)**.** If the AC and BD matrix products exist, then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

The mixed product property has the important consequence that for $A \in \mathbb{C}^{i\times j}$ and $B \in \mathbb{C}^{k\times l}$ the equality $A \otimes B = (I_i \otimes B)(A \otimes I_l) = (A \otimes I_k)(I_j \otimes B)$ holds.

Definition 2.22 (Trace of matrix)**.** *The trace of a matrix A is $tr(A) = \sum_i A_{ii}$.*

It can be easily shown that the trace has the cyclic property, i.e. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, and it is complex linear, i.e. $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$, $\mathrm{tr}(zA) = z \cdot \mathrm{tr}(A)$.

Two matrices A and B are similar if there exists an invertible matrix P such that $B = PAP^{-1}$, which means that A and B represent the same linear map in two different bases. The cyclic property of the trace implies that $\mathrm{tr}(B) = \mathrm{tr}(PAP^{-1}) = \mathrm{tr}(AP^{-1}P) = \mathrm{tr}(A)$, therefore the trace of an operator is well-defined as the trace of one of its matrix representations.

For operators in infinite-dimensional spaces we need another definition of the trace.

Definition 2.23 (Separable Hilbert space)**.** *The Hilbert space $\mathscr{H}$ is separable if it contains a countable dense subset. This condition can be shown to be equivalent with containing a countable orthonormal basis.*

Definition 2.24 (Trace of operator)**.** *If $\mathscr{H}$ is a separable Hilbert space with an orthonormal basis $(|\varphi_k\rangle)_{k=1}^{\infty}$ and $A : \mathscr{H} \to \mathscr{H}$ is a positive bounded linear operator, then the trace of A is defined as $tr(A) = \sum_{k=1}^{\infty}(A|\varphi_k\rangle, |\varphi_k\rangle) = \sum_{k=1}^{\infty}\langle\varphi_k|A|\varphi_k\rangle$.*

As in the finite-dimensional case, the trace is also independent of the choice of basis in the countably infinite-dimensional case. It is well-known that for an arbitrary linear operator A the operator $|A| := \sqrt{A^{\dagger}A}$ is a positive semidefinite operator.

Definition 2.25 (Trace-class operator)**.** *A bounded linear operator $A : \mathscr{H} \to \mathscr{H}$ is trace-class if $tr(|A|) = tr(\sqrt{A^{\dagger}A}) < \infty$.*

The trace-norm of a trace-class operator A is $||A||_1 := tr(|A|)$. The set of trace-class operators on $\mathscr{H}$ is denoted as $\mathscr{T}_1(\mathscr{H})$.

Definition 2.26 (Hilbert–Schmidt inner product)**.** *Let $\mathscr{H}$ be a Hilbert space. The set of linear operators $L_{\mathscr{H}} = \{A : \mathscr{H} \to \mathscr{H},\ A \text{ is linear}\}$ forms a complex vector space. This vector space can be equipped with the inner product $(A,B) := tr(A^{\dagger}B)$, under which it becomes a Hilbert space. This inner product is called the Hilbert–Schmidt or trace inner product.*

Definition 2.27 (Trace norm)**.** *The trace norm of a matrix A is defined as $||A||_1 = tr(\sqrt{A^{\dagger}A})$.*

2.4 Foundation of quantum mechanics

After summarizing the most important concepts of linear algebra, let us continue with an overview of the fundamental principles and definitions of quantum mechanics.

Quantum mechanics is a mathematical framework used to describe physical systems where classical mechanics fails, such systems on the microscopic scale. A complete introduction to the quantum mechanical formalism would far exceed the scope of this thesis, so we will focus on presenting the basic principles and the concepts that will be applied later.

Postulate 2.28 (1. State space)**.** To any isolated physical system (it may not exchange matter or energy with its surroundings) we can associate a complex Hilbert space called the state space of the system. The state of the system is described by a state vector, a unit vector in the state space.

The simplest quantum mechanical system is the qubit, which is a system with a two-dimensional Hilbert space. The elements of the standard orthonormal basis called the computational basis are usually denoted as $|0\rangle = (1,0)$ and $|1\rangle = (0,1)$. The state vectors of the phase space are of the form $|\psi\rangle = z_0|0\rangle + z_1|1\rangle$, where $z_1, z_2 \in \mathbb{C}$ and $|\psi\rangle$ is a unit vector, i.e. $|z_1|^2 + |z_2|^2 = 1$.

In general, we call $\sum_i z_i |\psi_i\rangle$ the superposition of the states $|\psi_i\rangle$, where the state $|\psi_i\rangle$ has amplitude z_i.

Postulate 2.29 (2. Evolution)**.** The time evolution of an isolated quantum system can be described by a unitary operator U, i.e. if $|\psi_t\rangle$ denotes the state of the system at time t, then $|\psi_{t_2}\rangle = U|\psi_{t_1}\rangle$.

As we have already mentioned before, the Pauli matrices are unitary. The Pauli matrix X is also called the bit flip matrix, because it maps $|0\rangle$ to $|1\rangle$ and vice versa:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The Pauli matrix Z is usually referred to as the phase flip channel, because it fixes the $|0\rangle$ state and maps $|1\rangle$ to $-|1\rangle$.

The second postulate has another form, which describes the evolution of the system in continuous time with a differential equation called the Schrödinger equation:

$$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle,$$

where $\hbar$ is a constant called the reduced Planck constant, $|\psi\rangle$ is the state of the system and H is a Hermitian operator called the Hamiltonian of the system.

The operator H is Hermitian and hence normal, so it has a spectral decomposition $H = \sum_E E|E\rangle\langle E|$, where the vectors $|E\rangle$ are normalized. We call the states $|E\rangle$ the energy eigenstates or stationary states with energy E. The lowest energy that is the smallest eigenvalue of H is called the ground state energy and its corresponding state is the ground state.

We get a Cauchy-type differential equation, which we can easily solve formally, however computing the exact value of $\psi(t)$ often leads to difficult problems depending on the complexity of the Hamiltonian H.

$$\begin{cases} i\hbar\partial_t \psi(t) = H\psi(t) \\ \psi(0) = \psi_0 \end{cases}$$

The solution has the form $\psi(t) = e^{-i\frac{H}{\hbar}t} \cdot \psi_0$.

Now let us consider the Schrödinger equation of a free particle (the Hamiltonian does not contain a potential term, only a kinetic term) in one dimension, which has the form

$$i\hbar\partial_t \psi(t,x) = -\frac{\hbar^2}{2m}\partial_x^2 \psi(t,x).$$

We use the method of variable separation, so we are looking for a solution of the form $\psi(t,x) = f(t)g(x)$ and we assume that $\psi(x,t) \neq 0$.

The Schrödinger equation becomes

$$i\hbar f'(t)g(x) = -\frac{\hbar^2}{2m}g''(x)f(t).$$

If we divide both sides by $\psi(t,x)$, then we can observe that the left side only depends on the variable t and the right side only depends on the variable x, which means that they can be equal for all t and x values only if both sides are equal to a constant, which will be denoted as E.

$$i\hbar\frac{f'(t)}{f(t)} = -\frac{\hbar^2}{2m}\frac{g''(x)}{g(x)} := E.$$

We can solve the two ordinary differential equations separately. The first equation is

$$i\hbar\frac{f'(t)}{f(t)} = E \Leftrightarrow f'(t) = -\frac{iE}{\hbar}f(t),$$

which has the solution $f(t) = f(0) \cdot e^{-\frac{iE}{\hbar}t}$.

The second equation is

$$-\frac{\hbar^2}{2m}\frac{g''(x)}{g(x)} = E \Leftrightarrow g''(x) = -\frac{2mE}{\hbar^2}g(x),$$

which is a second order linear homogenous ordinary differential equation with constant coefficients.

Introducing the notation $k := \sqrt{\frac{2mE}{\hbar^2}}$, the solution is $g(x) = Ae^{ikx} + Be^{-ikx}$. (We can derive it from the characteristic equation $r^2 + k^2 = 0 \Rightarrow r = \pm ik$ and the fact that if the roots are $\alpha \pm i\beta$, then the solutions have the form $e^{\alpha x}(c_1\cos(\beta x) + c_2\sin(\beta x))$.)

We have obtained that the general form of the solution is

$$\psi(t,x) = e^{-\frac{i}{\hbar}Et}(Ae^{ikx} + Be^{-ikx}),$$

where the coefficients A, B can be determined from the boundary condition $\psi(0,x) = \psi_0(x)$.

We successfuly described the evolution of a particular **isolated** quantum system, however when we observe a quantum mechanical system, our measuring devices interact with the system, which causes changes in the initial system, so it is no longer an isolated system.

The third postulate defines measurement in quantum mechanical systems.

Postulate 2.30 (3. Quantum measurement)**.** A quantum measurement is a set of linear operators $\{M_m\}_{m\in J}$ called measurement operators. J is the set of possible outcomes of the measurement.

The probability of the outcome $m \in J$ in the measurement of a system in state $|\psi\rangle$ is given by

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle\,.$$

If the outcome is m, then the state of the system collapses into

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}} = \frac{M_m|\psi\rangle}{\sqrt{p(m)}}.$$

Of course we must require that the sum of the probabilities of the possible outcomes is 1, i.e.

$$\sum_{m\in J} p(m) = \sum_{m\in J} \langle\psi|M_m^\dagger M_m|\psi\rangle = 1.$$

Claim 2 (Completeness equation)**.** The condition $\sum_{m\in J} M_m^\dagger M_m = I$ called the completeness equation is equivalent to the condition that $\sum_{m\in J}\langle\psi|M_m^\dagger M_m|\psi\rangle = 1$ for all states $|\psi\rangle$.

Proof. If $\sum_{m\in J} M_m^\dagger M_m = I$, then for any state $|\psi\rangle$ we have
$\sum_{m\in J}\langle\psi|M_m^\dagger M_m|\psi\rangle = \big\langle\psi\big|\sum_{m\in J} M_m^\dagger M_m\big|\psi\big\rangle = \langle\psi|I|\psi\rangle = 1.$

On the other hand, $\sum_{m\in J} M_m^\dagger M_m$ is positive and hence normal, so it has a spectral decomposition. The condition $\sum_{m\in J}\langle\psi|M_m^\dagger M_m|\psi\rangle = 1$ for all states implies that all the eigenvalues of $\sum_{m\in J} M_m^\dagger M_m$ are 1, therefore $\sum_{m\in J} M_m^\dagger M_m = I$. □

To understand this concept let us consider a simple example, the measurement of a qubit in the computational basis. We have a state vector $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where $|\alpha|^2 + |\beta|^2 = 1$ and a quantum measurement consisting of two measurement operators,

$$M_0 = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } M_1 = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

$$M_0^\dagger M_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } M_1^\dagger M_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

so applying the above formulae for the measurement outcomes we get

$$p(0) = \begin{pmatrix} \overline{\alpha} & \overline{\beta} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = |\alpha|^2 \text{and} p(1) = \begin{pmatrix} \overline{\alpha} & \overline{\beta} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = |\beta|^2.$$

The state after the measurement becomes

$$\frac{M_0|\psi\rangle}{\sqrt{p(0)}} = \frac{1}{|\alpha|} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{\alpha}{|\alpha|}|0\rangle$$

if the outcome was 0 and

$$\frac{M_1|\psi\rangle}{\sqrt{p(1)}} = \frac{1}{|\beta|} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{\beta}{|\beta|}|1\rangle$$

if the outcome was 1.

An important question is whether certain quantum states are distinguishable with the help of quantum measurement. By distinctiveness of a set of quantum states we mean that there exists a quantum measurement such that if we measure an arbitrary state from our set, then we can determine the measured state from the outcome of the measurement with zero error probability.

When the given set of states $(|\psi_i\rangle)_{i\in J}$ is orthonormal, they are distinguishable with the measurement $M = \cup_{i\in J}\{M_i\} \cup \{M_0\}$, where $M_i = |\psi_i\rangle\langle\psi_i|$ for $i \in J$ and for the completeness equation $M_0 = I - \sum_{i\in J}|\psi_i\rangle\langle\psi_i|$. Indeed, for each $i \in J$ we have

$$p(i) = \langle\psi_i|M_i^\dagger M_i|\psi_i\rangle = \langle\psi_i|\left(|\psi_i\rangle\langle\psi_i||\psi_i\rangle\langle\psi_i|\right)|\psi_i\rangle = \langle\psi_i|\psi_i\rangle^3 = 1,$$

i.e. for the state $|\psi_i\rangle$ the outcome of the measurement is i with probability 1, so we can distinguish the states by applying this measurement.

However, it can be proved that two non-orthogonal quantum states cannot be distinguished with any measurement. Frequently used measurements are the so-called projective measurements.

Definition 2.31 (Projective measurement)**.** *A projective measurement (or von Neumann measurement) is a Hermitian operator A, which has a spectral decomposition* $A = \sum_{i\in J} \lambda_i P_i$*, where each* P_i *is a projector operator onto the eigenspace of the operator M with eigenvalue* λ_i*.*

Projective measurements are special cases of quantum measurements if we choose the set of measurement operators $\{M_m\}_{m\in J}$ so that each M_m is a projector and they are pairwise orthogonal, i.e. $M_{m_1}M_{m_2} = \delta m_1, m_2 M_{m_1}$.

The possible measurement outcomes are the eigenvalues of A and the probability of each outcome when measuring on a system being in the state $|\psi\rangle$ is given by
$p(\lambda_i) = \langle\psi|P_i^\dagger P_i|\psi\rangle = \langle\psi|P_i|\psi\rangle$, because by definition the projector operators P_i are Hermitian ($P_i = P_i^\dagger$) and $P_i^2 = P_i$.
If the measurement outcome is λ_i, the state collapses into $\frac{P_i|\psi\rangle}{\sqrt{p(i)}} = \frac{P_i|\psi\rangle}{\langle\psi|P_i|\psi\rangle}$.
The expected value $\mathbb{E}(A)$ of the projective measurement A on a system being in state $|\psi\rangle$ can be computed as

$$\langle A\rangle_\psi = \mathbb{E}_\psi(A) = \sum_{i\in J} \lambda_i p(\lambda_i) = \sum_{i\in J} \lambda_i \langle\psi|P_i|\psi\rangle = \langle\psi| \left(\sum_{i\in J} \lambda_i P_i\right) |\psi\rangle = \langle\psi|A|\psi\rangle .$$

A common notation for the expected value of a measurement is $\langle M\rangle := \mathbb{E}(M)$, and the standard deviation of the measurement outcomes is defined as $\Delta(M) := \sqrt{\langle M^2\rangle - \langle M\rangle^2}$, which corresponds to the standard deviation commonly used in probability theory.

In quantum mechanics we call self-adjoint operators **observables**. The possible outcomes of the measurement described by an observable are its eigenvalues, which are all real.

When investigating the statistics of a measurement, i.e. the probabilities of the different outcomes, it is useful to introduce some new concepts called global and relative phases.

Definition 2.32 (Global phase factor)**.** *The state* $|\psi\rangle$ *is equal to the state* $e^{i\vartheta}|\psi\rangle$ *up to the global phase factor, where* $\vartheta \in \mathbb{R}$*.*

States that are equal up to the global phase factor are physically indistinguishable, because for an arbitrary measurement the respective outcome probabilities are all equal:
$(e^{i\vartheta}|\psi\rangle)^\dagger = \langle\psi|e^{-i\vartheta}$ and $p_\psi(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle = \langle\psi|e^{-i\vartheta}M_m^\dagger M_m e^{i\vartheta}|\psi\rangle = p_{e^{i\vartheta}\psi}(m)$.

Definition 2.33 (Relative phase factor)**.** *Two states $|\psi\rangle$ and $|\xi\rangle$ in a given basis differ by a relative phase if for each basis element $|m\rangle$ there exists a $\vartheta_m \in \mathbb{R}$ such that for the corresponding coordinates ψ_m and ξ_m the equality $\psi_m = e^{i\vartheta_m}\xi_m$ holds.*

The next postulate tells us how we can describe composite quantum systems which consist of smaller systems.

Postulate 2.34 (4. Composite systems)**.** The Hilbert space of a composite quantum physical system is the tensor product of the Hilbert spaces of its component systems.

If $|x\rangle$ is a state in the system A and $|y\rangle$ is a state in the system B, then their tensor product $|x\rangle \otimes |y\rangle$ is a state in the composite system AB. However, not all elements can be expressed in such a form. To see this, let us consider the two qubit state space, which has the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ defined as
$|00\rangle = |0\rangle \otimes |0\rangle = (1,0,0,0)^T$, $|01\rangle = |0\rangle \otimes |1\rangle = (0,1,0,0)^T$,
$|10\rangle = |1\rangle \otimes |0\rangle = (0,0,1,0)^T$, $|11\rangle = |1\rangle \otimes |1\rangle = (0,0,0,1)^T$.
Now we show that the state $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ cannot be expressed as the tensor product of two single-qubit states.

Let us indirectly suppose that $|\psi\rangle = |x\rangle \otimes |y\rangle$, where $|x\rangle = \alpha|0\rangle + \beta|1\rangle, |y\rangle = \gamma|0\rangle + \delta|1\rangle$, $|\alpha|^2 + |\beta|^2 = 1, |\gamma|^2 + |\delta|^2 = 1$.

$$|x\rangle \otimes |y\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} \otimes \begin{pmatrix}\gamma\\ \delta\end{pmatrix} = \begin{pmatrix}\alpha\gamma\\ \alpha\delta\\ \beta\gamma\\ \beta\delta\end{pmatrix} = \begin{pmatrix}1/\sqrt{2}\\ 0\\ 0\\ 1/\sqrt{2}\end{pmatrix} = |\psi\rangle.$$

This means that $\alpha\gamma = \frac{1}{\sqrt{2}}, \beta\delta = \frac{1}{\sqrt{2}}, \alpha\delta = 0, \beta\gamma = 0$, which is a contradiction.

States of a composite system that cannot be expressed as the tensor product of states of the components are called entangled states. The phenomenon of the quantum entanglement plays a

fundamental role in several application areas of quantum information, for example in superdense coding and quantum teleportation [14, page 96].

So far we have summarized the most important concepts of quantum mechanics, and now we continue by describing an alternative approach using density operators instead of state vectors, which makes a wider range of applications possible. The main difference of this approach from the previous one is that now we do not know the current state of the observed system with certainty. Instead, we only have a set $\{p_i, |\psi_i\rangle\}_{i\in J}$ called an ensemble of quantum states. We interpret this ensemble as for each $i \in J$ the system is in the state $|\psi_i\rangle$ with probability p_i.

Definition 2.35 (Density operator)**.** *For a given ensemble of states $\{p_i, |\psi_i\rangle\}$ the density operator is defined as $\rho = \sum_{i\in J} p_i |\psi_i\rangle \langle\psi_i|$.*

With the help of the density operator we can describe the evolution of a quantum system. According to the second postulate the evolution is described by a unitary operator U, and if the system was in state $|\psi_i\rangle$ with probability p_i, then after the action of the evolution operator it will be in state $U|\psi_i\rangle$ with probability p_i. Therefore the evolution takes the operator $\rho = \sum_{i\in J} p_i |\psi_i\rangle \langle\psi_i|$ to $\sum_{i\in J} p_i U|\psi_i\rangle \langle\psi_i| U^\dagger = U\rho U^\dagger$.

Definition 2.36 (Pure and mixed states)**.** *A quantum system is in a pure state ρ if its density operator has rank 1, i.e. $\rho = |\psi\rangle\langle\psi|$. If the system is not in a pure state, we call its state a mixed state.*

The density operator ρ is said to be the mixture of states ρ_i with probabilities p_i if $\rho = \sum_i p_i \rho_i$. We defined the density operator with the help of an ensemble of state vectors and probabilities, but our aim is to provide a description of quantum mechanics which does not rely on state vectors at all. The following theorem is a characterisation of density operators, which points in this direction.

Theorem 2.37. *For an operator ρ there exists an ensemble $\{p_i, |\psi_i\rangle\}$ if and only if $tr(\rho) = 1$ and ρ is a positive operator.*

We can distinguish pure states from mixed states with the following observation.

Theorem 2.38. *For any density operator ρ we have $tr(\rho^2) \leq 1$ and equality holds if and only if ρ is a pure state.*

Proof. ρ is a positive operator, so it has a spectral decomposition $\rho = \sum_i p_i |i\rangle \langle i|$ in an orthonormal basis. The positivity of ρ implies that $p_i \geq 0$ for all i and we also know from the trace condition for density operators that $1 = \mathrm{tr}(\rho) = \mathrm{tr}\left(\sum_i p_i |i\rangle \langle i|\right) = \sum_i \mathrm{tr}(p_i |i\rangle \langle i|) = \sum_i p_i$. Now we know that $0 \leq p_i \leq 1$ for all i.

Using the orthonormality of the basis we get

$$\rho^2 = \rho \cdot \rho = \left(\sum_i p_i |i\rangle \langle i|\right) \cdot \left(\sum_j p_j |j\rangle \langle j|\right) = \sum_{i,j} p_i p_j |i\rangle \langle i|j\rangle \langle j| = \sum_{i,j} p_i p_j |i\rangle \langle j| \, \delta_{i,j} = \sum_i p_i^2 |i\rangle \langle i|,$$

so $\mathrm{tr}(\rho^2) = \mathrm{tr}\left(\sum_i p_i^2 |i\rangle \langle i|\right) = \sum_i p_i^2 \mathrm{tr}(|i\rangle \langle i|) = \sum_i p_i^2 \leq \sum_i p_i \leq 1$,
and equality holds if and only if $p_i^2 = p_i$ for all i, which means that all states have probability 1 or 0, therefore ρ is a pure state. □

The following theorem tells us about what sets of states generate the same density operator, which is important in quantum information.

Theorem 2.39 (Unitary freedom for density operators)**.** *The ensembles* $\{p_i, |\psi_i\rangle\}$ *and* $\{q_j, |\xi_j\rangle\}$ *generate the same density operator* ρ*, i.e.* $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i| = \sum_j q_j |\xi_j\rangle \langle \xi_j|$ *if and only if there exists a unitary matrix* $U = (u_{ij})$ *such that for all* i, j*, the equality* $\sqrt{p_i} |\psi_i\rangle = \sum_j u_{ij} \sqrt{q_j} |\xi_j\rangle$ *holds.*

The statement of the theorem remains valid even if the two sets of states have different cardinalities, as the smaller set can be extended by adding states with corresponding probabilities zero.

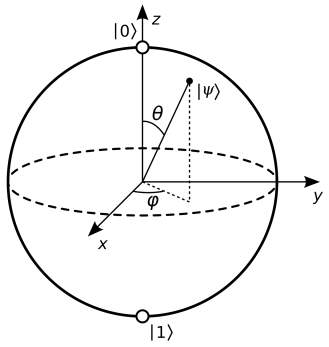
The introduction of the Bloch sphere provides a useful tool for a more precise understanding of the qubit state space. The Bloch sphere gives us a geometric interpretation of qubit states and graphically illustrates some quantum operations, for example certain quantum channels, which we will describe later in details.

A state vector in the qubit phase space has the form $|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$,
where $|\alpha|^2 + |\beta|^2 = 1$. From the normalizing condition we can rewrite it as
$|\psi\rangle = e^{i\gamma} \left(\cos\left(\frac{\vartheta}{2}\right) |0\rangle + e^{i\varphi} \sin\left(\frac{\vartheta}{2}\right) |1\rangle\right)$, where $\gamma, \vartheta, \varphi \in \mathbb{R}$.
States equal up to the global phase factor are indistinguishable, sowe can ignore the coefficient $e^{i\gamma}$

and describe the qubit states with only two real parameters ϑ and φ, as

$$|\psi\rangle = \cos\left(\frac{\vartheta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\vartheta}{2}\right)|1\rangle\,, \text{ where } \varphi \in [0,2\pi] \text{ and } \vartheta \in [0,\pi].$$

With this notation the quantum states can be visualized on the surface of the unit sphere in $\mathbb{R}^3$ called the Bloch sphere [2].



The Bloch sphere representation can be generalized for mixed states, too.

Theorem 2.40 (Bloch sphere representation for mixed states)**.** *If $\rho \in \mathbb{C}^2$ is a density matrix, then it can be written in the form $\rho = \frac{I+\vec{r}\cdot\vec{\sigma}}{2}$, where $\vec{r} \in \mathbb{R}^3$, $||\vec{r}|| \leq 1$, $\sigma_x, \sigma_y, \sigma_z$ are the Pauli matrices and $\vec{r}\cdot\vec{\sigma} = r_1\cdot\sigma_x + r_2\cdot\sigma_y + r_3\cdot\sigma_z$.*

Proof. We know that the four Pauli matrices form a basis of the 2×2 complex matrices over the complex numbers and hence any $\rho \in \mathbb{C}^{2\times 2}$ can be written in the form
$\rho = z_1 I + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z$ where $z_1, z_2, z_3, z_4 \in \mathbb{C}$.
Density operators are positive and hence Hermitian and Pauli matrices are also Hermitian, so $\rho^\dagger = \rho = \overline{z_1}I + \overline{z_2}\sigma_x + \overline{z_3}\sigma_y + \overline{z_4}\sigma_z$. The coefficients of a vector in a basis are unique, so $\overline{z_j} = z_j$ for all $j \in \{1,2,3,4\}$, i.e. the coefficients are real.
The trace of a density operator is 1 and the trace is linear, so
$1 = \mathrm{tr}(\rho) = \mathrm{tr}(z_1 I + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z) = z_1\mathrm{tr}(I) = 2z_1 \Rightarrow z_1 = \frac{1}{2}$.
Calculating ρ^2 using that $I^2 = \sigma_x^2 = \sigma_y^2 = \sigma_z^2 = I$, $\sigma_x\sigma_y = -\sigma_y\sigma_x, \sigma_x\sigma_z = -\sigma_z\sigma_x, \sigma_y\sigma_z = -\sigma_z\sigma_y$,

we get

$$\begin{aligned}\rho^2 &= \left(\frac{1}{2}I + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z\right)^2 \\ &= \frac{1}{4}I^2 + z_2^2\sigma_x^2 + z_3^2\sigma_y^2 + z_4\sigma_z^2 + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z + \\ &+ z_2z_3\sigma_x\sigma_y + z_2z_3\sigma_y\sigma_x + z_2z_4\sigma_x\sigma_z + z_2z_4\sigma_z\sigma_x + z_3z_4\sigma_y\sigma_z + z_3z_4\sigma_z\sigma_y \\ &= \left(\frac{1}{4} + z_2^2 + z_3^2 + z_4^2\right)I + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z.\end{aligned}$$

Now we can calculate the trace of the operator ρ^2 and use the previously proven fact that $(\rho^2) \leq 1$ for any density operator ρ:

$\mathrm{tr}(\rho^2) = 2 \cdot \left(\frac{1}{4} + z_2^2 + z_3^2 + z_4^2\right) \leq 1 \Leftrightarrow z_2^2 + z_3^2 + z_4^2 \leq \frac{1}{4} \Leftrightarrow \sqrt{z_2^2 + z_3^2 + z_4^2} \leq \frac{1}{2}$.

So $\rho = \frac{1}{2}\left(I + z_2\sigma_x + z_3\sigma_y + z_4\sigma_z\right)$, where

$$||\vec{r}|| = \sqrt{(2z_2)^2 + (2z_3)^2 + (2z_4)^2} = 2\sqrt{z_2^2 + z_3^2 + z_4^2} \leq 1$$

. □

It is also not difficult to see that ρ is a pure state if and only if $||\vec{r}|| = 1$.

Investigating composite quantum systems often requires describing the state of certain subsystems, which makes it necessary to define the so-called reduced density operator. For this definition we also need the definition of the partial trace of an operator.

Definition 2.41 (Partial trace [9])**.** *Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be two Hilbert spaces with orthonormal bases $(|\psi_i\rangle)_i \subset \mathscr{H}_1$ and $(|\xi_j\rangle)_j \subset \mathscr{H}_2$. Then $(|\psi_i, \xi_j\rangle)_{i,j} = (|\psi_i\rangle \otimes |\xi_j\rangle)_{i,j}$ is an orthonormal basis in $\mathscr{H}_1 \otimes \mathscr{H}_2$. For a linear operator $A = \sum_{i,j,k,l} A_{i,j,k,l} |\psi_i, \xi_j\rangle \langle\psi_k, \xi_l|$ on $\mathscr{H}_1 \otimes \mathscr{H}_2$ the partial trace of A over $\mathscr{H}_1$ is a linear operator on $\mathscr{H}_2$ defined as*

$$tr_{\mathscr{H}_1}[A] = \sum_{j,l}\left(\sum_i A_{i,j,i,l}\right)|\xi_j\rangle\langle\xi_l|,$$

and the partial trace of A over $\mathscr{H}_2$ is a linear operator on $\mathscr{H}_1$ defined as

$$tr_{\mathscr{H}_2}[A] = \sum_{i,k}\left(\sum_j A_{i,j,k,j}\right)|\psi_i\rangle\langle\psi_k|.$$

It can be proved that the partial traces do not depend on the choice of the bases.

Claim 3. For any $X \in \mathscr{L}(\mathscr{H}_1), Y \in \mathscr{L}(\mathscr{H}_2), A \in \mathscr{L}(\mathscr{H}_1 \otimes \mathscr{H}_2)$:

(i) $\mathrm{tr}_{\mathscr{H}_1 \otimes \mathscr{H}_2}\left[\left(X \otimes I_{\mathscr{H}_2}\right) A\right] = \mathrm{tr}_{\mathscr{H}_1}\left[X \cdot \mathrm{tr}_{\mathscr{H}_2}[A]\right].$

(ii) $\mathrm{tr}_{\mathscr{H}_1 \otimes \mathscr{H}_2}\left[\left(I_{\mathscr{H}_1} \otimes Y\right) A\right] = \mathrm{tr}_{\mathscr{H}_2}\left[Y \cdot \mathrm{tr}_{\mathscr{H}_1}[A]\right].$

Proof. We only prove (i), the case (ii) is similar.

$$\begin{aligned}
&\mathrm{tr}_{\mathscr{H}_1 \otimes \mathscr{H}_2}\left[\left(X \otimes I_{\mathscr{H}_2}\right) A\right] \\
&= \sum_{i,j} \langle \psi_i \otimes \xi_j | (X \otimes I_{\mathscr{H}_2}) A | \psi_i \otimes \xi_j \rangle = \sum_{i,j} \langle \psi_i | X \cdot \langle \xi_j | A | \xi_j \rangle | \psi_i \rangle \\
&= \sum_i \langle \psi_i | X \sum_j \langle \xi_j | A | \xi_j \rangle | \psi_i \rangle = \sum_i \langle \psi_i | X \cdot \mathrm{tr}_{\mathscr{H}_2}[A] | \psi_i \rangle = \mathrm{tr}_{\mathscr{H}_1}[X \cdot \mathrm{tr}_{\mathscr{H}_2}[A]].
\end{aligned}$$

□

Definition 2.42 (Reduced density operator [15])**.** *If we have two systems described by the Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ and the density operator of the composite system is ρ_{12}, then the reduced density operator for system $\mathscr{H}_1$ is $\rho_1 := tr_{\mathscr{H}_2}[\rho_{12}]$ and the reduced density operator for system $\mathscr{H}_2$ is $\rho_2 := tr_{\mathscr{H}_1}[\rho_{12}]$.*

To understand these definitions, let us take an example. One of the Bell states on the two-qubit system is defined as $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, and hence its density operator is

$$\rho = |\Phi^+\rangle\langle\Phi^+| = \frac{1}{2}(|00\rangle + |11\rangle)(\langle 00| + \langle 11|) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

If ρ_1 is the reduced density operator for the first qubit, then we have that $A_{0,0,0,0} = A_{1,1,1,1} = A_{0,0,1,1} = A_{1,1,0,0} = \frac{1}{2}$ and the other coordinates are all zero, hence

$$\rho_1 = \mathrm{tr}_2(\rho) = \sum_{i,k}\sum_{j} A_{i,j,k,j} \left|\psi_i\right\rangle \left\langle\psi_k\right| = A_{0,0,0,0} \left|0\right\rangle \left\langle 0\right| + A_{1,1,1,1} \left|1\right\rangle \left\langle 1\right|$$

$$= \frac{1}{2}(\left|0\right\rangle \left\langle 0\right| + \left|1\right\rangle \left\langle 1\right|) = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In the case of an n-particle quantum system, which is described by the tensor product $\otimes_{i=1}^{n} \mathscr{H}_i$ of n Hilbert spaces we can also take the partial trace corresponding to the i-th system as $\mathrm{tr}_{\mathscr{H}_i}[A] \in \mathscr{L}(\otimes_{j\neq i} \mathscr{H}_j)$, and for any $I \subseteq \{1,\ldots,n\}$ subset of indices the corresponding reduced density operator is $\rho_I = \mathrm{tr}_{\{1,\ldots,n\}\setminus I}[\rho]$.

The most frequently used example is the n-qubit system, which is the tensor product of the qubit Hilbert spaces $\mathscr{H}_i = \mathbb{C}^2$. Each of the subsystems has the previously defined basis $\{\left|0\right\rangle, \left|1\right\rangle\}$ and the computational basis of the n-qubit system is defined as $\{\left|x\right\rangle\}_{x\in\{0,1\}^n} \subseteq (\mathbb{C}^2)^{\otimes n}$.

3 Quantum mechanical approach to optimal transport

3.1 Different approaches to the transport problem

Now that we have introduced the basic concepts of both optimal transport theory and quantum mechanics, we can continue by connecting these two areas. For making it easier to notice the similarities between the classical and the quantum formalism, we make a table which compares these two. In quantum mechanical context $\mathscr{L}(\mathscr{H})$ denotes the set of linear operators $A : \mathscr{H} \to \mathscr{H}$, $\mathscr{O}(\mathscr{H})$ stands for the set of observables (linear self-adjoint operators), $\mathscr{S}(\mathscr{H})$ denotes the set of density operators (self-adjoint, positive, unit trace linear operators) and $\mathscr{P}_1(\mathscr{H})$ stands for the set of pure states $\rho = |\psi\rangle\langle\psi|$.

concept	classical notation	quantum notation
underlying set	X, Y	Hilbert space $\mathscr{H}$
elements of the set	x, y	$\|\varphi\rangle \in \mathscr{H}$
complex-valued functions	$f : X \to \mathbb{C}$	$A \in \mathscr{L}(\mathscr{H})$ linear operator
adjoint	f^*	A^*
absolute value	$\|f\|^2$	A^*A
real-valued functions	$f : X \to \mathbb{R}$	$A \in \mathscr{O}(\mathscr{H})$ observable
non-negative functions	$f : X \to [0,\infty)$	$A \in \mathscr{O}(\mathscr{H}), A \geq 0$ positive observable
expectation	$\int_X f(x)\, d\mu(x)$	$\mathrm{tr}[\rho A]$
state of the system	p probability measure on set X	$\rho \in \mathscr{S}(\mathscr{H})$ density operator (state)
extremal states	Dirac measure:$\delta_y : X \to \mathbb{R}$, $\delta_y(x) = \delta_{xy} = \mathbf{1}_{x=y}$	pure state: $\|\varphi\rangle\langle\varphi\|$, where $\langle\varphi\|\varphi\rangle = 1$
composite systems	Cartesian product $X \times Y$	tensor product $\mathscr{H}_1 \otimes \mathscr{H}_2$

The following two statements illustrate the analogy between marginal distributions in classical optimal transport and partial traces in quantum optimal transport.

Claim 4. A probability measure $\pi \in \text{Prob}(X \times Y)$ has first marginal $(\pi)_1 = \mu$ and second marginal $(\pi)_2 = \nu$ if and only if

$$\iint_{X\times Y} f(x)d\pi(x,y) = \int_X f(x)d\mu(x)$$

and

$$\iint_{X\times Y} g(y)d\pi(x,y) = \int_Y g(y)d\nu(y)$$

for all $f \in \mathscr{C}_b(X), g \in \mathscr{C}_b(Y)$ continuous, bounded functions.

Claim 5. A linear operator $A \in \mathscr{L}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ has partial traces $\text{tr}_{\mathscr{H}_2}[A] = A_1$ and $\text{tr}_{\mathscr{H}_1}[A] = A_2$ if and only if

$$\text{tr}_{\mathscr{H}_1\otimes\mathscr{H}_2}[(X \otimes I_{\mathscr{H}_2})A] = \text{tr}_{\mathscr{H}_1}[X \cdot A_1]$$

and

$$\text{tr}_{\mathscr{H}_1\otimes\mathscr{H}_2}[(I_{\mathscr{H}_1} \otimes Y)A] = \text{tr}_{\mathscr{H}_2}[Y \cdot A_2]$$

for all linear operators $X \in \mathscr{L}(\mathscr{H}_1), Y \in \mathscr{L}(\mathscr{H}_2)$.

Various distances can be defined between probability distributions, many of which also have quantum analogues. For example, let us consider the total variation distance defined on a countable probability space X between probability measures μ and ν as

$$||\mu - \nu||_{TV} := \sup_{A\subset X} |\mu(A) - \nu(A)|.$$

Theorem 3.1 ([17])**.** *If X is a finite or countable probability space, then*
$||\mu - \nu||_{TV} := \sup_{A\subset X} |\mu(A) - \nu(A)| = \frac{1}{2}\sum_{x\in X} |\mu(x) - \nu(x)|$.

Proof. $B := \{x \in X : \mu(x) \geq \nu(x)\}$. For an arbitrary $A \subset X$ we have $\mu(A) = \mu(A\cap B) + \mu(A\backslash B)$ and $\nu(A) = \nu(A\cap B) + \nu(A\backslash B)$ and from the definition of the set B we know that $\mu(A\backslash B) \leq \nu(A\backslash B)$, hence $\mu(A) - \nu(A) \leq \mu(A\cap B) - \nu(A\cap B) \leq \mu(B) - \nu(B)$. This yields $\sup_{A\subset X} \mu(A) - \nu(A) =$

$\mu(B) - \nu(B)$. For symmetry reasons, we similarly obtain that
$\mu(A) - \nu(A) \geq \mu(A \cap B^c) - \nu(A \cap B^c) \geq \mu(B^c) - \nu(B^c)$ and hence $\sup_{A \subset X} \nu(A) - \mu(A) = \nu(B^c) - \mu(B^c)$.
Both measures are probability measures, so $\mu(B) + \mu(B^c) = \nu(B) + \nu(B^c) = 1$, so $\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c)$, therefore $\sup_{A \subset X} |\mu(A) - \nu(A)| = \mu(B) - \nu(B)$ and also $|\mu(B) - \nu(B)| = |\mu(B^c) - \nu(B^c)| = \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)|$. Putting these two equations together we get the statement of the theorem. □

In the case of the quantum analogues of the classical distances, we define the distance between quantum states on separable Hilbert spaces instead of between probability measures.

Theorem 3.2. $||A||_1 = tr(\sqrt{A^\dagger A}) = \sum_j |\lambda_j|$, *where* λ_j *are the eigenvalues (with multiplicity) of the matrix A.*

The quantum counterpart of the total variation distance is the trace distance defined as

$$D_{\mathrm{tr}}(\rho, \sigma) = \frac{1}{2} \mathrm{tr} \sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)} = \frac{1}{2} ||\rho - \sigma||_1.$$

Theorem 3.3 (Unitary invariance)**.** *The trace norm of a matrix (and hence the total variation distance) is unitarily invariant, i.e. for any unitary matrix U and arbitrary matrix A the equality* $||UAU^\dagger||_1 = ||A||_1$ *holds.*

Proof. $||A||_1 = \sum_j |\lambda_j|$, where λ_j are the eigenvalues of A, so it is sufficient to show that $UAU^\dagger$ and A have the same eigenvalues. Supposing $A|v\rangle = \lambda |v\rangle$ for a vector $|v\rangle \neq \vec{0}$, we have $(UAU^\dagger)(U|v\rangle) = UA|v\rangle = U\lambda |v\rangle = \lambda (U|v\rangle)$, so λ is also an eigenvalue of $UAU^\dagger$. Similarly, if $(UAU^\dagger)|v\rangle = \lambda |v\rangle$, then $AU^\dagger |v\rangle = U^\dagger \lambda |v\rangle = \lambda U^\dagger |v\rangle$, so A has eigenvalue λ. □

3.2 Quantum channels

The quantum analogues of the Markov operators are quantum channels. Let us take a closer look at these objects. For understanding Markov operators, we need the definition of a special type of mapping between measurable spaces.

Definition 3.4 (Markov kernel)**.** *Let $(X,\mathscr{A})$ and $(Y,\mathscr{B})$ be measurable spaces. κ is a Markov kernel with source $(X,\mathscr{A})$ and target $(Y,\mathscr{B})$, if it is a mapping $\kappa : \mathscr{B}\times X \to [0,1]$ with the following two properties:*

- *for all $B_0 \in B$ the function $x \mapsto \kappa(B_0,x)$ is $\mathscr{A}$-measurable,*
- *for all $x_0 \in X$ the function $B \mapsto \kappa(B,x_0)$ is a probability measure on $(Y,\mathscr{B})$.*

An insightful example of a Markov kernel can be provided by reformulating a Markov process on a countable space X. If $X = Y$ and $\mathscr{A} = \mathscr{B} = \mathscr{P}(X)$ and κ is a Markov chain, then the σ-additivity of measures ensures that we only have to define κ on the singleton sets $\{y\} \in Y$ for each $x \in X$, because $\kappa(B,x) = \sum_{y\in B} \kappa(\{y\},x)$ for all $x \in X$ and $B \in \mathscr{B}$. Taking $\kappa(\{y\},x) := \mathbb{P}(y,x)$, where $\mathbb{P}(y,x)$ is the transition probability from state x to state y we obtain a Markov kernel, furthermore, for each Markov kernel we get a proper (possibly infinite) transition matrix for a Markov process.

Definition 3.5 (Markov operator)**.** *Let $(X,\mathscr{A})$ be a measurable space and let V be the set of real measurable functions $f : (X,\mathscr{A}) \to (\mathbb{R},\mathscr{B}(\mathbb{R}))$. The linear operator P on the set V is a Markov operator if the following properties hold:*

- *P maps bounded functions to bounded functions,*
- *$P(\mathbf{1}) = \mathbf{1}$, where $\mathbf{1}$ is the constant 1 function,*
- *for each $f \geq 0$ function $Pf \geq 0$, i.e. it conserves positivity.*

If certain topological properties hold for the measurable space, then each Markov operator P can be written in the form $(Pf)(x) = \int_X f(y)\kappa(x,dy)$, where $\kappa(x,A)$ is a Markov kernel [7]. We call this integral form the kernel representation of the Markov operator P.

Now we can move on to the definiton of a quantum channel, but before that we need the definition of a completely positive operator, which is a stronger condition for an operator than positivity.

Definition 3.6 (Completely positive operator)**.** *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are Hilbert spaces and $\phi : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ is a linear map, then it induces the map $id_k \otimes \phi : \mathbb{C}^{k\times k} \otimes \mathscr{B}(\mathscr{H}_1) \to \mathbb{C}^{k\times k} \otimes \mathscr{B}(\mathscr{H}_2)$,*

where the domain of the map can be interpreted as $k \times k$ matrices with elements in $\mathscr{B}(\mathscr{H}_1)$ and ϕ transforms each of the matrix elements a_{ij} into $\phi(a_{ij})$.

ϕ is said to be k-positive if $id_k \otimes \phi$ is a positive map and completely positive if it is k-positive for all $k \in \mathbb{Z}_+$.

Definition 3.7 (Quantum channel)**.** *Let $\mathscr{H}_1$ and $\mathscr{H}_2$ be Hilbert spaces and $\Phi : \mathscr{T}_1(\mathscr{H}_1) \to \mathscr{T}_1(\mathscr{H}_2)$ an operator. Φ is a quantum channel if it is*

- *linear,*
- *trace preserving, i.e. $tr[\Phi(A)] = tr[A]$ for all $A \in \mathscr{T}_1(\mathscr{H}_1)$,*
- *completely positive.*

This definition explains why quantum channels are often abbreviated as CPTP (completely positive, trace preserving) maps.

The following theorem was proved by Choi in 1975 [4] and independently from his work applied for quantum mechanics by Kraus in 1971 [12]. Let $\mathscr{B}(\mathscr{H}_1, \mathscr{H}_2)$ denote the set of bounded linear operators from $\mathscr{H}_1$ to $\mathscr{H}_2$.

Theorem 3.8 (Choi–Kraus factorisation theorem [1])**.** *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are finite-dimensional Hilbert spaces, then for a linear map $\Phi : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ these two conditions are equivalent:*

1. *Φ is completely positive,*
2. *there exists $N \in \mathbb{N}$ and operators $\{K_i\}_{i=1}^N \subseteq \mathscr{B}(\mathscr{H}_1, \mathscr{H}_2)$ such that $\Phi(A) = \sum_{i=1}^N K_i A K_i^\dagger$ for all $A \in \mathscr{H}_1$.*

As a corollary of this theorem, we can provide a characterisation of quantum channels.

Theorem 3.9 (Characterisation of quantum channels in the finite-dimensional case)**.** *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are finite-dimensional Hilbert spaces, then for a linear map $\Phi : \mathscr{B}(\mathscr{H}_1) \to \mathscr{B}(\mathscr{H}_2)$ these two conditions are equivalent:*

1. *Φ is a quantum channel,*

2. *there exists* $N \in \mathbb{N}$ *and operators* $\{K_i\}_{i=1}^N \subseteq \mathscr{B}(\mathscr{H}_1, \mathscr{H}_2)$ *such that* $\Phi(A) = \sum_{i=1}^N K_i A K_i^\dagger$ *for all* $A \in \mathscr{H}_1$, *and* $\sum_{i=1}^N K_i^\dagger K_i = \mathbf{1}_{\mathscr{H}_1}$.

Definition 3.10 (Adjoint of quantum channel)**.** *If* $\Phi : \mathscr{L}(\mathscr{H}_1) \to \mathscr{L}(\mathscr{H}_2)$ *is a quantum channel with Kraus representation* $\Phi(A) = \sum_{i=1}^N K_i A K_i^\dagger$ *and* $\sum_{i=1}^N K_i^\dagger K_i = \mathbf{1}_{\mathscr{H}_1}$, *then its adjoint* $\Phi^\dagger : \mathscr{L}(\mathscr{H}_2) \to \mathscr{L}(\mathscr{H}_1)$ *with respect to the Hilbert–Schmidt inner product is given as* $\Phi^\dagger(A) := \sum_{i=1}^N K_i^\dagger A K_i$.

As $\Phi^\dagger(\mathbf{1}_{\mathscr{H}_2}) = \sum_{i=1}^N K_i^\dagger \mathbf{1}_{\mathscr{H}_2} K_i = \sum_{i=1}^N K_i^\dagger K_i = \mathbf{1}_{\mathscr{H}_1}$ for all $A \in \mathscr{L}(\mathscr{H}_2)$, the adjoint is unital.

To grasp how quantum channels work, let us take a look at some standard examples on qubit states, i.e. $\mathbb{C}^{2\times 2}$ density operators.

Example 3.11 (Identity channel)**.** The identity channel on qubit state acts as $\Phi(\rho) = \rho$, and its Kraus representation is given by only one Kraus operator, $K_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Indeed, $\Phi(\rho) = K_1 \rho K_1^\dagger = \rho$ and $K_1 K_1^\dagger = \mathbb{I}$.

Example 3.12 (Depolarizing channel)**.** The depolarizing channel maps the quantum state into itself with probability $1-p$ and into the completely mixed state $\frac{I}{2}$ with probability p. In general the completely mixed state is the state proportional to the identity matrix and it describes maximal randomness as it is the uniform distribution over all the orthonormal basis states. The depolarizing channel models the noise in the quantum system, because with nonzero probability the initial state is replaced by some uniform noise.

The formula for the depolarizing channel is $\Phi(\rho) = (1-p)\rho + \frac{pI}{2}$.

Taking the Bloch representation of a quantum state $\rho = \frac{1}{2}(I + \vec{r}\cdot\vec{\sigma})$ and using the identities $XXX = X, XYX = -Y, XZX = -Z, YXY = -X, YYY = Y, YZY = -Z, ZXZ = -X, ZYZ = -Y$, $ZZZ = Z$, we obtain

$$\begin{cases} X\rho X = \frac{1}{2}(I + r_x X - r_y Y - r_z Z) \\ Y\rho Y = \frac{1}{2}(I - r_x X + r_y Y - r_z Z) \\ Z\rho Z = \frac{1}{2}(I - r_x X - r_y Y + r_z Z), \end{cases}$$

and therefore summing up these equations we get $\frac{1}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z) = \frac{I}{2}$.

Substituting this into the initial equation yields

$$\Phi(\rho) = (1-p)\rho + p\left(\frac{1}{4}(\rho + X\rho X + Y\rho Y + Z\rho Z)\right) = \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}(X\rho X + Y\rho Y + Z\rho Z).$$

So one possible choice of the Kraus operators is

$$K_1 = \sqrt{1 - \frac{3p}{4}}I, K_2 = \frac{\sqrt{p}}{2}X, K_3 = \frac{\sqrt{p}}{2}Y, K_4 = \frac{\sqrt{p}}{2}Z.$$

Given a quantum mechanical system described by a self-adjoint time independent Hamiltonian H, we can define a unitary quantum channel via the time evolution of the system.
For proving this we need a lemma about the properties of the adjoint.

Lemma 3.13 (Adjoint of matrix exponential)**.** *For any operator A, we have* $\left(e^A\right)^\dagger = e^{A^\dagger}$.

Proof. By definition, $e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!} = 1 + A + \frac{A^2}{2!} + \cdots$

Well-known properties of the adjoint are $(AB)^\dagger = B^\dagger A^\dagger, (A+B)^\dagger = A^\dagger + B^\dagger, (cA)^\dagger = c^* A^\dagger$.

Using the first one it is easy to see that $(A^n)^\dagger = (A \cdots A)^\dagger = (A^\dagger)^n$, and therefore

$$(e^A)^\dagger = \left(\sum_{n=0}^{\infty} \frac{A^n}{n!}\right)^\dagger = \sum_{n=0}^{\infty} \left(\frac{A^n}{n!}\right)^\dagger = \sum_{n=0}^{\infty} \frac{(A^n)^\dagger}{n!} = \sum_{n=0}^{\infty} \frac{(A^\dagger)^n}{n!} = e^{A^\dagger}.$$

□

Now we can prove the following theorem.

Theorem 3.14 (Quantum channel of Hamiltonian)**.** *If H is a self-adjoint, time independent Hamiltonian and* $U(t) = e^{-iHt}$*, then for the operator* $\Phi : \mathbb{R} \times \mathscr{S}(\mathscr{H}) \to \mathscr{S}(\mathscr{H})$ *we have that* $\Phi(t,\rho) = \Phi_t(\rho) = U(t)\rho(0)U^\dagger(t)$ *is a quantum channel for all* $t \in \mathbb{R}$.

Proof. Using the characterisation theorem for quantum channels we only have to show that U is a unitary bounded operator.
Applying the previous lemma we get $(U(t))^\dagger = (e^{-iHt})^\dagger = e^{(-iHt)^\dagger} = e^{iH^\dagger t} = e^{iHt}$, and therefore $U(t)^\dagger U(t) = e^{iHt}e^{-iHt} = 1$, so $U^\dagger U = I$.
Every unitary operator has norm 1 and hence is bounded.

□

We also provide a counterexample to show that not all linear, trace preserving maps $\Phi : \mathscr{L}(\mathscr{H}) \to \mathscr{L}(\mathscr{H})$ are quantum channels.

Definition 3.15 (Transpose)**.** *With standard braket notation, for any operator $A \in \mathscr{L}(\mathscr{H})$, its transpose is a map $A^T \in \mathscr{L}(\mathscr{H}^*)$ such that $A^T(\langle\psi|) := \langle\psi|A$, i.e. $\varphi(Ax) = (A^T\varphi)(x)$ for all $\varphi \in \mathscr{H}^*, x \in \mathscr{H}$. It is easy to check that for finite matrices this definition coincides with the usual matrix transpose definition.*

For $\mathbb{C}^{n\times n}$ matrices the transpose Φ is a linear, trace preserving map, however it is not completely positive. To see that, it is sufficient to show that $\mathrm{id}_2 \otimes \Phi$ is not a positive map.

$$(\mathrm{id}_2 \otimes \Phi)\left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

It is sufficient to see that $A = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right)$ is positive semidefinite and $B = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$ is not.

On one hand,

$(a,b,c,d)A(a,b,c,d)^T = (a,b,c,d)(a+d,0,0,a+d)^T = a^2+ad+da+d^2 = (a+d)^2 \geq 0,$

on the other hand, $(a,b,c,d)B(a,b,c,d)^T = (a,b,c,d)(a,c,b,d)^T = a^2+2bc+d^2$, so for the vector $(a,b,c,d) = (0,1,-1,0)$ we have $(0,1,-1,0)B(0,1,-1,0)^T = -2 < 0$.

Important types of quantum channels on single-qubit systems are the flip channels, which can be described with the help of the Pauli matrices and which play an important role for instance in quantum error correction.

Definition 3.16 (Bit flip channel)**.** *The bit flip channel is defined as* $\Phi(\rho) = (1-p)\rho + pX\rho X$*, so its Kraus representation is* $K_1 = \sqrt{1-p}I, K_2 = \sqrt{p}X$*. The name bit flip originates from the fact that it flips the states* $|0\rangle$ *and* $|1\rangle$ *with probability p.*

Indeed, $X|0\rangle = |1\rangle$ *and* $X|1\rangle = |0\rangle$*.*

Definition 3.17 (Phase flip channel)**.** *The phase flip channel is defined as* $\Phi(\rho) = (1-p)\rho + pZ\rho Z$*, so its Kraus representation is* $K_1 = \sqrt{1-p}I, K_2 = \sqrt{p}Z$*. The name of the channel is phase flip because it switches the phase of the state* $|1\rangle$ *with probability p.*

Indeed, $Z|0\rangle = |0\rangle, Z|1\rangle = -|1\rangle$*.*

Definition 3.18 (Bit-phase flip channel)**.** *The bit-phase flip channel is defined as* $\Phi(\rho) = (1-p)\rho + pY\rho Y$*, so its Kraus representation is* $K_1 = \sqrt{1-p}I, K_2 = \sqrt{p}Y$*. The bit-flip channel is the combination of the previous two channels because* $Y = iXZ$*, so it flips both the basis states and their phases.*

Indeed, $Y|0\rangle = iXZ|0\rangle = iX|0\rangle = i|1\rangle$ *and* $Y|1\rangle = iXZ|1\rangle = iX(-|1\rangle) = -i|0\rangle$*.*

3.3 Optimal transport with quantum channels

In classical optimal transport theory, the transference plans π can be viewed as Markov kernels, because the mapping $B \to \pi(x_0, B)$ describes the amount transported from x_0 to $B \subseteq Y$, and hence defines a probability measure on Y for all $x_0 \in X$.

In quantum optimal transport theory, quantum channels play the roles of Markov kernels: the complete positivity of quantum channels corresponds to the nonnegativity of probability measures and the trace-preserving property corresponds to the normalization condition $\pi(X \times Y) = 1$.

Definition 3.19 (Quantum optimal transport plan)**.** *A quantum optimal transport plan* Φ *from a state* $\rho \in \mathscr{S}(\mathscr{H}_1)$ *to a state* $\omega \in \mathscr{S}(\mathscr{H}_2)$ *is a quantum channel from* $\mathscr{T}_1(\mathscr{H}_1)$ *to* $\mathscr{T}_1(\mathscr{H}_2)$ *such that* $\Phi(\rho) = \omega$*.*

The following important theorem claims that any state on a Hilbert space can be represented as a partial trace of a pure state on a larger Hilbert space.

Theorem 3.20 (Purification of a state)**.** *Let $\mathscr{H}$ be a finite-dimensional Hilbert space. For any $\rho \in \mathscr{S}(\mathscr{H})$ there is a quantum system $\mathscr{K}$ and a pure state $|\psi\rangle\langle\psi| \in \mathscr{S}(\mathscr{H}\otimes\mathscr{K})$ such that $tr_{\mathscr{K}}[|\psi\rangle\langle\psi|] = \rho$.*

Proof. Every state $\rho \in \mathscr{S}(\mathscr{H})$ can be written as $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ where $(|\psi_i\rangle)_i \subset \mathscr{H}$ is an orthonormal set, $p_i \geq 0, \sum_i p_i = 1$. By choosing $\mathscr{K} = \mathscr{H}^*$ and
$|\Psi\rangle := \sum_i \sqrt{p_i}|\psi_i\rangle \otimes \langle\psi_i| \in \mathscr{H}\otimes\mathscr{H}^*$ we have
$|\Psi\rangle\langle\Psi| = \sum_{i,j}\sqrt{p_i p_j}(|\psi_i\rangle\otimes\langle\psi_i|)(\langle\psi_j|\otimes|\psi_j\rangle)$, and hence
$\mathrm{tr}_{\mathscr{H}^*}[|\Psi\rangle\langle\Psi|] = \sum_i p_i |\psi_i\rangle\langle\psi_i| = \rho$.

□

The vector $|\Psi\rangle$ we constructed in the proof is called the canonical purification of the state ρ and is denoted as $||\varphi\rangle\rangle$.

The density operator associated to a quantum optimal transport plan Φ from ρ to σ is $\pi_\Phi = (\Phi\otimes\mathbf{1}_{\mathscr{L}(\mathscr{H}^*)})(|\Psi\rangle\langle\Psi|) \in \mathscr{S}(\mathscr{H}\otimes\mathscr{H}^*)$, where $|\Psi\rangle = ||\rho\rangle\rangle$. From this definition we obtain $\mathrm{tr}_{\mathscr{H}}[\pi_\Phi] = \rho^T$ and $\mathrm{tr}_{\mathscr{H}^*}[\pi_\Phi] = \Phi(\rho) = \sigma$.

Definition 3.21 (Quantum coupling)**.** *A quantum coupling between the states $\rho, \sigma \in \mathscr{S}(\mathscr{H})$ is a state $\pi \in \mathscr{S}(\mathscr{H}\otimes\mathscr{H}^*)$ such that $tr_{\mathscr{H}}[\pi] = \rho^T$ and $tr_{\mathscr{H}^*}[\pi] = \sigma$. The set of quantum couplings between ρ and σ is denoted as $C(\rho,\sigma)$.*

The cost operator is given via a set of observables $\mathscr{A} = \{A_1, \ldots, A_n\}$. In this thesis, we focus exclusively on quadratic costs, which are defined as

$$C_{\mathscr{A}} = \sum_{j=1}^{n} (A_j \otimes I^T - I \otimes A_j^T)^2.$$

The quantum Wasserstein distance is defined analogously to the classical optimal transport case.

Definition 3.22 (Quantum Wasserstein distance[8])**.** *The quantum Wasserstein distance corresponding to the cost operator generated by the set of observables $\mathscr{A}$ is*

$$D^2_{\mathscr{A}}(\rho,\sigma) = \inf_{\pi\in C(\rho,\sigma)} \left(tr_{\mathscr{H}\otimes\mathscr{H}^*}[\pi C_{\mathscr{A}}]\right).$$

4 Quantum Wasserstein isometries

4.1 Summary of known results

As in the case of basic quantum mechanical concepts, it is again convenient to create a table which compares the definitions of the classical and the quantum formulation of the transport problem. In the classical formulation we take the case $X = Y$. The following pairs of concepts are equivalents of each other in the two formulations: probability measures and density operators, marginal measures and partial traces of density operators, integrals and traces. The correspondence between more subtle concepts is described in the table below.

concept	classical notation	quantum notation
quadratic transport cost	$c(x,y) = d^2(x,y)$, where $d(x,y)$ is a metric on $X \times X$	for an $\mathscr{A} = \{A_1, \ldots, A_n\}$ set of observables the generated cost is $C_{\mathscr{A}} = \sum_{j=1}^{n} (A_j \otimes I^T - I \otimes A_j^T)^2$
transport plans	set of transference plans $\Pi(\mu,\nu)$: $\pi : X \times X \to [0,1]$ probability measure with $\pi[A \times X] = \mu(A)$ and $\pi[X \times B] = \nu(B)$	set of couplings $C(\rho,\sigma)$: $\pi \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H}^*)$: $\mathrm{tr}_{\mathscr{H}^*}[\pi] = \sigma, \mathrm{tr}_{\mathscr{H}}[\pi] = \rho^T$
quadratic Wasserstein distance	$W_2^2(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int_{X \times X} d^2(x,y) d\pi(x,y) \right)$	with the given set $\mathscr{A}$ of observables: $D^2_{\mathscr{A}}(\rho,\sigma) = \inf_{\pi \in C(\rho,\sigma)} (\mathrm{tr}_{\mathscr{H} \otimes \mathscr{H}^*}[\pi C_{\mathscr{A}}])$

An important difference between the classical and the quantum Wasserstein distance is that while the classical Wasserstein distance is a proper metric, the quantum Wasserstein distance is not. In fact, the distance of a state of itself can be greater than zero. In order to avoid this problem,

the Wasserstein divergence was introduced.

Definition 4.1 (Wasserstein divergence)**.** *The Wasserstein divergence generated by a set $\mathscr{A}$ of observables is defined as*

$$d_{\mathscr{A}}(\rho,\sigma) := \sqrt{D^2_{\mathscr{A}}(\rho,\sigma) - \frac{1}{2}(D^2_{\mathscr{A}}(\rho,\rho) + D^2_{\mathscr{A}}(\sigma,\sigma))}.$$

Now it is trivial that for all states ρ the divergence is $d_{\mathscr{A}}(\rho,\rho) = 0$. Moreover, it was recently proved that under certain further conditions it is a proper metric [3]. We have some results about the isometries in the special case $\mathbb{C}^2$. A we have seen before, the state $\rho \in \mathscr{S}(\mathbb{C}^2)$ can be represented as $\rho = \frac{1}{2}(I + x\sigma_x + y\sigma_y + z\sigma_z)$, where $\vec{b}_\rho = (x,y,z) \in \mathbb{R}^3, ||\vec{b}_\rho|| \leq 1$ is the Bloch vector of ρ.

Claim 6. For a state $\rho \in \mathscr{S}(\mathbb{C}^2)$ we have $\vec{b}_\rho = (\mathrm{tr}(\sigma_x\rho), \mathrm{tr}(\sigma_y\rho), \mathrm{tr}(\sigma_z\rho))$.

Proof. The set $\{I, \sigma_x, \sigma_y, \sigma_z\}$ is orthonormal and hence $\sigma_x\rho = \frac{1}{2}\sigma_x(I + x\sigma_x + y\sigma_y + z\sigma_z) = \frac{1}{2}xI$, therefore $\mathrm{tr}(\sigma_x\rho) = \mathrm{tr}(\frac{1}{2}xI) = x$. From similar calculations we get $\mathrm{tr}(\sigma_y\rho) = y$ and $\mathrm{tr}(\sigma_z\rho) = z$. □

Let us consider the cost operator C_{sym} generated by the set of Pauli matrices $\mathscr{A} = \{\sigma_x, \sigma_y, \sigma_z\}$, which is referred to as "symmetric", because it is constructed using all three matrices in a symmetric manner.

Claim 7. $C_{sym} = \begin{pmatrix} 4 & 0 & 0 & -4 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ -4 & 0 & 0 & 4 \end{pmatrix}$.

Proof. By definition, $C_{sym} = (\sigma_x \otimes I^T - I \otimes \sigma_x^T)^2 + (\sigma_y \otimes I^T - I \otimes \sigma_y^T)^2 + (\sigma_z \otimes I^T - I \otimes \sigma_z^T)^2$. Computing the three terms of the sum we obtain the matrix above. □

Let us introduce the following notations: for $\mathscr{A} = \{\sigma_x, \sigma_y, \sigma_z\}$ the corresponding Wasserstein distance is $D^2_{sym}(\rho,\xi) = D^2_{\mathscr{A}}(\rho,\xi)$ and the Wasserstein divergence is $d_{sym}(\rho,\xi) = d_{\mathscr{A}}(\rho,\xi)$. The set of pure states is $\mathscr{P}_1(\mathbb{C}^2) := \{\rho \in \mathscr{S}(\mathbb{C}^2) : \rho = |\psi\rangle\langle\psi|, |\psi\rangle \in \mathbb{C}^2\}$.

Definition 4.2 (Wasserstein isometry)**.** *A map $\Phi : \mathscr{S}(\mathscr{H}) \to \mathscr{S}(\mathscr{H})$ is an isometry of the quantum Wasserstein divergence $d_{\mathscr{A}}$ corresponding to the cost operator $C_{\mathscr{A}}$, if $d_{\mathscr{A}}(\Phi(\rho),\Phi(\xi)) = d_{\mathscr{A}}(\rho,\xi)$ for all $\rho,\xi \in \mathscr{S}(\mathscr{H})$.*

Similarly, Φ is an isometry of the quantum Wasserstein distance corresponding to $C_{\mathscr{A}}$, if $D_{\mathscr{A}}(\Phi(\rho),\Phi(\xi)) = D_{\mathscr{A}}(\rho,\xi)$ for all $\rho,\xi \in \mathscr{S}(\mathscr{H})$.

According to Theorem 2.18., in the space of bounded linear operators on a finite-dimensional Hilbert space $\mathscr{H}$ unitary operators are exactly those which preserve the inner product, i.e. for all $x,y \in \mathscr{H}$ we have $\langle Ux|Uy\rangle = \langle x|y\rangle$. This definition gives meaning to the introduction of anti-unitary operators in the following way.

Definition 4.3 (Anti-unitary operator)**.** *On a finite-dimensional Hilbert space $\mathscr{H}$ the bounded linear operator U is anti-unitary if for all $x,y \in \mathscr{H}$ we have $\langle Ux|Uy\rangle = \overline{\langle x|y\rangle} = \langle y|x\rangle$.*

The following theorem by Richárd Simon and Dániel Virosztek gives a characterisation of isometries of the quantum Wasserstein divergence d_{sym} [18].

Theorem 4.4. (i) *Let $\Phi : \mathscr{S}(\mathbb{C}^2) \to \mathscr{S}(\mathbb{C}^2)$ be an isometry of d_{sym} such that Φ maps pure states to pure states. Then Φ is a conjugation by a unitary or anti-unitary operator, i.e. $\Phi(\rho) = U\rho U^\dagger$ for all $\rho \in \mathscr{S}(\mathbb{C}^2)$, where U is a unitary or an anti-unitary operator on $\mathbb{C}^2$.*

(ii) *Unitary and anti-unitary conjugations are d_{sym}-isometries which map pure states to pure states.*

Now let us consider the cost operator C_z generated by the singleton set $\mathscr{A} = \{\sigma_z\}$.

Claim 8. $C_z = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0. \end{pmatrix}$.

Proof.

$$C_z = (\sigma_z \otimes I^T - I \otimes \sigma_z^T)^2 = \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right)^2$$

$$= \left(\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \right)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

□

With D_z denoting the Wasserstein distance corresponding to the cost operator C_z and $\mathbf{b}_\rho$ denoting the Bloch vector of the state $\rho \in \mathscr{S}(\mathbb{C}^2)$, we have the following equivalent conditions in the qubit state space [18].

Theorem 4.5. *Let $\Phi : \mathscr{S}(\mathbb{C}^2) \to \mathscr{S}(\mathbb{C}^2)$ be a map. The following two conditions are equivalent.*

(i) *Φ is a quantum Wasserstein isometry with respect to the Wasserstein distance D_z.*

(ii) *For all $\rho \in \mathscr{S}(\mathbb{C}^2)$ states $|\mathbf{b}_{\Phi(\rho)}| = |\mathbf{b}_\rho|$, and either $(\mathbf{b}_{\Phi(\rho)})_3 = (\mathbf{b}_\rho)_3$ for all states or $(\mathbf{b}_{\Phi(\rho)})_3 = -(\mathbf{b}_\rho)_3$ for all states.*

4.2 New results

The aim of current research is to generalise the previous results to other quantum systems and to other cost generating sets of operators.

Theorem 4.6. *Let $\mathscr{H}$ be a finite-dimensional Hilbert space with* $\dim \mathscr{H} = n$ *and let $A = A^\dagger \in \mathscr{O}(\mathscr{H})$ be an observable.*

(i) *Let $\Phi : \mathscr{S}(\mathscr{H}) \to \mathscr{S}(\mathscr{H})$ be defined as $\Phi(\rho) = e^{it_\rho A} \rho e^{-it_\rho A}$,*
where $t_\rho \in \mathbb{R}$ for all $\rho \in \mathscr{S}(\mathscr{H})$ and $\rho \mapsto t_\rho$ is an arbitrary map from $\mathscr{S}(\mathscr{H})$ to $\mathbb{R}$.

Then Φ is a quantum Wasserstein isometry, i.e. for all states $\rho, \omega \in \mathscr{S}(\mathscr{H})$ we have $D^2_{\{A\}}(\Phi(\rho), \Phi(\omega)) = D^2_{\{A\}}(\rho, \omega)$.

(ii) *If $\Phi : \mathscr{S}(\mathscr{H}) \to \mathscr{S}(\mathscr{H})$ is a $D_{\{A\}}$-isometry and both the largest and the smallest eigenvalues of the operator A have multiplicity* 1*, then*

(a) Either $tr_{\mathscr{H}}[\Phi(\rho)(A^2 - 2\lambda_1 A)] = tr_{\mathscr{H}}[\rho(A^2 - 2\lambda_1 A)]$ for all $\rho \in \mathscr{S}(\mathscr{H})$, where λ_1 is the largest eigenvalue of A, or
$tr_{\mathscr{H}}[\Phi(\rho)(A^2 - 2\lambda_n A)] + \lambda_n^2 = tr_{\mathscr{H}}[\rho(A^2 - 2\lambda_1 A)] + \lambda_1^2$ for all $\rho \in \mathscr{S}(\mathscr{H})$, where λ_1 is the largest and λ_n is the smallest eigenvalue of A.

(b) Moreover, $\sum_{j,k}(\lambda_j - \lambda_k)^2 |\langle \varphi_j | \sqrt{\Phi(\rho)} | \varphi_k \rangle|^2 = \sum_{j,k}(\lambda_j - \lambda_k)^2 |\langle \varphi_j | \sqrt{\rho} | \varphi_k \rangle|^2$, where λ_k are the eigenvalues of A with the corresponding eigenvectors φ_k.

Proof. (i): The Wasserstein distance induced by A is

$$D^2_{\{A\}}(\rho, \omega) = \inf_{\pi \in C(\rho,\omega)} \left(\mathrm{tr}_{\mathscr{H} \otimes \mathscr{H}^*}[\pi (A \otimes I^T - I \otimes A^T)^2] \right),$$

so $\pi \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H}^*)$, $\mathrm{tr}_{\mathscr{H}^*}[\pi] = \omega, \mathrm{tr}_{\mathscr{H}}[\pi] = \rho^T$.
Let us use the notation $t := t_\omega, s := t_\rho, U_A(x) = e^{ixA}$,
hence $\Phi(\rho) = U_A(s)\rho U_A^\dagger(s), \Phi(\omega) = U_A(t)\omega U_A^\dagger(t)$.
Any coupling $\pi \in \mathscr{S}(\mathscr{H} \otimes \mathscr{H}^*)$ has the form $\pi = \sum_k \alpha_k \otimes \beta_k^T$, where $\alpha_k, \beta_k \in \mathscr{L}(\mathscr{H})$.
$U_A(t,s) := U_A(t) \otimes (U_A^\dagger(s))^T$ is a unitary operator on $\mathscr{H} \otimes \mathscr{H}^*$.
We are going to show that if $\pi \in C(\rho, \omega)$, then $U_A(t,s)\pi U_A^\dagger(t,s) \in C(\Phi(\rho), \Phi(\omega))$.
Indeed,

$$\begin{aligned}
&\mathrm{tr}_{\mathscr{H}^*}[U_A(t,s)\pi U_A^\dagger(t,s)] \\
&\qquad = \sum_k \mathrm{tr}_{\mathscr{H}^*} \left[(U_A(t) \otimes (U_A^\dagger(s))^T)(\alpha_k \otimes \beta_k^T)(U_A^\dagger(t) \otimes U_A^T(s)) \right] \\
&\qquad = \sum_k \mathrm{tr}_{\mathscr{H}^*} \left[(U_A(t)\alpha_k U_A^\dagger(t)) \otimes ((U_A^\dagger(s))^T \beta_k^T U_A^T(s)) \right] \\
&\qquad = \sum_k U_A(t)\alpha_k U_A^\dagger(t) \cdot \mathrm{tr}_{\mathscr{H}^*} \left[(U_A^\dagger(s))^T \beta_k^T U_A^T(s) \right] \\
&\qquad = U_A(t) \left(\sum_k \alpha_k \cdot \mathrm{tr}_{\mathscr{H}^*} \left[(U_A^\dagger(s))^T \beta_k^T U_A^T(s) \right] \right) U_A^\dagger(t).
\end{aligned}$$

Using the cyclic property of the trace we have
$\mathrm{tr}_{\mathscr{H}^*}\left[(U_A^\dagger(s))^T\beta_k^T U_A^T(s)\right] = \mathrm{tr}_{\mathscr{H}^*}\left[\beta_k^T U_A^T(s)(U_A^\dagger(s))^T\right] = \mathrm{tr}_{\mathscr{H}^*}\left[\beta_k^T\right]$, and hence

$$\begin{aligned}
&U_A(t)\left(\sum_k \alpha_k \cdot \mathrm{tr}_{\mathscr{H}^*}\left[(U_A^\dagger(s))^T\beta_k^T U_A^T(s)\right]\right)U_A^\dagger(t)\\
&= U_A(t)\left(\sum_k \alpha_k \mathrm{tr}_{\mathscr{H}^*}[\beta_k^T]\right)U_A^\dagger(t)\\
&= U_A(t)\mathrm{tr}_{\mathscr{H}^*}\left[\sum_k \alpha_k\otimes\beta_k^T\right]U_A^\dagger(t)\\
&= U_A(t)\mathrm{tr}_{\mathscr{H}^*}[\pi]U_A^\dagger(t)\\
&= U_A(t)\omega U_A^\dagger(t) = \Phi(\omega).
\end{aligned}$$

We could show with similar computation that
$\mathrm{tr}_{\mathscr{H}}[U_A(t,s)\pi U_A^\dagger(t,s)] = (U_A(s)\rho U_A^\dagger(s))^T = \Phi(\rho)^T$.
The operator $U_A(t,s)$ is invertible and therefore with the same reasing we can show that if $U_A(t,s)\pi U_A^\dagger(t,s) \in C(\Phi(\rho),\Phi(\omega))$, then $\pi \in C(\rho,\omega)$.

Now we only need to show that π and $U_A(t,s)\pi U_A^\dagger(t,s)$ generate the same transport cost. The trace is linear, so it is sufficient to check this for couplings $\pi = \alpha\otimes\beta^T$, where $\alpha,\beta\in\mathscr{L}(\mathscr{H})$. The operator e^{itA} commutes with both A and I and therefore $U_A^\dagger(t,s)(A\otimes I^T - I\otimes A^T) = (e^{-itA}\otimes (e^{isA})^T)(A\otimes I^T - I\otimes A^T) = (A\otimes I^T - I\otimes A^T)U_A^\dagger(t,s)$. Using this and the cyclic property of the trace we have

$$\begin{aligned}
&\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[(U_A(t,s)\pi U_A^\dagger(t,s))(A\otimes I^T - I\otimes A^T)^2\right]\\
&= \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[U_A(t,s)\pi(A\otimes I^T - I\otimes A^T)^2 U_A^\dagger(t,s)\right]\\
&= \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[U_A^\dagger(t,s)U_A(t,s)\pi(A\otimes I^T - I\otimes A^T)^2\right]\\
&= \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\pi(A\otimes I^T - I\otimes A^T)^2\right].
\end{aligned}$$

(ii): (a): First let us consider the case when one of the states $\rho,\omega\in\mathscr{S}(\mathscr{H})$ is pure. In this case the only coupling is the trivial coupling $\pi = \omega\otimes\rho^T$.

$$
\begin{aligned}
&D^2_{\{A\}}(\rho,\omega)\\
&\qquad = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[(\omega\otimes\rho^T)(A\otimes I^T - I\otimes A^T)^2\right]\\
&\qquad = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[(\omega\otimes\rho^T)(A^2\otimes I^T - 2A\otimes A^T + I\otimes (A^T)^2\right]\\
&\qquad = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega A^2\otimes\rho^T\right] - 2\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega A\otimes\rho^T A^T\right] + \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega\otimes\rho^T(A^T)^2\right].
\end{aligned}
$$

Using $\mathrm{tr}_{\mathscr{H}}[\rho] = \mathrm{tr}_{\mathscr{H}}[\omega] = 1$ and the identity
$\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[X\otimes Y^T\right] = \mathrm{tr}_{\mathscr{H}}[X]\cdot\mathrm{tr}_{\mathscr{H}^*}\left[Y^T\right] = \mathrm{tr}_{\mathscr{H}}[X]\cdot\mathrm{tr}_{\mathscr{H}}[Y]$, we have

$$
\begin{aligned}
&\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega A^2\otimes\rho^T\right] - 2\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega A\otimes\rho^T A^T\right] + \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega\otimes\rho^T(A^T)^2\right]\\
&\qquad = \mathrm{tr}_{\mathscr{H}}\left[\omega A^2\right]\cdot\mathrm{tr}_{\mathscr{H}}[\rho] - 2\mathrm{tr}_{\mathscr{H}}[\omega A]\cdot\mathrm{tr}_{\mathscr{H}}[\rho A] + \mathrm{tr}_{\mathscr{H}}[\omega]\cdot\mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right]\\
&\qquad = \mathrm{tr}_{\mathscr{H}}\left[\omega A^2\right] - 2\mathrm{tr}_{\mathscr{H}}[\omega A]\cdot\mathrm{tr}_{\mathscr{H}}[\rho A] + \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right].
\end{aligned}
$$

Any self-adjoint operator A can be written in the form $A = \sum_{j=1}^n \lambda_j p_j$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of A with corresponding eigenvectors $|\varphi_j\rangle$ and eigenprojections $p_j = |\varphi_j\rangle\langle\varphi_j|$.

Using the fact that

$$
\mathrm{tr}_{\mathscr{H}}[p_1 A] = \mathrm{tr}_{\mathscr{H}}[|\varphi_1\rangle\langle\varphi_1|A] = \mathrm{tr}_{\mathscr{H}}[\langle\varphi_1|A|\varphi_1\rangle] = \langle\varphi_1|A|\varphi_1\rangle = \lambda_1
$$

and

$$
\mathrm{tr}_{\mathscr{H}}\left[p_1 A^2\right] = \mathrm{tr}_{\mathscr{H}}\left[|\varphi_1\rangle\langle\varphi_1|A^2\right] = \mathrm{tr}_{\mathscr{H}}\left[\langle\varphi_1|A^2|\varphi_1\rangle\right] = \langle\varphi_1|A^2|\varphi_1\rangle = \lambda_1^2,
$$

we have that for an arbitrary $\rho\in\mathscr{S}(\mathscr{H})$:

$$
\begin{aligned}
&D^2_{\{A\}}(p_1,\rho)\\
&\qquad = \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] - 2\mathrm{tr}_{\mathscr{H}}[\rho A]\cdot\mathrm{tr}_{\mathscr{H}}[p_1 A] + \mathrm{tr}_{\mathscr{H}}\left[p_1 A^2\right]\\
&\qquad = \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] - 2\lambda_1\mathrm{tr}_{\mathscr{H}}[\rho A] + \lambda_1^2.
\end{aligned}
$$

Rearranging the terms we obtain $\mathrm{tr}_{\mathscr{H}}\left[\rho(A^2 - 2\lambda_1 A)\right] = D^2_{\{A\}}(p_1,\rho) - \lambda_1^2$, and with similar computations we get $\mathrm{tr}_{\mathscr{H}}[\rho(A^2 - 2\lambda_n A)] = D^2_{\{A\}}(p_n,p) - \lambda_n^2$.

Using the decomposition $A = \sum_{j=1}^n \lambda_j |\varphi_j\rangle\langle\varphi_j|$ and the fact that
$I = \sum_{j=1}^n |\varphi_j\rangle\langle\varphi_j|$ we obtain

$$
\begin{aligned}
&(A\otimes I^T - I\otimes A^T)^2 \\
&\quad= \left[\left(\sum_{j=1}^n \lambda_j |\varphi_j\rangle\langle\varphi_j|\right)\otimes\left(\sum_{k=1}^n |\varphi_k\rangle\langle\varphi_k|\right)^T - \left(\sum_{j=1}^n |\varphi_j\rangle\langle\varphi_j|\right)\otimes\left(\sum_{k=1}^n \lambda_k |\varphi_k\rangle\langle\varphi_k|\right)^T\right]^2 \\
&\quad= \left(\sum_{j=1}^n\sum_{k=1}^n \lambda_j |\varphi_j\rangle\langle\varphi_j|\otimes(|\varphi_k\rangle\langle\varphi_k|)^T - \sum_{j=1}^n\sum_{k=1}^n \lambda_k |\varphi_j\rangle\langle\varphi_j|\otimes(|\varphi_k\rangle\langle\varphi_k|)^T\right)^2 \\
&\quad= \left(\sum_{j=1}^n\sum_{k=1}^n (\lambda_j-\lambda_k) |\varphi_j\rangle\langle\varphi_j|\otimes(|\varphi_k\rangle\langle\varphi_k|)^T\right)^2.
\end{aligned}
$$

With $\{\tilde{\varphi}_j\}_{j=1}^n$ denoting the dual basis of $\{\varphi_j\}_{j=1}^n$ we have

$$
\begin{aligned}
&\left(\sum_{j=1}^n\sum_{k=1}^n (\lambda_j-\lambda_k) |\varphi_j\rangle\langle\varphi_j|\otimes(|\varphi_k\rangle\langle\varphi_k|)^T\right)^2 \\
&\qquad= \left(\sum_{j,k=1}^n (\lambda_j-\lambda_k) |\varphi_j\otimes\tilde{\varphi}_k\rangle\langle\varphi_j\otimes\tilde{\varphi}_k|\right)^2 \\
&\qquad= \sum_{j,k=1}^n (\lambda_j-\lambda_k)^2 |\varphi_j\otimes\tilde{\varphi}_k\rangle\langle\varphi_j\otimes\tilde{\varphi}_k|.
\end{aligned}
$$

We can write A with the eigenprojections $p_j = |\varphi_j\rangle\langle\varphi_j|$ as $A = \sum_{j=1}^n \lambda_j p_j$, therefore

$$
(A\otimes I^T - I\otimes A^T)^2 = \sum_{j,k=1}^n (\lambda_j-\lambda_k)^2 |\varphi_j\otimes\tilde{\varphi}_k\rangle\langle\varphi_j\otimes\tilde{\varphi}_k| = \sum_{j,k=1}^n (\lambda_j-\lambda_k)^2 p_j\otimes p_k^T.
$$

From this decomposition we can see that the operator norm of $(A\otimes I^T - I\otimes A^T)^2$ is its largest eigenvalue $(\lambda_1-\lambda_n)^2$, i.e. $||(A\otimes I^T - I\otimes A^T)^2||_\infty = (\lambda_1-\lambda_n)^2$. By definition, for any $\pi\in C(\rho,\omega)$ we have $||\pi||_1 = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}[\pi] = \mathrm{tr}_{\mathscr{H}}[\mathrm{tr}_{\mathscr{H}^*}[\pi]] = \mathrm{tr}_{\mathscr{H}}[\omega] = 1$.

We can apply the Hölder inequality $|\mathrm{tr}(XY)| \leq ||X||_1\cdot||Y||_\infty$ for $X=\pi, Y=(A\otimes I^T - I\otimes A^T)^2$ to get

$$
\mathrm{tr}[\pi(A\otimes I^T - I\otimes A^T)^2] \leq ||\pi||_1\cdot||(A\otimes I^T - I\otimes A^T||_\infty = (\lambda_1-\lambda_n)^2.
$$

Therefore $D^2_{\{A\}}(\rho,\omega) \leq (\lambda_1-\lambda_n)^2$ and for $\rho = p_1, \omega = p_n$ the equality holds, because

$$
D^2_{\{A\}}(p_1,p_n) = \mathrm{tr}\left[(p_n\otimes p_1^T)(\sum_{j,k=1}^n (\lambda_j-\lambda_k)^2 p_j\otimes p_k^T)\right] = (\lambda_n-\lambda_1)^2.
$$

For the following calculations we use that λ_1 and λ_n are eigenvalues with multiplicity one. We want to show that $D^2_{\{A\}}(\rho,\omega) = (\lambda_1-\lambda_n)^2$ can only hold if $\{\rho,\omega\} = \{p_1,p_n\}$.

Indeed,

$$\begin{aligned}
&D^2_{\{A\}}(\rho,\omega) \\
&\qquad \leq \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[(\omega\otimes\rho^T)(A\otimes I^T - I\otimes A^T)^2\right] \\
&\qquad = \sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2 \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[(\omega\otimes\rho^T)(p_j\otimes p_k^T)\right] \\
&\qquad = \sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2 \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\omega p_j\otimes\rho^T p_k^T\right] \\
&\qquad = \sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2 \mathrm{tr}_{\mathscr{H}}\left[\omega p_j\right]\cdot \mathrm{tr}_{\mathscr{H}}\left[\rho p_k\right] \\
&\qquad \leq \sum_{j,k=1}^{n} (\lambda_1-\lambda_n)^2 \mathrm{tr}_{\mathscr{H}}\left[\omega p_j\right]\cdot \mathrm{tr}_{\mathscr{H}}\left[\rho p_k\right] \\
&\qquad \leq (\lambda_1-\lambda_n)^2,
\end{aligned}$$

where the last inequality holds because the terms $\mathrm{tr}_{\mathscr{H}}\left[\omega p_j\right]\cdot \mathrm{tr}_{\mathscr{H}}\left[\rho p_k\right]$ form a convex combination of the values $(\lambda_j-\lambda_k)^2$. Combining this with our assumption that λ_1 and λ_n have multiplicity one, equality can only hold if all the terms $\mathrm{tr}_{\mathscr{H}}\left[\omega p_j\right]\cdot \mathrm{tr}_{\mathscr{H}}\left[\rho p_k\right]$ are zero except when $\{j,k\} = \{1,n\}$. In that case, $\omega = \alpha p_1 + (1-\alpha)p_n, \rho = \beta p_1 + (1-\beta)p_n$ and

$$\begin{aligned}
(\lambda_1-\lambda_1)^2\alpha\beta + (\lambda_1-\lambda_n)^2\alpha(1-\beta) + (\lambda_n-\lambda_1)^2(1-\alpha)\beta + (\lambda_n-\lambda_n)^2(1-\alpha)(1-\beta) = \\
= (\lambda_1-\lambda_n)^2(\alpha+\beta-2\alpha\beta) = (\lambda_1-\lambda_n)^2(\alpha-\beta)^2 = (\lambda_1-\lambda_n)^2,
\end{aligned}$$

and therefore either $\omega = p_1$ and $\rho = p_n$ or $\omega = p_n$ and $\rho = p_1$.

As a corollary, if Φ is a quantum Wasserstein isometry, then either $\Phi(p_1) = p_1$ and $\Phi(p_n) = p_n$ or $\Phi(p_1) = p_n$ and $\Phi(p_n) = p_1$.

If $\Phi(p_1) = p_1$ and $\Phi(p_n) = p_n$, then $D^2_{\{A\}}(p_1,\rho) = D^2_{\{A\}}(\Phi(p_1),\Phi(\rho)) = D^2_{\{A\}}(p_1,\Phi(\rho))$ for all states $\rho\in\mathscr{S}(\mathscr{H})$, and therefore

$$\mathrm{tr}_{\mathscr{H}}[\Phi(\rho)(A^2-2\lambda_1 A)] = D^2_{\{A\}}(p_1,\Phi(\rho)) - \lambda_1^2 = D^2_{\{A\}}(p_1,\rho) - \lambda_1^2 = \mathrm{tr}_{\mathscr{H}}[\rho(A^2-2\lambda_1 A)].$$

If If $\Phi(p_1) = p_n$ and $\Phi(p_n) = p_1$, then

$$\begin{aligned}\mathrm{tr}_{\mathscr{H}}[\Phi(\rho)(A^2-2\lambda_n A)]+\lambda_n^2 & \\ &= D^2_{\{A\}}(p_n,\Phi(\rho)) = D^2_{\{A\}}(\Phi(p_1),\Phi(\rho)) = D^2_{\{A\}}(p_1,\rho)\\ &= \mathrm{tr}_{\mathscr{H}}[\rho(A^2-2\lambda_1 A)]+\lambda_1^2.\end{aligned}$$

(b): For an arbitrary state $\rho\in\mathscr{S}(\mathscr{H})$ according to Corollary 1 in [5] the equality

$$D^2_{\{A\}}(\rho,\rho) = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(A\otimes I^T - I\otimes A^T)^2\right]$$

holds.

Using the definition of canonical purification, the cyclic property of the trace and the identity $X\otimes Y^T\,||Z\rangle\rangle = ||XZY\rangle\rangle$, we have

$$\begin{aligned}&\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(A\otimes I^T - I\otimes A^T)^2\right]\\ &\quad= \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(A^2\otimes I^T)\right] + \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(I\otimes (A^T)^2)\right]\\ &\quad- 2\mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(A\otimes A^T)\right]\\ &\quad= \mathrm{tr}_{\mathscr{H}}\left[\mathrm{tr}_{\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\right]A^2\right] + \mathrm{tr}_{\mathscr{H}^*}\left[\mathrm{tr}_{\mathscr{H}}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\right](A^T)^2\right]\\ &\quad- 2\langle\langle\sqrt{\rho}||A\otimes A^T\,||\sqrt{\rho}\rangle\rangle\\ &\quad= \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] + \mathrm{tr}_{\mathscr{H}^*}\left[\rho^T(A^T)^2\right] - 2\langle\langle\sqrt{\rho}||Aqrt\rho A\rangle\rangle\\ &\quad= \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] + \mathrm{tr}_{\mathscr{H}^*}\left[\rho^T(A^T)^2\right] - 2\mathrm{tr}_{\mathscr{H}}\left[(\sqrt{\rho})^\dagger A\sqrt{\rho}A\right]\\ &\quad= \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] + \mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] - 2\mathrm{tr}_{\mathscr{H}}\left[(\sqrt{\rho}A)^2\right].\end{aligned}$$

We obtained that for all $\rho\in\mathscr{S}(\mathscr{H})$ the equality $D^2_{\{A\}}(\rho,\rho) = 2\left(\mathrm{tr}_{\mathscr{H}}\left[\rho A^2\right] - \mathrm{tr}_{\mathscr{H}}\left[(\sqrt{\rho}A)^2\right]\right)$ holds.

Decomposing the state ρ as

$$\rho = \sum_{j,k=1}^{n}\langle\varphi_j|\rho|\varphi_k\rangle\,|\varphi_j\rangle\langle\varphi_k| = \sum_{j,k=1}^{n}\langle\varphi_j|\rho|\varphi_k\rangle\,\varphi_j\otimes\tilde{\varphi}_k$$

and its square root $\sqrt{\rho}$ as

$$\sqrt{\rho} = \sum_{j,k=1}^{n}\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle\,|\varphi_j\rangle\langle\varphi_k| = \sum_{j,k=1}^{n}\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle\,\varphi_j\otimes\tilde{\varphi}_k$$

and putting our results together we have

$$
\begin{aligned}
&D^2_{\{A\}}(\rho,\rho) \\
&\quad = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[||\sqrt{\rho}\rangle\rangle\langle\langle\sqrt{\rho}||\,(A\otimes I^T - I\otimes A^T)^2\right] \\
&\quad = \mathrm{tr}_{\mathscr{H}\otimes\mathscr{H}^*}\left[\langle\langle\sqrt{\rho}||\,(A\otimes I^T - I\otimes A^T)^2\,||\sqrt{\rho}\rangle\rangle\right] \\
&\quad = \left\langle\left\langle \sum_{j,k=1}^{n} \langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle\,\varphi_j\otimes\tilde{\varphi}_k \middle|\middle| \sum_{l,m=1}^{n} (\lambda_l-\lambda_m)^2\,|\varphi_l\otimes\tilde{\varphi}_m\rangle\langle\varphi_l\otimes\tilde{\varphi}_m| \middle|\middle| \sum_{r,s=1}^{n} \langle\varphi_r|\sqrt{\rho}|\varphi_s\rangle\,\varphi_r\otimes\tilde{\varphi}_s \right\rangle\right\rangle \\
&\quad = \sum_{j,k=1}^{n}\sum_{l,m=1}^{n}\sum_{r,s=1}^{n} \overline{\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle}\,\langle\varphi_r|\sqrt{\rho}|\varphi_s\rangle\,(\lambda_l-\lambda_m)^2\,\langle\langle\varphi_j\otimes\tilde{\varphi}_k||\varphi_k\otimes\tilde{\varphi}_m\rangle\rangle\,\langle\langle\varphi_l\otimes\tilde{\varphi}_m||\varphi_r\otimes\tilde{\varphi}_s\rangle\rangle \\
&\quad = \sum_{j,k=1}^{n}\sum_{l,m=1}^{n}\sum_{r,s=1}^{n} \overline{\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle}\,\langle\varphi_r|\sqrt{\rho}|\varphi_s\rangle\,(\lambda_l-\lambda_m)^2\cdot\delta_{jl}\cdot\delta_{km}\cdot\delta_{lr}\cdot\delta_{ms} \\
&\quad = \sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2\,|\,\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle\,|^2.
\end{aligned}
$$

We know that $D^2_{\{A\}}(\Phi(\rho),\Phi(\rho)) = D^2_{\{A\}}(\rho,\rho)$, and computing $D^2_{\{A\}}(\Phi(\rho),\Phi(\rho))$ similarly yields

$$\sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2\,|\,\langle\varphi_j|\sqrt{\rho}|\varphi_k\rangle\,|^2 = \sum_{j,k=1}^{n} (\lambda_j-\lambda_k)^2\,|\,\langle\varphi_j|\sqrt{\Phi(\rho)}|\varphi_k\rangle\,|^2.$$

□

References

[1] *Representations of quantum channels*. https://www.uio.no/studier/emner/matnat/math/MAT4430/v23/timeplan/lecture4.pdf.

[2] *Wikipedia article bloch sphere*, Accessed: June 2023.

[3] Gergely Bunth, József Pitrik, Tamás Titkos, and Dániel Virosztek, *Metric property of quantum wasserstein divergences*, Physical Review A **110** (2024), no. 2, 022211.

[4] Man-Duen Choi, *Completely positive linear maps on complex matrices*, Linear algebra and its applications **10** (1975), no. 3, 285–290.

[5] Giacomo De Palma and Dario Trevisan, *Quantum optimal transport with quantum channels*, Annales henri poincaré, 2021, pp. 3199–3234.

[6] ———, *Quantum optimal transport: Quantum channels and qubits*, Optimal transport on quantum structures, 2024, pp. 203–239.

[7] Tanja Eisner, Bálint Farkas, Markus Haase, and Rainer Nagel, *Operator theoretic aspects of ergodic theory*, Vol. 272, Springer, 2015.

[8] György Pál Gehér, József Pitrik, Tamás Titkos, and Dániel Virosztek, *Quantum wasserstein isometries on the qubit state space*, Journal of Mathematical Analysis and Applications **522** (2023), no. 2, 126955.

[9] Fumio Hiai and Dénes Petz, *Introduction to matrix analysis and applications*, Springer Science & Business Media, 2014.

[10] Kazuki Ikeda, *Foundation of quantum optimal transport and applications*, Quantum Information Processing **19** (2020), no. 1, 25.

[11] Leonid V Kantorovich, *On the translocation of masses*, Dokl. akad. nauk. ussr (ns), 1942, pp. 199–201.

[12] Karl Kraus, *General state changes in quantum theory*, Annals of Physics **64** (1971), no. 2, 311–335.

[13] Gaspard Monge, *Mémoire sur la théorie des déblais et des remblais*, Mem. Math. Phys. Acad. Royale Sci. (1781), 666–704.

[14] Michael A Nielsen and Isaac L Chuang, *Quantum computation and quantum information*, Cambridge university press, 2010.

[15] Dénes Petz, *Quantum information theory and quantum statistics*, Springer Science & Business Media, 2007.

[16] Gabriel Peyré, Marco Cuturi, et al., *Computational optimal transport: With applications to data science*, Foundations and Trends® in Machine Learning **11** (2019), no. 5-6, 355–607.

[17] K. R. Sahasranand, *Techniques in discrete probability – lecture 1*, 2018.

[18] Richárd Simon and Dániel Virosztek, *Isometries of the qubit state space with respect to quantum wasserstein distances*, Linear Algebra and its Applications **714** (2025), 1–14.

[19] Zsigmond Tarcsay, *Funkcionálanalízis* (2018). in Hungarian.

[20] Cédric Villani, *Topics in optimal transportation*, Vol. 58, American Mathematical Soc., 2021.

MI alapú eszközök használatáról szóló nyilatkozat

Alulírott Szabó Eszter nyilatkozom, hogy szakdolgozatom elkészítése során az alább felsorolt feladatok elvégzésére a megadott MI alapú eszközöket alkalmaztam:

Feladat	**Felhasznált eszköz**	**Felhasználás helye**	**Megjegyzés**
LaTeX szintaktika	GPT-4o	Teljes dolgozat	-
Nyelvhelyesség ellenőrzése	Writefull	Teljes dolgozat	-

A felsoroltakon túl más MI alapú eszközt nem használtam.