

Approach to optimal quantum transport via states over time

Matt Hoogsteder-Riera¹, John Calsamiglia¹, and Andreas Winter^{1,2,3,4}

¹Grup d'Informació Quàntica, Departament de Física,
Universitat Autònoma de Barcelona, 08193 Bellaterra (BCN), Spain

²ICREA—Institució Catalana de Recerca i Estudis Avançats,
Pg. Lluís Companys, 23, 08010 Barcelona, Spain

³Department Mathematik/Informatik—Abteilung Informatik,
Universität zu Köln, Albertus-Magnus-Platz, 50923 Köln, Germany

⁴Institute for Advanced Study, Technische Universität München,
Lichtenbergstraße 2a, 85748 Garching, Germany
(04 April 2025)

We approach the problem of constructing a quantum analogue of the immensely fruitful classical transport cost theory of Monge from a new angle. Going back to the original motivations, by which the transport is a bilinear function of a mass distribution (without loss of generality a probability density) and a transport plan (a stochastic kernel), we explore the quantum version where the mass distribution is generalised to a density matrix, and the transport plan to a completely positive and trace preserving map. These two data are naturally integrated into their Jordan product, which is called *state over time* (“stote”), and the transport cost is postulated to be a linear function of it. We explore the properties of this transport cost, as well as the optimal transport cost between two given states (simply the minimum cost over all suitable transport plans). After that, we analyse in considerable detail the case of unitary invariant cost, for which we can calculate many costs analytically. These findings suggest that our quantum transport cost is qualitatively different from Monge’s classical transport.

1 Introduction

The Monge problem, and the field of optimal transport it has spawned [1, 2, 3, 4], asks for the best way to move mass, modeled as probability distributions, according to some cost functional. Explicitly, the optimal transport cost between two probability distributions μ, ν defined over space X is

$$c(\mu, \nu) := \inf_{\pi \in C(\mu, \nu)} \int_{X \times X} k(x, y) d\pi(x, y), \quad (1.1)$$

where $k(x, y)$ is a (usually real and positive) cost function and $C(\mu, \nu)$ is the set of *couplings* between μ and ν , that is the probability distributions on $X \times X$ with marginals μ and ν .

This approach allows for properties of the space X to be reflected in the cost through the function $k : X \times X \rightarrow \mathbb{R}$. In contrast, typical distinguishability measures on probability spaces, such as the total variation distance,¹ or the Kullback-Leibler divergence, can be computed focusing solely on the probability distributions and ignoring the underlying properties of the space X . As an example, let $X = \mathbb{R}$ and $\mu_0 = \delta_0$, $\mu_1 = \delta_1$ and $\nu = \delta_{-1}$. The total variation distances are equal: $\Delta_{TV}(\mu_0, \nu) = \Delta_{TV}(\mu_1, \nu) = 1$, but we can define an optimal transport cost that reflects the Euclidean metric of the reals by choosing $k(x, y) = |x - y|$. With this cost function, $\Delta_{OT}(\mu_0, \nu) = 1$, but $\Delta_{OT}(\mu_1, \nu) = 2$. This

Matt Hoogsteder-Riera: hoogstedermatt@gmail.com

John Calsamiglia: john.calsamiglia@uab.cat

Andreas Winter: andreas.winter@uab.cat

¹Note that the total variation distance can be defined in terms of optimal transport cost, by choosing the cost to be the trivial metric, $k(x, y) = 1 - \delta_{xy}$.

difference of the distances between 0 and -1 and between 1 and -1 is ignored by the total variation distance, but taken into account in the optimal transport cost.

Physically, the cost function k could be reflecting things like energy or information related quantities such as the number of bit flips necessary to transform a bit string into another one (known as the Hamming distance), the latter being a special case of a metric on X .

In the present paper, we make (yet another) attempt to quantise the theory of optimal classical transport from probability densities to quantum states. Our interpretation of the coupling $\pi(x, y)$ appearing in Eq. (1.1) is that it defines a stochastic matrix that yields output ν for input μ . We can recover this transformation explicitly using Bayes' Theorem: $p(y|x) = \pi(x, y)/\mu(x)$, and this is the “transport plan”: the map describing which fraction of mass at each given point is transferred to a given target point. The cost appearing on the right hand side of Eq. (1.1) is then bilinear in μ and $p(y|x)$, and we attempt to preserve this feature in our quantum version.

There have been attempts to generalise optimal transport to quantum systems dating back to the late 1990s [5, 6]. Some more recent approaches look at the problem directly through the primal formulation [7, 8, 9, 10, 11] (some form of couplings), through the dual formulation [12, 13] (which (1.1) has via linear programming duality) and through the continuous formulation [14, 15, 16, 17] (of certain dynamical semigroups having trajectories that are geodesic for a suitably defined optimal transport distance). Our work looks at the problem through the primal formulation. Our objective here is to define a quantum optimal transport where couplings have a straightforward physical interpretation as bilinear combinations of quantum states and quantum channels, similar to our interpretation of the classical couplings, leading to a cost assigned to each transport plan that is bilinear in the initial density and the quantum channel.

In Section 2 we motivate the mathematical and conceptual idea behind our formulation, in Section 3 we make the basic definitions, in Section 4 we show our main results regarding properties of our proposed quantum optimal transport, in Section 5 we discuss a specific cost metric that results from imposing unitary invariance and in Section 6 we discuss the messages to take away and open problems derived from our work.

1.1 Mathematical preliminaries

The Choi and Jamiołkowski isomorphisms [18, 19] are a relation between CPTP maps from $\mathcal{B}(\mathcal{H}_A)$ to $\mathcal{B}(\mathcal{H}_B)$ and bipartite states on the joint state space $\mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$. These isomorphisms are relevant to this work because they allow us to treat maps as operators, allowing for operations that would not be possible otherwise. In more general terms, these isomorphisms allow us to translate complete positivity of a map to positivity of an operator, with the latter usually being much easier to check. We also make use of this fact in Section 3 and Section 5.

Theorem 1.1 (Choi isomorphism [19]). *Let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ be a linear map and $|\Phi_+\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ the non-normalised maximally entangled vector on the composite system of A and a copy $A' \simeq A$ of it. The Choi operator, defined as*

$$C_{\mathcal{E}} = (\text{id}_A \otimes \mathcal{E})(|\Phi_+\rangle\langle\Phi_+|) \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (1.2)$$

recovers the action of $\mathcal{E}$ as

$$\mathcal{E}(x) = \text{Tr}_A \left[(x^T \otimes \mathbb{1}_B) C_{\mathcal{E}} \right]. \quad (1.3)$$

Moreover, $\mathcal{E}$ is completely positive if and only if $C_{\mathcal{E}}$ is positive semi-definite and $\mathcal{E}$ is trace preserving if and only if $\text{Tr}_B [C_{\mathcal{E}}] = \mathbb{1}_A$.

Jamiołkowski's statement of the same result is just the Choi isomorphism with a partial transpose.

Theorem 1.2 (Jamiołkowski isomorphism [18]). *Let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ be a linear map and $\mathcal{S} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_{A'})$ the swap operator on the joint system $A \otimes A'$. The Jamiołkowski operator, defined as*

$$J_{\mathcal{E}} = (\text{id}_A \otimes \mathcal{E})(\mathcal{S}) \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (1.4)$$

recovers the action of $\mathcal{E}$ as

$$\mathcal{E}(x) = \text{Tr}_A [(x \otimes \mathbb{1}_B)J_{\mathcal{E}}]. \quad (1.5)$$

Moreover, $\mathcal{E}$ is completely positive if and only if $J_{\mathcal{E}}^{T_A}$ is positive² and $\mathcal{E}$ is trace preserving if and only if $\text{Tr}_B [J_{\mathcal{E}}] = \mathbb{1}_A$.

It is clear from the theorems that if C and J are Choi and Jamiołkowski operators associated to the same CPTP map, then

$$J = C^{T_A}. \quad (1.6)$$

These theorems clarify how states and channels relate and how the two isomorphisms relate to each other. Each has its own advantages: the Choi matrix is manifestly positive, which is convenient when positivity must be checked explicitly, while the Jamiołkowski form is basis-independent. In the Choi construction one uses a maximally entangled state and the transpose on $\mathcal{H}_A$. Both of those steps depend on a choice of basis, so the Choi matrix is basis-dependent. In contrast, using the Jamiołkowski form avoids that dependence. Moreover, the swap operator plays a useful computational role in the Jamiołkowski picture, facilitating manipulations like exchanging tensor factors or expressing compositions more neatly.

We use the notation C and J for Choi and Jamiołkowski operators, respectively, throughout this work. Moreover, we generally refer to Choi or Jamiołkowski operators associated to CPTP map. Because our work deals with finite dimensional systems only, we also use the terms Choi and Jamiołkowski matrix interchangeably with operator. We usually use the Jamiołkowski operator, but the Choi operator will sometimes be more convenient to use in some circumstances. We always assume that the transpose is taken with respect to the canonical basis, or whichever privileged basis the problem presents. The basis dependence of the Choi matrix is only a problem when dealing with practical cases and never for general theoretical derivations.

Throughout the paper we use A^T for the transpose of a matrix A , $\bar{A}$ for the element-wise complex conjugate, and A^* for the conjugate transpose. With this notation, $A^* = \bar{A}^T$.

2 Jordan product motivation and properties

In this section we want to motivate our choice of coupling for the quantum optimal transport, since this choice is what differentiates our approach from the rest. The main observation was done in the previous section: a classical joint probability distribution can be interpreted as both a correlation function of a composite system and a map for the evolution over time of a single subsystem. The first interpretation can be generalised to quantum systems with bipartite states, as seen in, for example, [9, 11]. We choose the other interpretation, that of a stochastic map on a system. In the context of quantum mechanics this corresponds to quantum channels acting on a system. As we explain in the following, these questions on how to write states encoding the correlations of an input system with its output in quantum mechanics had been asked before in the context of quantum foundations.

The concept of *states over time* as introduced in [20] is based on previous attempts to formalise causal correlations in quantum theory, see [21, 22] and references within [20]. Its main motivation is to find an operator in the tensor product of the state spaces associated to two time points, which would capture the process transforming the initial state to the final one.

For this purpose, several properties have been put forward as desirables for a state over time: as a function of the initial state and the quantum channel it should be Hermitian preserving, bilinear in the two arguments, it should contain the classical case, reproduce the initial and final states as marginals, and, finally, keep the composability of channels. Through the later works of Fullwood and Parzygnat [23, 24, 25] and [26], we know that the Jordan product (also called symmetric bloom or Fullwood-Parzygnat state over time function in the literature) is the unique function that fulfils a slightly stronger set of axioms [26] that imply the properties above (that had been proposed in [20]).

²For simplicity we usually use positive instead of the more rigorous positive semi-definite.

The set of desiderata introduced in the paper are, in our opinion, well suited to the study of quantum optimal transport. Notably, the bilinearity, associativity (which translates to clear composability of the channels) and clear reduction to the classical case. Hermiticity and preservation of marginal states are useful mathematically as they allow us to define the problem as an SDP. Other approaches mentioned in [20], the LS state over time and the W state over time fail to fulfil bilinearity and composability; and preservation of the classical limit, respectively. That said, other works [8, 9] have approached this problem using the LS state over time, even if not explicitly acknowledged in the manuscripts, with some success. Our approach takes the generalised FJV state over time (from [20]) due to its nice properties not shared with other states over time.

In the context of quantum optimal transport, a state over time gives an initial state and a process (in the form of a CPTP map). As previously mentioned, this is similar to how in classical (optimal) transport a coupling $\pi(x, y)$ gives an initial state and final states (the marginals) and a process through the conditional probability formula $\pi(y|x) = \pi(x, y)/\pi(x)$. In this sense, a classical joint probability distribution acts as both a joint state in space and time, something that does not happen for quantum joint states. In this work we take the state over time interpretation of a classical joint probability distribution and extend this to the quantum case.

The Jordan product is defined as one half of the anti-commutator: given two operators A, B , we denote $A \star B := \frac{1}{2}\{A, B\} = \frac{1}{2}(AB + BA)$ [27, 28, 29, 30]. In the context of transport maps, this operation has several desirable properties, as previously mentioned here and further expanded upon in [20]. Moreover, while the Jordan product is not associative in general, it was shown in [23] to fulfil the associative property for products of matrices of the form $A_{01} \equiv A_{01} \otimes \mathbb{1}_{23}$, $B_{12} \equiv B_{12} \otimes \mathbb{1}_{03}$ and $C_{23} \equiv \mathbb{1}_{01} \otimes C_{23}$: $A_{01} \star (B_{12} \star C_{23}) = (A_{01} \star B_{12}) \star C_{23}$. This form of product is what we encounter in our formalism, see Subsection 2.2 for the details on the associativity.

2.1 States over two times

Firstly, we formally define a state over time, for which we propose the handy abbreviation *stote* (see Fig. 1):

Definition 2.1. Let $\mathcal{H}_A, \mathcal{H}_B$ be finite dimensional Hilbert spaces and $\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)$ the set of Jamiołkowski matrices between these two spaces. Let $\rho \in \mathcal{S}(\mathcal{H}_A)$ and $J \in \mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)$. The associated state over time (*stote*) is defined as $Q = (\rho \otimes \mathbb{1}) \star J$. Typically we will omit the identity and write this as $Q = \rho \star J$. The set of states over time between two Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$ is the set of all operators of this form:

$$\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B) = \{\rho \star J \mid J \in \mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B) \text{ and } \rho \in \mathcal{S}(\mathcal{H}_A)\}. \quad (2.1)$$

This definition has the same interpretation as the classical probability coupling. Initial and final states appear as marginals: $\text{Tr}_B [Q] = \rho$ and $\text{Tr}_A [Q] = \sigma$, and the map connecting one to the other can be reconstructed from Q as well (see later in this section).

In the literature of states over time, the Jamiołkowski matrix has been used instead of the Choi matrix, and we will generally do the same. We denote Choi matrices by C and Jamiołkowski matrices by J , and they are related by $C = J^{TA}$, as seen in Section 1.1.

Similar to the classical case, we are concerned with the set of channels that map a given state to another given state. We formalise this in the following definition.

Definition 2.2. Let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A), \mathcal{S}(\mathcal{H}_B)$ and $\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B) = \{J \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), J^{TA} \geq 0, \text{Tr}_B [J] = \mathbb{1}\}$ be the set of Jamiołkowski matrices. The set of states over time between ρ and σ is

$$\begin{aligned} \mathcal{Q}(\rho, \sigma) &= \{\rho \star J \mid J \in \mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B) \text{ and } \text{Tr}_A [\rho J] = \sigma\} \\ &= \{Q \in \mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B) \mid \text{Tr}_B [Q] = \rho, \text{Tr}_A [Q] = \sigma\}. \end{aligned} \quad (2.2)$$



Figure 1: Stoat (also stote in old spelling), *Mustela erminea*; not to be confused with the common weasel, *Mustela nivalis*, or other species of weasels, ferrets, minks or martens.

The last condition, $\text{Tr}_A[Q] = \sigma$, specifies the image of ρ under the associated quantum channel. That is, we are selecting the channels (in matrix form) that map ρ to σ . We can also go in the opposite direction. That is, given a state over time, an associated initial state and Jamiołkowski matrix (and therefore final state) can be found explicitly. This has been also shown in [24, 26].

Theorem 2.3. *Let ω be a Hermitian operator on $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\rho = \text{Tr}_B[\omega] \geq 0$. Then let $B = \{|ik\rangle\}$ be a product basis of $\mathcal{H}_A \otimes \mathcal{H}_B$ such that $\{|i\rangle\}$ is a diagonal basis of ρ with associated eigenvalues $\{p_i \geq 0\}$. Finally, consider an operator³ J such that, if p_i or p_j are nonzero then*

$$\langle ik|J|j\ell\rangle = \frac{2}{p_i + p_j} \langle ik|\omega|j\ell\rangle. \quad (2.3)$$

Then, $\omega = \rho \star J$. Moreover, if ρ is faithful then J is Hermitian and the unique operator such that $\omega = \rho \star J$ for the given ω .

Proof. This is immediate from the construction of $\rho \star J$ in a product basis that contains an eigenbasis of ρ . Let J be written in a product basis $\{|ik\rangle\}$ whose $\mathcal{H}_A$ component is a diagonal basis of ρ ,

$$\rho = \sum_i p_i |i\rangle\langle i| \quad (2.4)$$

$$J = \sum_{ikj\ell} \langle ik|J|j\ell\rangle |ik\rangle\langle j\ell|. \quad (2.5)$$

In this basis we can calculate the Jordan product

$$\begin{aligned} \rho \star J &= \frac{1}{2} \left(\sum_{ikj\ell i'} p_{i'} |i'\rangle\langle i'| \langle ik|J|j\ell\rangle |ik\rangle\langle j\ell| + \langle ik|J|j\ell\rangle |ik\rangle\langle j\ell| p_i |i\rangle\langle i| \right) \\ &= \sum_{ikj\ell} \frac{1}{2} (p_i + p_j) \langle ik|J|j\ell\rangle |ik\rangle\langle j\ell|. \end{aligned} \quad (2.6)$$

Therefore, if p_i or p_j are nonzero, the coefficients of J from this matrix are

$$\langle ik|J|j\ell\rangle = \frac{2}{p_i + p_j} \langle ik|\rho \star J|j\ell\rangle. \quad (2.7)$$

If ρ is faithful, this fully characterises every coefficient of J in the chosen basis, and therefore J is unique. $\square$

³Note we are looking for a Jamiołkowski operator here associated to a CPTP, but the theorem works regardless. The check is performed later, depending on whether ρ is faithful or not.

If ρ is faithful, from Theorem 2.3 we can check if ρ and J are a state and a Jamiołkowski matrix, respectively, to conclude whether or not ω is a state over time. Note that the trace condition is guaranteed:

$$\begin{aligned}\mathrm{Tr}_B[J] &= \sum_{ikjl} \langle ik|J|j\ell\rangle |ik\rangle\langle j\ell| = \sum_{ikjl} \frac{2}{p_i + p_j} \langle ik|\omega|j\ell\rangle |i\rangle\langle j| \langle \ell|k\rangle \\ &= \sum_{ij} \frac{2}{p_i + p_j} \langle i| \left(\sum_k \langle k|\omega|k\rangle \right) |j\rangle |i\rangle\langle j| = \sum_{ij} \frac{2}{p_i + p_j} \langle i|\mathrm{Tr}_B[\omega]|j\rangle |i\rangle\langle j| \\ &= \sum_{ij} \frac{2}{p_i + p_j} \langle i|\rho|j\rangle |i\rangle\langle j| = \sum_{ij} \frac{2}{p_i + p_j} \delta_{ij} p_i |i\rangle\langle j| = \mathbb{1},\end{aligned}\quad (2.8)$$

where we have used that $\rho = \mathrm{Tr}[\omega]$ and its faithfulness.

In the case where ρ is not faithful we end up with a matrix completion problem that can be solved with the following semidefinite programming (SDP) problem [31, 32]

$$\begin{aligned}\min_J \quad & f(J) \\ \text{s.t.} \quad & \begin{cases} \langle ik|J|j\ell\rangle = \frac{2}{p_i + p_j} \langle ik|\omega|j\ell\rangle, & \forall i, j \in B \mid p_i + p_j \neq 0, \forall k, \ell \in B \\ \mathrm{Tr}_B J = \mathbb{1} \\ J^{T_A} \geq 0 \end{cases}\end{aligned}\quad (2.9)$$

Here, f is any linear function, since we are not interested in minimising a specific function but just in finding a matrix that fulfils the given conditions (a feasible solution). For numerical calculation purposes, we can rewrite this feasibility problem by adding an extra real variable x . This is useful because numerical solvers require the feasible set to have a nonempty interior. In some cases (like when ρ is faithful) the set of feasible Jamiołkowski matrices can have an empty interior and adding the dummy variable x allows us to expand the feasible set. x is added as follows:

$$\begin{aligned}\min_{(x,J)} \quad & -x \\ \text{s.t.} \quad & \begin{cases} \langle ik|J|j\ell\rangle = \frac{2}{p_i + p_j} \langle ik|\omega|j\ell\rangle, & \forall i, j \in B \mid p_i + p_j \neq 0, \forall k, \ell \in B \\ \mathrm{Tr}_B J = \mathbb{1} \\ J^{T_A} \geq x\mathbb{1} \end{cases}\end{aligned}\quad (2.10)$$

From this it is clear that if the output of the SDP is a nonnegative x then the associated matrix J will be a Jamiołkowski matrix.

In case that ρ is faithful, the following basis independent expression from [33] can also be used:

$$J = \int_0^\infty e^{-\frac{t}{2}\rho} \omega e^{-\frac{t}{2}\rho} dt. \quad (2.11)$$

As a corollary of this form we can see that this map is CP:

Corollary 2.4. *Let $\rho \in \mathcal{S}(\mathcal{H})$ be a faithful state of a Hilbert space $\mathcal{H}$. Then the map*

$$x \mapsto \int_0^\infty e^{-\frac{t}{2}\rho} x e^{-\frac{t}{2}\rho} dt \quad (2.12)$$

is completely positive.

Proof. $e^{-\frac{t}{2}\rho}$ will be Hermitian because ρ is. For a fixed t , let $K = K^\dagger = e^{-\frac{t}{2}\rho}$ the map associated to this fixed t , $x \mapsto KxK^\dagger$, will be CP because it is written in Kraus form [34, Ch. 8], [35]. The integral of CP maps will be CP, thus the original map is CP. $\square$

Still for a faithful state ρ , starting from the canonical basis, the following expression is also equivalent and will be useful later:

$$J = (U_\rho \otimes \mathbb{1}) \left(U_\rho^* \rho U_\rho \star \left((U_\rho^* \otimes \mathbb{1}) \omega (U_\rho \otimes \mathbb{1}) \right)^\Theta \right)^\Theta (U_\rho^* \otimes \mathbb{1}), \quad (2.13)$$

where U_ρ is a unitary that diagonalises ρ from the canonical basis and Θ symbolises the Hadamard (entry-wise) inverse. To show it is equal we need to see that the equation yields the correct coefficients. First, we can remove the enveloping $(U_\rho \otimes \mathbb{1}) \cdot (U_\rho^* \otimes \mathbb{1})$ and work in the diagonal basis of ρ , as done in Theorem 2.3. Then, note that $(U_\rho^* \otimes \mathbb{1}) \omega (U_\rho \otimes \mathbb{1})$ is just ω written in the diagonal basis of ρ in the first subsystem and the canonical basis in the second, that is

$$\left((U_\rho^* \otimes \mathbb{1}) \omega (U_\rho \otimes \mathbb{1}) \right)_{ikj\ell} = \langle ik | \omega | j\ell \rangle. \quad (2.14)$$

We then invert it element-wise and multiply by half the diagonal element of $|i\rangle$ and $|j\rangle$, that is

$$\left(U_\rho^* \rho U_\rho \star \left((U_\rho^* \otimes \mathbb{1}) \omega (U_\rho \otimes \mathbb{1}) \right)^\Theta \right)_{ikj\ell} = \frac{1}{2} \frac{(p_i + p_j)}{\langle ik | \omega | j\ell \rangle}. \quad (2.15)$$

Now, we only need to invert it element-wise.

Finally, we want to point out that Theorem 2.3 recovers Bayes' Theorem in the classical case:

Remark 2.5. Let $\rho = \sum_i p_i |i\rangle\langle i|$ and $J_{A \rightarrow B} = \sum_{ij} M_{i \rightarrow j} |ij\rangle\langle ij|$, where $M_{i \rightarrow j}$ is a classical stochastic map. Then $Q = \sum_{ij} M_{i \rightarrow j} p_i |ij\rangle\langle ij| = \sum_{ij} r_{ij} |ij\rangle\langle ij|$, where $r_{ij} = M_{i \rightarrow j} p_i$ is a joint probability distribution. We can now apply Theorem 2.3 considering B the input space. The partial trace will be $\text{Tr}_A [Q] = \sum_j (\sum_i r_{ij}) |j\rangle\langle j| = \sum_j q_j |j\rangle\langle j|$. Q is already diagonal in a product basis of the required form so we can directly find $J_{A \leftarrow B} = \sum_{ij} r_{ij}/q_j |ij\rangle\langle ij| = \sum_{ij} N_{i \leftarrow j} |ij\rangle\langle ij|$. Joining everything together we recover Bayes' Theorem: $r_{ij} = N_{i \leftarrow j} q_j = M_{i \rightarrow j} p_i$.

Note that, similarly to Theorem 2.3, in the case where $q_j = 0$ $N_{i \leftarrow j}$ remains undefined. In the classical case it is possible to choose any values for $N_{i \leftarrow j}$ such that N is a stochastic map since these coefficients do not affect the system, as $N_{i \leftarrow j} q_j = 0$ regardless of the chosen value.

We will also be interested in the cone generated by a set of states over time, $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$. A cone is a subset of a real vector space that is closed under product by positive constant and convex combination. We consider the cone and not the convex hull because we are interested in studying the positivity of the quantum transport cost, and the cone gives us a better framework to study it later. The normalisation condition is linear and therefore much easier to deal with.

Definition 2.6. Let $\mathcal{H}_A, \mathcal{H}_B$ be finite dimensional Hilbert spaces and $\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)$ the set of Jamiołkowski matrices between these two spaces. The cone of states over time is

$$\begin{aligned} \hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B) &= \text{cone}(\{\rho \star J \mid J \in \mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B) \text{ and } \rho \in S(\mathcal{H}_A)\}) \\ &= \text{cone}(\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)) = \text{cone} \left(\bigcup_{\substack{\rho \in S(\mathcal{H}_A) \\ \sigma \in S(\mathcal{H}_B)}} \mathcal{Q}(\rho, \sigma) \right). \end{aligned} \quad (2.16)$$

2.2 States over multiple times

If instead of a single channel we have n channels in sequence we can write the space of states over time associated to this process as a recursive hierarchy:

Definition 2.7. Consider Hilbert spaces $\mathcal{H}_i$, $i \geq 0$. Then

$$\begin{aligned} \mathcal{Q}(\mathcal{H}_0 : \dots : \mathcal{H}_n) &= (\mathcal{Q}(\mathcal{H}_0 : \dots : \mathcal{H}_{n-1}) \otimes \mathbb{1}_n) \star \left(\bigotimes_{j=0}^{n-2} \mathbb{1}_j \otimes \mathcal{J}(\mathcal{H}_{n-1} \rightarrow \mathcal{H}_n) \right) \\ &\equiv \mathcal{Q}(\mathcal{H}_0 : \dots : \mathcal{H}_{n-1}) \star \mathcal{J}(\mathcal{H}_{n-1} \rightarrow \mathcal{H}_n). \end{aligned} \quad (2.17)$$

By taking partial traces on specific subsystems we can ‘forget’ about the state of the system at that slot, that is that $\text{Tr}_i [\mathcal{Q}(\mathcal{H}_0 : \dots : \mathcal{H}_n)] = \mathcal{Q}(\mathcal{H}_0 : \dots : \mathcal{H}_{i-1} : \mathcal{H}_{i+1} : \dots : \mathcal{H}_n)$. This is true because given two Jamiołkowski matrices $J_{i-1,i}$, $J_{i,i+1}$ we can construct a Jamiołkowski matrix $\tilde{J}_{i-1,i+1} = \text{Tr}_i [(J_{i-1,i} \otimes \mathbb{1}_{i+1}) \star (\mathbb{1}_{i-1} \otimes J_{i,i+1})]$ such that the associated channels fulfil $\tilde{\mathcal{E}}_{i-1,i+1} = \mathcal{E}_{i,i+1} \circ \mathcal{E}_{i-1,i}$ [36].

Given a state over n times, we can also reconstruct the initial and all instantaneous states of the system, as well as the CPTP maps linking different times. Indeed, by tracing out every subsystem except a consecutive pair, we can reduce the problem to the two-time scenario and use Theorem 2.3.

3 Quantum transport cost and optimal transport

Before introducing the optimal transport cost between two states, it is useful to define the cost of transporting a single state through a given quantum channel.

Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces, let $\rho \in \mathcal{S}(\mathcal{H}_A)$ be a quantum state, and let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B)$ be a quantum channel. For a given cost matrix $K \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we define the *quantum transport cost* of ρ under $\mathcal{E}$ as

$$\kappa(\rho, \mathcal{E}) = \text{Tr} [K Q_{\rho, \mathcal{E}}], \quad (3.1)$$

where $Q_{\rho, \mathcal{E}} = \rho \star J_{\mathcal{E}}$ denotes the state over time (or *coupling*) associated with sending ρ through the channel $\mathcal{E}$, with Jamiołkowski operator $J_{\mathcal{E}}$.

This quantity measures the expected cost of transporting the state ρ from system A to system B when the process is fixed to be $\mathcal{E}$. In general, there may exist many channels $\mathcal{E}$ that transform ρ into a target state σ . The *quantum optimal transport cost* identifies, among all such admissible transformations, one that achieves the smallest possible cost:

Definition 3.1. Let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_{A,B})$ and $K \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ a Hermitian operator. The *quantum transport cost with cost matrix K between ρ and σ* is

$$\mathcal{K}(\rho, \sigma) = \min_{\mathcal{E} \mid Q_{\rho, \mathcal{E}} \in \mathcal{Q}(\rho, \sigma)} \kappa(\rho, \mathcal{E}) = \min_{Q \in \mathcal{Q}(\rho, \sigma)} \text{Tr}[KQ]. \quad (3.2)$$

Sometimes, when it is clear from the context, we refer to this quantity as just the cost, with the implication that it is quantum, optimal and transport based. Importantly, this definition explicitly depends on the choice of cost matrix K , which serves as the analogue of the cost function in classical optimal transport. The relevance of $\mathcal{K}(\rho, \sigma)$ and the properties it inherits are determined by this choice.

As we will discuss later in Section 4, the choice of cost matrix K determines the properties of the transport cost. In particular, we are interested in the properties that define a distance (cost 0 for identical inputs, positivity, triangle inequality and symmetry) as well as some generically useful properties in quantum information (additivity, joint convexity). Ideally, we would be able to find cost matrices K with good properties, notably the weaker ones: cost 0 for identical inputs and positivity, that allows us to encode some physical property we are interested in considering in a distinguishability measure. In this regard, transport costs differ from traditional distinguishability measures that rely purely on the vector space structure of the states.

Note that the quantum optimal transport cost can be calculated with an SDP [37, 32]. Let, again, K be some cost matrix and ρ and σ states. Then the SDP associated to finding the quantum optimal transport cost between ρ and σ with cost matrix K is

$$\begin{aligned} \min_J \quad & \text{Tr} [(K \star \rho) J] \\ \text{s.t.} \quad & \begin{cases} \text{Tr}_B J = \mathbb{1} \\ \text{Tr}_A [\rho J] = \sigma \\ J^{T_A} \geq 0 \end{cases} \end{aligned} \quad (3.3)$$

This form can provide some problems computationally due to the partial transpose in the positivity condition. To move these partial transposes to the constants of the problem, we present the SDP with the optimisation over the channel being in Choi form

$$\begin{aligned} \min_C \quad & \text{Tr}[(K \star \rho)^{T_A} C] \\ \text{s.t.} \quad & \begin{cases} \text{Tr}_B C = \mathbb{1} \\ \text{Tr}_A [\rho^T C] = \sigma \\ C \geq 0 \end{cases} \end{aligned} \quad (3.4)$$

and its dual

$$\begin{aligned} \max_{Y_1, Y_2} \quad & \text{Tr}[Y_1] + \text{Tr}[\sigma Y_2] \\ \text{s.t.} \quad & \begin{cases} Y_1 \otimes \mathbb{1} + \rho^T \otimes Y_2 \leq (K \star \rho)^{T_A} \\ Y_1, Y_2 \text{ Hermitian} \end{cases} \end{aligned} \quad (3.5)$$

The formulation of the primal in Choi form allows us to see the relation between the primal and dual problems. We use the notation of [32]. In this notation, the primal SDP is given by the triple $(\Phi(X), A, B) = ((\text{Tr}_B[X], \text{Tr}_A[\rho^T X]), -(K \star \rho)^{T_A}, (\mathbb{1}, \sigma))$. Then $\Phi^*((Y_1, Y_2)) = Y_1 \otimes \mathbb{1} + \rho^T \otimes Y_2$.

The primal expression of the SDP further shows the connection between the coupling and the channel. In fact, the coupling is only implicitly in the SDP through $\text{Tr}[(K \star \rho)J] = \text{Tr}[K(\rho \star J)] = \text{Tr}[KQ]$. We use the Jamiołkowski matrix in the SDP instead of the coupling in the SDP because it is unclear how the couplings can be characterised through semidefinite expressions.

4 Properties

In classical probabilistic settings, it is easy and of supreme interest to use transport costs to define metrics on probability distributions. Recall that a metric on a set Ω is a function $D : \Omega \times \Omega \rightarrow \mathbb{R}$ (the distance) that satisfies the following properties for all $x, y, z \in \Omega$:

$$D(x, y) \geq 0 \quad (4.1)$$

$$D(x, y) = D(y, x), \quad (4.2)$$

$$D(x, y) = 0 \text{ if and only if } x = y, \quad (4.3)$$

$$D(x, z) \leq D(x, y) + D(y, z). \quad (4.4)$$

Note that not all properties described here are equal. While symmetry and the triangle inequality can be nice, a function without these properties can still be a divergence, a very important and useful concept in information theory [38, 39]. Thus, finding conditions of K for which the associated quantum optimal transport cost fulfils these properties is an important question.

The main goal of this section is to study sufficient conditions on K for which the associated cost has desirable properties. This is the main theoretical framework and general results of our approach to quantum optimal transport. A better characterisation of which cost matrices induce certain properties in the associated cost would allow us to define physically relevant costs with desirable properties. Additionally, we examine general mathematical properties of interest in this context, namely convexity properties of the cost under ensemble input states as well as additivity/multiplicity under tensor products.

In the next section, we will use these results to focus on particular choices of cost functions under further physically relevant constraints, namely unitary invariance of the cost.

4.1 Cost of identity

For a quantum transport cost to be reasonable we want that the cost associated with the identity channel is 0, regardless of the input state. We could ask that there exists a state over time such that the cost is 0 if $\rho = \sigma$, like in Eq. (4.4), but physically it makes sense to ask that doing nothing results in zero cost. The following result characterises the cost matrices that fulfil this property.

Proposition 4.1. *Given a finite dimensional Hilbert space $\mathcal{H}$, a cost matrix K assigns cost 0 to the identity map (with any input) if and only if*

$$\text{Tr}_B [\mathcal{S} \star K] = 0, \quad (4.5)$$

where $\mathcal{S}$ is the swap operator.

Proof. The Jamiołkowski operator associated to the identity channel is the swap operator $\mathcal{S}$, clearly from Section 1.1: $J_{\text{id}} = (\text{id} \otimes \text{id})(\mathcal{S}) = \mathcal{S}$. Now, let K be a matrix such that $\kappa(\rho, \text{id}) = \text{Tr}[(\rho \star \mathcal{S})K] = 0 \forall \rho \geq 0$. We can transform the left-hand side in the following way, using the definition of the Jordan product, the cyclic property of the trace and properties of partial traces:

$$\text{Tr}[(\rho \otimes \mathbb{1}) \star \mathcal{S})K] = \frac{1}{2} \text{Tr}[(\rho \otimes \mathbb{1})\mathcal{S} + \mathcal{S}(\rho \otimes \mathbb{1})K] = \frac{1}{2} \text{Tr}[(\rho \otimes \mathbb{1})(\mathcal{S}K + K\mathcal{S})] \quad (4.6)$$

$$= \text{Tr}[\rho \text{Tr}_B [\mathcal{S} \star K]] = 0. \quad (4.7)$$

The set of positive matrices generates the whole space [40, 41], therefore this is equivalent to saying that the Hilbert-Schmidt inner product of $\text{Tr}_B [\mathcal{S} \star K]$ with all other elements is 0. Therefore $\text{Tr}_B [\mathcal{S} \star K] = 0$. The converse is immediate, so the proof is done. $\square$

4.2 Positivity

A natural requirement for a cost matrix K is that the associated quantum optimal transport cost be nonnegative, *i.e.* $\mathcal{K}(\rho, \sigma) \geq 0$ for all $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A, \mathcal{H}_B)$. Note that in the definition of the cost matrix K is always Hermitian. Because K and the states over time $Q \in \mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ are Hermitian the cost is a real number. Moreover, since $\text{Tr}[KQ] = \langle Q, K \rangle_{HS}$, where $\langle \cdot, \cdot \rangle_{HS}$ denotes the Hilbert-Schmidt inner product, the k for which the cost is positive is the dual cone to $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ with respect to the Hilbert-Schmidt inner product. We have no closed-form characterisation of this dual cone, but we can provide some partial results. Throughout this section, whenever we refer to duality, we mean it with respect to the Hilbert-Schmidt inner product.

We will be working with the cone generated by the set of states over time: $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B) = \text{cone}(\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B))$, as seen in Definition 2.6.

Remark 4.2. The cone can also be obtained by adding an ancillary system R and considering the CPTP maps that take inputs in the composite system AR . More precisely, consider finite dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_R$ and the following: let

$$\rho = \sum_t p_t \rho_t \in \mathcal{S}(\mathcal{H}_A), \quad \{\mathcal{E}_t^{A \rightarrow B}\}, \quad (4.8)$$

where $\mathcal{E}_t^{A \rightarrow B}$ are quantum channels from $\mathcal{H}_A$ to $\mathcal{H}_B$. Now consider the extension to a conditional quantum channel and the following extended state

$$\rho^{AR} = \sum_t p_t \rho_t^A \otimes |t\rangle\langle t|^R, \quad \mathcal{E}^{AR \rightarrow B}(X_A \otimes X_R) = \sum_t \mathcal{E}_t^{A \rightarrow B}(X_A) \langle t|X_R|t\rangle, \quad (4.9)$$

where the definition of $\mathcal{E}^{AR \rightarrow B}$ is extended to general elements by linearity. The Jamiołkowski operator associated to $\mathcal{E}^{AR \rightarrow B}$ is $J^{AR \rightarrow B} = \sum_t J_t^{A \rightarrow B} \otimes |t\rangle\langle t|^R$. Now

$$J^{AR \rightarrow B} \star \rho^{AR} = \sum_t p_t J_t^{A \rightarrow B} \star \rho_t^A \otimes |t\rangle\langle t|^R, \quad (4.10)$$

and the partial trace (removing R) of this state over time is

$$\mathrm{Tr}_R \left[J^{AR \rightarrow B} \star \rho^{AR} \right] = \sum_t p_t J_t^{A \rightarrow B} \star \rho_t^A, \quad (4.11)$$

which is an arbitrary convex combination of states over time on $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$.

First, we want to characterise the dual cone to the set of Jamiołkowski matrices $\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)^*$. For this, we need some technical results about convex cones, starting with the precise definition of dual cone in inner product vector spaces.

Definition 4.3. *Let $X \subseteq V$ be a subset of a real inner product vector space V . The dual cone to X with respect to the inner product is a cone in V , denoted by X^* , and defined as*

$$X^* = \{v \in V \mid \langle v|x \rangle \geq 0 \forall x \in X\}. \quad (4.12)$$

In the context of this work, V is the Hermitian subset of $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ equipped with the Hilbert-Schmidt inner product. Note that with no further requirements for X , X^* fulfils the definition of cone.

Lemma 4.4. *Let I be an index set and $\{\mathcal{C}_i\}_{i \in I}$ a set of convex cones. Then*

$$\bigcap_{i \in I} \mathcal{C}_i^* = \left(\sum_{i \in I} \mathcal{C}_i \right)^*. \quad (4.13)$$

Proof. We can show the equality directly. Let $x \in \bigcap_{i \in I} \mathcal{C}_i^*$. Then,

$$\begin{aligned} x \in \mathcal{C}_i^* \forall i \in I &\Leftrightarrow \langle x|c_i \rangle \geq 0 \forall c_i \in \mathcal{C}_i \forall i \in I \Leftrightarrow \sum_{i \in I} \langle x|c_i \rangle \geq 0 \forall c_i \in \mathcal{C}_i \\ &\Leftrightarrow \left\langle x \left| \sum_{i \in I} c_i \right. \right\rangle \geq 0 \forall c_i \in \mathcal{C}_i \Leftrightarrow x \in \left(\sum_{i \in I} \mathcal{C}_i \right)^*. \end{aligned} \quad (4.14)$$

□

Recall that a convex cone $C \subseteq V$ is called *pointed* if for all nonzero $x \in C$, $-x \notin C$.

Lemma 4.5. *A cone C is pointed if and only if there exists an element f in the dual space such that $f(x) > 0$ for all nonzero $x \in C$.*

Proof. Let $x, -x \in C$ and $f \in C^*$ such that $f(x) > 0$ for all nonzero $x \in C$. Because f is linear $f(-x) = -f(x) < 0$, which is a contradiction.

Conversely, let C be pointed, then $(C \setminus \{0\}) \cap ((-C) \setminus \{0\}) = \emptyset$. By the Hahn-Banach Theorem (as formulated in [42, Theorem 7.8.4]), there exists a linear map such that $f(x) > 0$ for all $x \in C \setminus \{0\}$. □

Lemma 4.6. *Consider the convex cone $\mathcal{C}_2 = \{C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \mathrm{Tr}_B[C] \propto_{\mathbb{C}} \mathbb{1}\}$. Its dual is*

$$\mathcal{C}_2^* = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \mathbb{1} \in \mathcal{B}(\mathcal{H}_B), \mathrm{Tr}[A] = 0\}. \quad (4.15)$$

Proof. Let us call this set $\mathcal{A} = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \mathrm{Tr}[A] = 0\}$. The following calculation shows that $\mathcal{A} \subseteq \mathcal{C}_2^*$: let $A \otimes \mathbb{1} \in \mathcal{A}$ and $C \in \mathcal{C}_2$, then:

$$\mathrm{Tr}[(A \otimes \mathbb{1})C] = \mathrm{Tr}_A[\mathrm{Tr}_B[(A \otimes \mathbb{1})C]] = \mathrm{Tr}[A \mathrm{Tr}_B[\mathbb{1}C]] = z \mathrm{Tr}[A\mathbb{1}] = 0. \quad (4.16)$$

To see that they are equal, note that $\mathcal{C}_2$ (and thus the orthogonal $\mathcal{C}_2^*$ [41]) and $\mathcal{A}$ are real subspaces of the real vector space $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. We will calculate the dimension of each and see they are the

same. The real dimension of $\mathcal{A}$ is just the real dimension of $\mathcal{B}(\mathcal{H})$ minus the dimension subtracted by the two real (one complex) linear conditions $\text{Tr}[A] = 0$. That is

$$\dim \mathcal{A} = \dim \mathcal{B}(\mathcal{H}) - 2 = 2d_A^2 - 2. \quad (4.17)$$

To find the dimension of $\mathcal{C}_2^*$, we first find the dimension of $\mathcal{C}_2$. Recall that this set is defined by the condition $\text{Tr}_B[C] \propto_{\mathbb{C}} \mathbb{1}$. This corresponds to $2d_A(d_A - 1)$ equations (real and imaginary parts of non diagonal terms equal to 0) plus $2(d_A - 1)$. That is because the condition is proportionality, not equality, so we first fix the real and imaginary components of the first diagonal element and then every other diagonal element will have to have the same real and imaginary components, for a total of $2(d_A - 1)$. Thus the dimension is

$$\begin{aligned} \dim \mathcal{C}_2 &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A(d_A - 1) - 2d_A + 2 \\ &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A^2 + 2d_A - 2d_A + 2 \\ &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A^2 + 2. \end{aligned} \quad (4.18)$$

The dimension of the orthogonal complement is the dimension of the total space minus this, thus

$$\dim \mathcal{C}_2^* = \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) + 2d_A^2 - 2 = 2d_A^2 - 2. \quad (4.19)$$

Since this two sets $\mathcal{A}$ and $\mathcal{C}_2^*$ are real subspaces of the same dimension and $\mathcal{A} \subseteq \mathcal{C}_2^*$, they are the same:

$$\mathcal{C}_2^* = \mathcal{A} = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \text{Tr}[A] = 0\}. \quad (4.20)$$

□

Lemma 4.7. *The partial transpose map, denoted here by $T_A(\cdot)$, fulfils the following:*

$$\begin{aligned} \text{Tr}_A[T_A(K)C] &= \text{Tr}_A[K T_A(C)] \\ T_A(\text{Tr}_B[K C]) &= \text{Tr}_B[T_A(C) T_A(K)] \quad \forall K, C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B), \rho \in \mathcal{B}(\mathcal{H}_A). \\ T_A((\rho \otimes \mathbb{1})C) &= T_A(C)(\rho^T \otimes \mathbb{1}) \end{aligned} \quad (4.21)$$

Moreover, the partial transpose is self-adjoint with respect to the Hilbert-Schmidt inner product.

Proof. Recall that the transpose is a basis dependent operation. These two properties are trivial to check if we expand the equations in a product basis that includes the basis over which we are transposing. Alternatively, we can use tensor network notation [43] as shown in Fig. 2.

$$\begin{aligned} \text{Tr}[T_A(K)C] &= \text{Tr}[K C] = \text{Tr}[K T_A(C)] \\ T_A(\text{Tr}_B[K C]) &= \text{Tr}_B[T_A(C) T_A(K)] \\ T_A((\rho \otimes \mathbb{1})C) &= T_A(C)(\rho^T \otimes \mathbb{1}) \end{aligned}$$

Figure 2: Proofs of the expressions in Lemma 4.7 using tensor network notation.

To see that the partial transpose is self adjoint, apply the first equation to K^* and take the trace on both sides of the equation:

$$\begin{aligned} \text{Tr}_B \text{Tr}_A[T_A(K^*)C] &= \text{Tr}_B \text{Tr}_A[K^* T_A(C)] \\ &\Leftrightarrow \text{Tr}[T_A(K^*)C] = \text{Tr}[K^* T_A(C)] \\ &\Leftrightarrow \langle T_A(K^*), C \rangle_{HS} = \langle K^*, T_A(C) \rangle_{HS}. \end{aligned} \quad (4.22)$$

□

Lemma 4.8. *Let K be a convex cone and A an invertible linear map. Then,*

$$A(K)^* = (A^*)^{-1}(K^*). \quad (4.23)$$

Proof. Let $x \in A(K)^*$. Then,

$$\begin{aligned} x \in A(K)^* &\Leftrightarrow \langle x, A(y) \rangle \geq 0 \quad \forall y \in K \Leftrightarrow \langle A^*(x), y \rangle \geq 0 \quad \forall y \in K \\ &\Leftrightarrow A^*(x) \in K^* \Leftrightarrow x \in (A^*)^{-1}(K^*). \end{aligned} \quad (4.24)$$

□

Corollary 4.9. *The dual of the cone of Jamiołkowski operators is the partial transpose of the dual cone of the Choi operators, denoted by $\mathcal{C}$. In other words,*

$$\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)^* = T_A(\mathcal{C}(\mathcal{H}_A \rightarrow \mathcal{H}_B))^* = T_A(\mathcal{C}(\mathcal{H}_A \rightarrow \mathcal{H}_B)^*). \quad (4.25)$$

Proof. Note that the partial transpose is self adjoint from Lemma 4.7 and self inverse and apply the previous Lemma 4.8. □

First, we can show using Lemma 4.5 that the cones $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ and $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ are pointed and spanning. A cone $C \subset V$ is spanning if $C + (-C) = V$ [37].

Proposition 4.10. *The cone of states over time $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ and its dual $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ are pointed, spanning cones.*

Proof. First, we show that $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ is pointed and spanning. By definition of the elements $Q \in \hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$, $\text{Tr}[\mathbb{1}Q] = \text{Tr}[\rho \star J] = \text{Tr}[\rho] = 1 > 0$. By Lemma 4.5, $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ is pointed. The cone $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ is spanning because the set of product states $\{\rho \otimes \sigma | \rho, \sigma \in S(\mathcal{H})\}$ is contained in $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$. That is because the Jamiołkowski matrix of the replacement channel is $\mathbb{1} \otimes \sigma$. This set is spanning so as its superset $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ is also spanning.

The properties of pointed and spanning are such that if the primal cone has one, the dual has the other [37]. As we just showed that $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ is pointed and spanning, its dual is also pointed and spanning. □

We are mostly interested in the fact that $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ is spanning from Proposition 4.10. This shows that our search for the cone of cost matrices with positive associated costs is not futile since the set of matrices with this property is spanning.

Proposition 4.11. *Let $\mathcal{C} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be the minimal cone that contains the Choi matrices.⁴ Then,*

$$\begin{aligned} \mathcal{C}^* &= \overline{\mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B) + \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \text{Tr}[A] = 0\}} \\ &= \overline{\{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \text{Tr}[A] = 0\}} \\ &= \{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \text{Tr}[A] = 0\}. \end{aligned} \quad (4.26)$$

Proof. Consider the following:

$$\mathcal{C}_1 = \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (4.27)$$

$$\mathcal{C}_2 = \{C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \text{Tr}_B[C] \propto_{\mathbb{C}} \mathbb{1}\}. \quad (4.28)$$

These two are closed cones and

$$\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2. \quad (4.29)$$

⁴Through the Choi isomorphism this would correspond to CP and trace *scaling* (by a real positive constant, instead of trace preserving) maps.

Moreover (the cone of psd matrices is self dual [41] and Lemma 4.6):

$$\mathcal{C}_1^* = \mathcal{C}_1 = \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \quad (4.30)$$

$$\mathcal{C}_2^* = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \text{Tr}[A] = 0\}. \quad (4.31)$$

Now, we can use Lemma 4.4, setting $I = \{1, 2\}$ and the duals in the theorem, to find the dual of $\mathcal{C}$:

$$\begin{aligned} \mathcal{C}^* &= (\mathcal{C}_1 \cap \mathcal{C}_2)^* = (\overline{\mathcal{C}_1} \cap \overline{\mathcal{C}_2})^* = \overline{\mathcal{C}_1^* + \mathcal{C}_2^*} \\ &= \overline{\mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B) + \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \text{Tr}[A] = 0\}}, \end{aligned} \quad (4.32)$$

where we used $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$ first; the closeness of $\mathcal{C}_1$ and $\mathcal{C}_2$ second, then Lemma 4.4; and finally the duals of $\mathcal{C}_1$ and $\mathcal{C}_2$. To see that the set

$$\{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \text{Tr}[A] = 0\} \quad (4.33)$$

is closed, note that both $\mathcal{C}_1^*$ and $\mathcal{C}_2^*$ are closed due to being dual cones. Moreover, let $\omega \in \mathcal{C}_1$ and $A \otimes \mathbb{1} \in \mathcal{C}_2$ have norm 1 (in the norm induced by the inner product). First note that the minimum eigenvalue of A , is lower bounded because $1 = \|A \otimes \mathbb{1}\| = \|A\| \|I\| = \sqrt{d} \|A\|$ and $\frac{1}{\sqrt{d}} = \|A\| = \sqrt{\sum_i \lambda_i^2}$. Due to $\text{Tr}[A] = 0$, the smallest possible eigenvalue of A , $\lambda_{\min}$ appears when A has two eigenvalues, $\lambda_{\min}$ and $-\lambda_{\min}$. Then

$$\frac{1}{\sqrt{d}} = \|A\| = \sqrt{\sum_i \lambda_i^2} = \sqrt{2\lambda_{\min}^2}, \quad (4.34)$$

so that $\lambda_{\min} = -\frac{1}{\sqrt{2d}}$. Now we can calculate the lower bound for the norm of the sum of these two operators:

$$\begin{aligned} \|w + A \otimes \mathbb{1}\| &= \sqrt{\text{Tr}\{[(w + A \otimes \mathbb{1})^* (w + A \otimes \mathbb{1})]\}} = \sqrt{\text{Tr}[\omega^2] + \text{Tr}[|A|^2 \otimes \mathbb{1}] + 2 \text{Tr}[\omega (A \otimes \mathbb{1})]} \\ &= \sqrt{2 + 2 \text{Tr}[\text{Tr}_B[\omega] (A \otimes \mathbb{1})]} \geq \sqrt{2 - 2\lambda_{\min}} = \sqrt{2 \frac{\sqrt{2d} - 1}{\sqrt{2d}}} > 0. \end{aligned} \quad (4.35)$$

By [44], the sum is closed, finishing the proof. $\square$

For the following theorem we use the notation D_s , where $s \in \mathbb{R}_+^n$ is a sequence on positive numbers, to indicate the diagonal matrix with s in its diagonal.

Theorem 4.12. *The dual to the set of states over time for finite dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ can be expressed as*

$$\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^* = \bigcap_{U \in \mathcal{U}(\mathcal{H}_A)} (U \otimes \mathbb{1}) \left(\bigcap_{s \in \mathbb{R}_+^{d_A}} \varphi_{D_s}^{-1}(\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)^*) \right) (U^* \otimes \mathbb{1}), \quad (4.36)$$

where $\varphi_\rho(X) = \rho \star X$ and $\mathcal{J}(\mathcal{H}_A \rightarrow \mathcal{H}_B)^*$ is the dual to the set of Jamiołkowski matrices.

Proof. For simplicity, we ignore the specific Hilbert space dependencies. We want to study $\hat{\mathcal{Q}}^* = \left(\sum_\rho (\rho \star \mathcal{J})\right)^*$. Given a non faithful ρ and a fixed $J \in \mathcal{J}$, we can arbitrarily approximate $\rho \star J$ with faithful ρ_ε as

$$\left(\frac{\rho + \varepsilon \mathbb{1}}{\text{Tr}[\rho + \varepsilon \mathbb{1}]} \right) \star J, \quad (4.37)$$

$\varepsilon > 0$. Let K be such that $\text{Tr}\{K(\rho \star J)\} < 0$. Then, by continuity, $\exists \varepsilon_0 > 0$ such that $\text{Tr}\{K(\rho_\varepsilon \star J)\} < 0$ for $0 < \varepsilon < \varepsilon_0$.⁵ Thus we can consider only faithful ρ in the definition of $\mathcal{Q}^*$. This allows us to take the inverse without problems.

Start with the definition of $\hat{\mathcal{Q}}$, then apply Lemma 4.4 and Lemma 4.8:

$$\hat{\mathcal{Q}}^* = \left(\sum_{\rho} (\rho \star \mathcal{J}) \right)^* = \left(\sum_{\rho} \varphi_{\rho}(\mathcal{J}) \right)^* = \bigcap_{\rho} \varphi_{\rho}(\mathcal{J})^* = \bigcap_{\rho} \varphi_{\rho}^{-1}(\mathcal{J}^*). \quad (4.38)$$

Note that we can use Lemma 4.8 because for a fixed ρ , φ_{ρ} is self dual and has linear inverse, as can be seen from the statement of the inverse in Theorem 2.3. From here, realise that choosing a state ρ is equivalent to choosing a spectrum and a basis or, equivalently, a spectrum $s \in \mathbb{R}_+^n$ and a unitary of $\mathcal{U}(n)$; such that $\rho = U_{\rho} D_{s_{\rho}} U_{\rho}^*$. Moreover, $\mathcal{J}^*$ is invariant under local unitaries, thus

$$\begin{aligned} \varphi_{\rho}^{-1}(\mathcal{J}^*) &= (U_{\rho} \otimes \mathbb{1}) \left(U_{\rho}^* \rho U_{\rho} \star \left((U_{\rho}^* \otimes \mathbb{1}) \mathcal{J}^* (U_{\rho} \otimes \mathbb{1}) \right)^{\Theta} \right)^{\Theta} (U_{\rho}^* \otimes \mathbb{1}) \\ &= (U_{\rho} \otimes \mathbb{1}) \left(D_{s_{\rho}} \star (\mathcal{J}^*)^{\Theta} \right)^{\Theta} (U_{\rho}^* \otimes \mathbb{1}) = (U_{\rho} \otimes \mathbb{1}) \varphi_{D_{s_{\rho}}}^{-1}(\mathcal{J}^*) (U_{\rho}^* \otimes \mathbb{1}). \end{aligned} \quad (4.39)$$

And we can insert this result into the expression of $\hat{\mathcal{Q}}^*$ to obtain that

$$\begin{aligned} \hat{\mathcal{Q}}^* &= \bigcap_{\rho} \varphi_{\rho}^{-1}(\mathcal{J}^*) = \bigcap_{U \in U(\mathcal{H}_A)} \bigcap_{s \in \mathbb{R}_+^{d_A}} (U \otimes \mathbb{1}) \varphi_{D_s}^{-1}(\mathcal{J}^*) (U^* \otimes \mathbb{1}) \\ &= \bigcap_{U \in U(\mathcal{H}_A)} (U \otimes \mathbb{1}) \left(\bigcap_{s \in \mathbb{R}_+^{d_A}} \varphi_{D_s}^{-1}(\mathcal{J}^*) \right) (U^* \otimes \mathbb{1}), \end{aligned} \quad (4.40)$$

which is the local unitarily invariant subset of $\bigcap_{s \in \mathbb{R}_+^{d_A}} \varphi_{D_s}^{-1}(\mathcal{J}^*)$. $\square$

Theorem 4.12 provides a partial characterisation of the dual to the set of states over time as the local unitary invariant subset of the intersection of the images of $\mathcal{J}^*$ with the inverse of the Jordan product with diagonal states (in a chosen basis). Even though this inverse was discussed in Theorem 2.3, we have not been able to use it to obtain a complete and succinct characterization of the dual cone.

Finally, we can assume both non-negativity of the cost and zero cost for the identity channel to obtain the following results in the case where $\mathcal{H}_A = \mathcal{H}_B$:

Theorem 4.13. *Let $\mathcal{H}$ be a finite dimensional Hilbert space. Then $K \in \mathcal{J}(\mathcal{H} \rightarrow \mathcal{H})^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star C] = 0\}$ if and only if*

$$K = T_A(\omega) - (\text{Tr}_B[\mathcal{S} \star T_A(\omega)] \otimes \mathbb{1}), \quad \omega \geq 0, \quad \omega \perp |\Phi_+\rangle\langle\Phi_+|. \quad (4.41)$$

In the previous theorem, K is a matrix that is dual to the Jamiołkowski matrices and generates cost 0 for the identity (see Proposition 4.1, Proposition 4.11 and Lemma 4.7).

Proof. Similarly to before, we ignore the Hilbert space dependencies for the proof. Note that even though the identity $T(\mathcal{C}_1 \cap \mathcal{C}_2) = T(\mathcal{C}_1) \cap T(\mathcal{C}_2)$ for a linear map T and convex cones $\mathcal{C}_1, \mathcal{C}_2$, is false in general, it is true for the partial transpose because that is an invertible map. Thus we can transform

⁵Note this argument is equivalent to removing part of the boundary of the set $\hat{\mathcal{Q}}$ and then noting that the dual cone is always the same for a set and its interior.

the target set as follows:

$$\begin{aligned}
& \mathcal{J}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star C] = 0\} \\
&= T_A(T_A(\mathcal{J}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star C] = 0\})) \\
&= T_A(T_A(\mathcal{J}^*) \cap T_A(\{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star C] = 0\})) \\
&= T_A(\mathcal{C}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star T_A(C)] = 0\}) \\
&= T_A(\mathcal{C}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid T_A(\text{Tr}_B[\mathcal{S} \star T_A(C)]) = T_A(0)\}) \\
&= T_A(\mathcal{C}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[T_A(\mathcal{S}) \star C] = 0\}) \\
&= T_A(\mathcal{C}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star C] = 0\}),
\end{aligned} \tag{4.42}$$

where we used Lemma 4.7.

Now consider an element of $\mathcal{C}^*$, that is a $K = \omega + A \otimes \mathbb{1}$, where $\omega \geq 0$ and $\text{Tr}[A] = 0$. We can now plug this expression in the equation that defines the other set of the intersection:

$$0 = \text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star K] = \text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star (\omega + A \otimes \mathbb{1})] = \text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star \omega] + A, \tag{4.43}$$

thus $A = -\text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star \omega]$. Moreover if we take the trace of this expression, since $\text{Tr}[A] = 0$, we find that $\langle\Phi_+|\omega|\Phi_+\rangle = 0$, *i.e.* $\omega \perp |\Phi_+\rangle\langle\Phi_+|$. Now, the initial set is the set defined by the partial transpose of this elements, that is

$$\begin{aligned}
& \mathcal{J}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \text{Tr}_B[\mathcal{S} \star C] = 0\} \\
&= T_A(\{\omega - \text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star \omega] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0, \omega \perp |\Phi_+\rangle\langle\Phi_+|\}) \\
&= \{T_A(\omega) - T_A(\text{Tr}_B[|\Phi_+\rangle\langle\Phi_+| \star \omega] \otimes \mathbb{1}) \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0, \omega \perp |\Phi_+\rangle\langle\Phi_+|\} \\
&= \{T_A(\omega) - \text{Tr}_B[T_A(|\Phi_+\rangle\langle\Phi_+|) \star T_A(\omega)] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0, \omega \perp |\Phi_+\rangle\langle\Phi_+|\} \\
&= \{T_A(\omega) - \text{Tr}_B[\mathcal{S} \star T_A(\omega)] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0, \omega \perp |\Phi_+\rangle\langle\Phi_+|\}.
\end{aligned} \tag{4.44}$$

□

Using a variation of Lemma 4.8 we can show how the dual cone to the cone of states over multiple times behaves under partial traces.

Theorem 4.14. *Let $\hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_n)$ be defined as*

$$\hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_n) = \text{cone}\left(\hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_{n-1}) \star \mathcal{J}(\mathcal{H}_{n-1} \rightarrow \mathcal{H}_n)\right). \tag{4.45}$$

The dual of this hierarchy fulfils

$$\text{Tr}_i\left[\hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_n)^*\right] \supseteq \text{Tr}_i\left[\hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_n)\right]^*. \tag{4.46}$$

Proof. We can show this for general cones using the proof of Lemma 4.8. Let $K \subseteq \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a cone and, then

$$\begin{aligned}
x \in \text{Tr}_B(K)^* &\Leftrightarrow \langle x, \text{Tr}_B(y) \rangle \geq 0 \quad \forall y \in K \Leftrightarrow \langle x \otimes \mathbb{1}, y \rangle \geq 0 \quad \forall y \in K \\
&\Leftrightarrow x \otimes \mathbb{1} \in K^* \Rightarrow x \in \text{Tr}_B(K^*).
\end{aligned} \tag{4.47}$$

We can now set $\mathcal{H}_A = \mathcal{H}_0 \otimes \dots \mathcal{H}_{i-1} \otimes \mathcal{H}_{i+1} \otimes \dots \mathcal{H}_n$, $\mathcal{H}_B = \mathcal{H}_i$ and $K = \hat{\mathcal{Q}}(\mathcal{H}_0 : \dots : \mathcal{H}_n)$ to complete the proof. □

4.3 Symmetry

We call an optimal quantum cost symmetric if

$$\mathcal{K}(\rho, \sigma) = \mathcal{K}(\sigma, \rho), \quad \forall \rho, \sigma \in \mathcal{S}(\mathcal{H}_{A,B}). \quad (4.48)$$

A naive guess or first thought might suggest that symmetry of the cost matrix under input–output exchange, i.e., $\mathcal{SKS} = K$, guarantees symmetry of the associated cost. We will see that this not the case. To understand this, we first address another natural question, namely whether the set of states over time itself is symmetric. In other words, we ask whether swapping the input and output space in the coupling always produces a valid state over time for the reversed pair of states, that is, whether $\mathcal{SQ}(\rho, \sigma)\mathcal{S} = \mathcal{Q}(\sigma, \rho)$.

There is a clear asymmetry in the case when, for example, ρ is full rank and σ is pure, where a single channel that brings ρ to σ exists. Despite that, as seen in Theorem 2.3, if the input state is not full rank a single state over time can correspond to multiple channels. Therefore, in the case where σ is pure, it could be that the single channel that brings ρ to σ inverts into all the channels that bring σ to ρ .

In the following example we perform specific calculations for the case where both states are qubits. We observe, in examples *iii*), *iv*), that some states over time cannot be time inverted, thus strongly suggesting that $\mathcal{SKS} = K$ is not a sufficient condition on the cost matrix to give rise to symmetry of the cost.

Example 4.15. *i) Replacement channel:* Let ρ, σ be any states and let $J_{A \rightarrow B}$ be the Jamiołkowski matrix associated to the constant channel $\mathcal{E}(\rho) = \text{Tr}(\rho)\sigma$, that is $J_{A \rightarrow B} = \mathbb{1} \otimes \sigma$. The associated state over time is $Q = \rho \star J_{A \rightarrow B} = \rho \otimes \sigma$. From the symmetry of the state over time we can see immediately that we can obtain the same result with $(\sigma, J_{A \leftarrow B} = \rho \otimes \mathbb{1})$.

ii) Identity channel: Let ρ be a qubit state with eigenvalues $\{p, 1-p\}$ and $J_{A \rightarrow B}$ be the Jamiołkowski matrix associated to the identity channel, $\mathcal{S}$. Then the associated state over time Q is, in (the tensor basis generated by) the diagonal basis of ρ ,

$$Q = \rho \star \mathcal{S} = \begin{bmatrix} p & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1-p \end{bmatrix}. \quad (4.49)$$

Similarly to before, the symmetry (under subsystem swap) allows us to easily show that the pair $(\sigma = \rho, J_{A \leftarrow B} = \mathcal{S})$ yields the same state over time.

iii) Depolarising channel: Let the initial state be a pure state, without loss of generality (w.l.o.g.), we will set $\rho = |0\rangle\langle 0|$. Let $J_{A \rightarrow B}$ be the Jamiołkowski matrix associated to the depolarising channel $\mathcal{E}(\rho) = (1-p)\rho + p \text{Tr}(\rho)\frac{\mathbb{1}}{2}$, that is $J_{A \rightarrow B} = (1-p)\mathcal{S} + \frac{p}{2}\mathbb{1}$. The resulting state over time is

$$Q = \rho \star J_{A \rightarrow B} = \frac{1}{2} \begin{bmatrix} 2-p & 0 & 0 & 0 \\ 0 & p & 1-p & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4.50)$$

From this channel, $\sigma = \mathcal{E}(\rho) = \frac{1}{2}(2-p)|0\rangle\langle 0| + \frac{p}{2}|1\rangle\langle 1|$. Applying Theorem 2.3 to Q yields

$$J_{A \leftarrow B}^{T_B} = \begin{bmatrix} 1 & 0 & 0 & 1-p \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-p & 0 & 0 & 0 \end{bmatrix}, \quad (4.51)$$

which can be shown through Sylvester's criterion to be not psd by taking the principal minor with $I = \{1, 4\}$ if $p \neq 1$. If $p = 1$, the depolarising channel becomes a replacement channel which we have seen is reversible.

iv) **Dephasing channel:** Let $\rho = |+\rangle\langle+|$ and $J_{A \rightarrow B}$ be the Jamiołkowski matrix associated to the dephasing channel $\mathcal{E}(\rho) = p\rho + (1-p)\sigma_z\rho\sigma_z$ for $p \in (0, 1)$, that is

$$J_{A \rightarrow B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{2p-1}{2} & 0 \\ 0 & \frac{2p-1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4.52)$$

Note that $\sigma = \mathcal{E}(\rho) = \frac{1}{2}(\mathbb{1} + (2p-1)\sigma_x)$, which has rank 2 for $p \in (0, 1)$. We can now calculate the associated state over time

$$Q = \rho \star J_{A \rightarrow B} = \frac{1}{8} \begin{bmatrix} 4 & 2p-1 & 2 & 0 \\ 2p-1 & 0 & 2(2p-1) & 2 \\ 2 & 2(2p-1) & 0 & 2p-1 \\ 0 & 2 & 2p-1 & 4 \end{bmatrix}. \quad (4.53)$$

We can now calculate $J_{A \leftarrow B}$ from Theorem 2.3,⁶ which yields

$$J_{A \leftarrow B} = \frac{1}{4} \begin{bmatrix} 4 & 0 & 2 & 1-2p \\ 0 & 0 & 2p-1 & 2 \\ 2 & 2p-1 & 0 & 0 \\ 1-2p & 2 & 0 & 4 \end{bmatrix}. \quad (4.54)$$

This matrix is clearly not psd under partial transposition of B since the principal minor $[J_{A \leftarrow B}]_{\{1,3\}}$ (which is unaffected by the partial transposition) has negative determinant, thus the matrix is not psd from Sylvester's criterion. For example, when $p = \frac{1}{2}$, the eigenvalues of $J_{A \leftarrow B}^{T_A}$ are $\{\frac{1}{2}(1 \pm \sqrt{2})\}$.

v) **Measure-and-prepare channel:** Let $\rho \in \mathcal{S}(\mathcal{H}_A)$ and $J_{A \rightarrow B}$ be the Jamiołkowski matrix of a measure-and-prepare channel, that is a channel of the form

$$\mathcal{E}(x) = \sum_i \text{Tr}[M_i x] \sigma_i, \quad (4.55)$$

where $\{M_i\}$ is a POVM on $\mathcal{H}_A$ and $\sigma_i \in \mathcal{S}(\mathcal{H}_B)$ are states. Additionally, we assume that $\text{supp } \sigma = \cup_i \text{supp } \sigma_i = \mathcal{H}_B$, else we could redefine $\mathcal{H}_B$ as $\text{supp } \sigma$. Then,

$$J_{A \rightarrow B} = \sum M_i \otimes \sigma_i \quad \text{and} \quad Q = \sum (\rho \star M_i) \otimes \sigma_i. \quad (4.56)$$

Because the map in Theorem 2.3 is CP, as seen in Corollary 2.4, $J_{A \leftarrow B}^{T_A}$ will be positive if Q is positive, and Q will be positive if every $\rho \star M_i$ is (with an if and only if when the σ_i are orthogonal). This will happen if the channel is classical-quantum channels, that is when $\{M_i\}$ is a projective measurement, and ρ is diagonal in a basis defined by this measurement.

As a particular example of this last case, let $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ and $J_{A \rightarrow B} = |0,0\rangle\langle 0,0| + |1,+\rangle\langle 1,+|$, the Jamiołkowski matrix corresponding to the classical-quantum channel that keeps $|0\rangle\langle 0|$ constant and yields $|+\rangle\langle+|$ on input $|1\rangle\langle 1|$. Then $\sigma = \frac{1}{2}(\mathbb{1} + p\sigma_z + (1-p)\sigma_x)$ and the resulting state over time is

$$Q = \rho \star J_{A \rightarrow B} = \frac{1}{2} \begin{bmatrix} 2p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & 1-p \\ 0 & 0 & 1-p & 1-p \end{bmatrix}. \quad (4.57)$$

⁶Note that even though ρ has rank 1, σ has rank 2 and therefore allows us to uniquely apply Theorem 2.3.

If we write $J_{A \leftarrow B}$ we get

$$J_{A \leftarrow B} = \frac{1}{2} \begin{bmatrix} 1+p & p-1 & 0 & 0 \\ p-1 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 1-p \\ 0 & 0 & 1-p & 1+p \end{bmatrix}, \quad (4.58)$$

which is positive and left invariant under partial transposition. Note that $J_{A \leftarrow B} = |0\rangle\langle 0| \otimes \sigma_0 + |1\rangle\langle 1| \otimes \sigma_1$, with

$$\sigma_0 = \frac{1}{2} \begin{bmatrix} 1+p & p-1 \\ p-1 & 1-p \end{bmatrix}, \quad \sigma_1 = \frac{1}{2} \begin{bmatrix} 1-p & 1-p \\ 1-p & 1+p \end{bmatrix}. \quad (4.59)$$

This Jamiołkowski matrix corresponds to another measure-and-prepare map, this time with non-orthogonal measurements since $\text{Tr}[\sigma_0 \sigma_1] = p(1-p)$.

The example shows a type of channels where the symmetry of the set of states over time is broken: channels which reduce the coherence of the input states. In Section 5 we discuss an example where this asymmetry is numerically shown in the cost function, rather than the set of couplings, acting as a proof that $K(\rho, \sigma)$ is not a symmetric function for symmetric cost matrices.

4.4 Triangle inequality

We will state a condition for the fulfilment of the triangle inequality. For this purpose, we consider $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B : \mathcal{H}_C)$ from Definition 2.7 and its dual, $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B : \mathcal{H}_C)^*$. This dual provides us with a cone with respect to which we can define a partial order for cost matrices:

Theorem 4.16. *Consider Hilbert subspaces $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_C$. The inequality*

$$\mathcal{K}_{AB}(\rho_A, \rho_B) + \mathcal{K}_{BC}(\rho_B, \rho_C) \geq \mathcal{K}_{AC}(\rho_A, \rho_C) \quad (4.60)$$

will be fulfilled for all $\rho_i \in \mathcal{S}(\mathcal{H}_i)$, $i \in \{A, B, C\}$ if the cost matrices fulfil the following identity:

$$K_{AB} \otimes \mathbb{1}_C + \mathbb{1}_A \otimes K_{BC} - K_{AC} \otimes \mathbb{1}_B \in \mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B : \mathcal{H}_C)^*. \quad (4.61)$$

Proof. Consider first an admissible Jamiołkowski matrix in systems AB and a cost matrix K_{AB} . Because $\mathbb{1}_B = \text{Tr}_C[J_{BC}]$ for any admissible Jamiołkowski matrix in systems BC , and the partial associativity of the Jordan product [20, 22] we can rewrite the cost as

$$\begin{aligned} \text{Tr}[K_{AB}(\rho \star J_{AB})] &= \text{Tr}[K_{AB}((\rho \star J_{AB}) \star (\mathbb{1}_A \otimes \text{Tr}_C[J_{BC}]))] \\ &= \text{Tr}[(K_{AB} \star (\rho \star J_{AB}))(\mathbb{1}_A \otimes \text{Tr}_C[J_{BC}])] \\ &= \text{Tr}[(K_{AB} \star (\rho \star J_{AB})) \otimes \mathbb{1}_C](\mathbb{1}_A \otimes J_{BC}) \\ &= \text{Tr}[(K_{AB} \otimes \mathbb{1}_C)((\rho \star (J_{AB} \otimes \mathbb{1}_C)) \star (\mathbb{1}_A \otimes J_{BC}))] \\ &= \text{Tr}[(K_{AB} \otimes \mathbb{1}_C)(\rho \star ((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})))]. \end{aligned} \quad (4.62)$$

Similarly, because if a channel yields σ as the image of ρ its Jamiołkowski matrix will fulfil $\text{Tr}_A[\rho \star J_{AB}] = \sigma$ we can operate the cost for any admissible Jamiołkowski matrices and cost K_{BC} as:

$$\begin{aligned} \text{Tr}[K_{BC}(\sigma \star J_{BC})] &= \text{Tr}[K_{BC}((\text{Tr}_A[\rho \star J_{AB}] \otimes \mathbb{1}_C) \star J_{BC})] \\ &= \text{Tr}[(J_{BC} \star K_{BC})(\text{Tr}_A[\rho \star J_{AB}] \otimes \mathbb{1}_C)] \\ &= \text{Tr}[(\mathbb{1}_A \otimes J_{BC}) \star (\mathbb{1}_A \otimes K_{BC})(\rho \star J_{AB} \otimes \mathbb{1}_C)] \\ &= \text{Tr}[(\mathbb{1}_A \otimes K_{BC})(\rho \star (J_{AB} \otimes \mathbb{1}_C)) \star (\mathbb{1}_A \otimes J_{BC})] \\ &= \text{Tr}[(\mathbb{1}_A \otimes K_{BC})(\rho \star ((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})))]. \end{aligned} \quad (4.63)$$

Finally, for systems AC , consider the link product [36] of any two Jamiołkowski matrices as the Jamiołkowski matrix of AC :

$$\begin{aligned}
\text{Tr}[K_{AC}(\rho \star J_{AC})] &= \text{Tr}[K_{AC}(\rho \star \text{Tr}_B[(J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})])] \\
&= \text{Tr}[(K_{AC} \star \rho) \text{Tr}_B[(J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})]] \\
&= \text{Tr}[(K_{AC} \otimes \mathbb{1}_B) \star \rho)((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}))] \\
&= \text{Tr}[(K_{AC} \otimes \mathbb{1}_B)(\rho \star ((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})))].
\end{aligned} \tag{4.64}$$

Let $K' = K_{AB} \otimes \mathbb{1}_C + \mathbb{1}_A \otimes K_{BC} - K_{AC} \otimes \mathbb{1}_B$. With these 3 equalities in hand, we can consider 3 optimal Jamiołkowski matrices, indicated by the superindex o , for the costs $\mathcal{K}_{AB}$ and $\mathcal{K}_{BC}$ and $\mathcal{K}_{AC}$. Then by using the previous expressions we can show that:

$$\begin{aligned}
\mathcal{K}_{AC} &= \text{Tr}[K_{AC}(\rho \star J_{AC}^o)] \leq \text{Tr}[K_{AC}(\rho \star J_{AC})] \\
&= \text{Tr}[K_{AC} \otimes \mathbb{1}_B(\rho \star ((J_{AB}^o \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}^o)))] \\
&= \text{Tr}[(K_{AB} \otimes \mathbb{1}_C + \mathbb{1}_A \otimes K_{BC} - K')(\rho \star ((J_{AB}^o \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}^o)))] \\
&= \text{Tr}[(\mathbb{1}_A \otimes K_{BC})(\rho \star ((J_{AB}^o \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}^o)))] \\
&\quad + \text{Tr}[(K_{AB} \otimes \mathbb{1}_C)(\rho \star ((J_{AB}^o \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}^o)))] \\
&\quad - \text{Tr}[K'(\rho \star (J_{AB}^o \star J_{BC}^o))] \\
&= \text{Tr}[K_{AB}(\rho \star J_{AB}^o)] + \text{Tr}[K_{BC}(\rho \star J_{BC}^o)] - \text{Tr}[K'(\rho \star (J_{AB}^o \star J_{BC}^o))] \\
&= \mathcal{K}_{AB} + \mathcal{K}_{BC} - \text{Tr}[K'(\rho \star (J_{AB}^o \star J_{BC}^o))].
\end{aligned} \tag{4.65}$$

Finally, because K' is in the dual of $\mathcal{Q}_3$, $\text{Tr}[K'(\rho \star (J_{AB}^o \star J_{BC}^o))] \geq 0$ and therefore $\mathcal{K}_{AB} + \mathcal{K}_{BC} \geq \mathcal{K}_{AC}$. $\square$

For the sake of completeness, this general statement can be converted to a statement about the triangle inequality for the cost $\mathcal{K}$ associated to a given cost matrix K :

Corollary 4.17. *Let $\mathcal{H}$ be a Hilbert space, $K \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ and $\mathcal{K}$ the quantum optimal cost associated to K . Then $\mathcal{K}$ fulfils the triangle inequality if*

$$K \otimes \mathbb{1}_3 + \mathbb{1}_1 \otimes K - K \otimes \mathbb{1}_2 \in \mathcal{Q}(\mathcal{H} : \mathcal{H} : \mathcal{H})^*, \tag{4.66}$$

where the subindices indicate different copies of the same Hilbert space $\mathcal{H}$.

4.5 Structural properties of the transport cost

We would like to prove subadditivity of the cost for both inputs, unfortunately we can only see it for the second input. The following proposition shows this result as well as two consequences of the triangle inequality.

Proposition 4.18. *Let p_x be a probability distribution. The optimal quantum cost fulfils the following:*

$$i) \text{ Convexity in the second argument: } \mathcal{K}(\rho, \sum_x p_x \sigma_x) \leq \sum_x p_x \mathcal{K}(\rho, \sigma_x).$$

Moreover, if the triangle inequality is fulfilled:

$$ii) \sum_x p_x \mathcal{K}(\rho_x, \sigma) \leq \mathcal{K}(\sum_x p_x \rho_x, \sigma) + \sum_x p_x \mathcal{K}(\rho_x, \sum_{x'} p_{x'} \rho_{x'}).$$

$$iii) \mathcal{K}(\sum_x p_x \rho_x, \sigma) \leq \sum_x p_x \mathcal{K}(\rho_x, \sigma) + \sum_{x'} p_{x'} \mathcal{K}(\sum_x p_x \rho_x, \rho_{x'}).$$

Proof. Consider optimal Jamiołkowski matrices J_o for $(\rho, \sum_x p_x \sigma_x)$ and J_x for (ρ, σ_x) . Note that $J_\Sigma = \sum_x p_x J_x$ is a Jamiołkowski matrix with an associated channel that fulfils $\mathcal{E}_{J_\Sigma}(\rho) = \sum_x p_x \sigma_x$. Thus,

$$\begin{aligned} \mathcal{K}\left(\rho, \sum_x p_x \sigma_x\right) &= \text{Tr}[K(\rho \star J_o)] \leq \text{Tr}[K(\rho \star J_\Sigma)] = \text{Tr}\left[K\left(\rho \star \left(\sum_x p_x J_x\right)\right)\right] \\ &= \sum_x p_x \text{Tr}[K(\rho \star J_x)] = \sum_x p_x \mathcal{K}(\rho, \sigma_x), \end{aligned} \quad (4.67)$$

where we used the bilinearity of the Jordan product [20] and the linearity of the trace.

The second property is a direct consequence of the triangle inequality:

$$\begin{aligned} \sum_x p_x \mathcal{K}(\rho_x, \sigma) &\leq \sum_x p_x \left(\mathcal{K}\left(\rho_x, \sum_{x'} p_{x'} \rho_{x'}\right) + \mathcal{K}\left(\sum_{x'} p_{x'} \rho_{x'}, \sigma\right) \right) \\ &= \mathcal{K}\left(\sum_x p_x \rho_x, \sigma\right) + \sum_x p_x \mathcal{K}\left(\rho_x, \sum_{x'} p_{x'} \rho_{x'}\right). \end{aligned} \quad (4.68)$$

Similarly, we can show the third property. Let $\rho = \sum_x p_x \rho_x$. Then, $\forall x$

$$\begin{aligned} p_x \mathcal{K}(\rho, \sigma) &\leq p_x \mathcal{K}(\rho, \rho_x) + p_x \mathcal{K}(\rho_x, \sigma) \\ \Rightarrow \sum_x p_x \mathcal{K}(\rho, \sigma) &\leq \sum_x p_x \mathcal{K}(\rho, \rho_x) + \sum_x p_x \mathcal{K}(\rho_x, \sigma) \\ \Rightarrow \mathcal{K}(\rho, \sigma) &\leq \sum_x p_x \mathcal{K}(\rho, \rho_x) + \sum_x p_x \mathcal{K}(\rho_x, \sigma), \end{aligned} \quad (4.69)$$

where we first used the triangle inequality and then we added all the inequalities together. $\square$

Remark 4.19. A similar proof does not work for convexity on the first input and joint subadditivity because of the following. We will use subadditivity on the first input as an example. Let J_o be the optimal Jamiołkowski matrix for $(\sum_x p_x \rho_x, \sigma)$. Starting on the left hand side we obtain

$$\mathcal{K}\left(\sum_x p_x \rho_x, \sigma\right) = \text{Tr}\left[K\left(\sum_x p_x \rho_x \star J_o\right)\right] = \sum_x p_x \text{Tr}[K(\rho_x \star J_o)]. \quad (4.70)$$

At this point we can observe that the channel associated to J_o does not necessarily have output σ for each ρ_x (unless σ is pure) and we can not upper bound the associated cost with anything defined with the optimal channels for the pairs (ρ_x, σ) . In contrast, in the proof of Proposition 4.18 it was possible to define the joint channel J_Σ because we could send ρ to each element σ_x of the ensemble $\{(p_x, \sigma_x)\}$ and that would in total define a channel that sends ρ to σ .

Moreover, we can show joint convexity is false in general. Let $\sigma_x = \mathcal{E}_{J_o}(\rho_x)$ and observe that letting $\sigma = \sum_x p_x \sigma_x$ in the previous equation we obtain:

$$\mathcal{K}\left(\sum_x p_x \rho_x, \sum_x p_x \sigma_x\right) \geq \sum_x p_x \mathcal{K}(\rho_x, \sigma_x). \quad (4.71)$$

This joint concavity is not general in the sense that we have the relation only for $\sigma_x = \mathcal{E}_{J_o}(\rho_x)$, where the ensemble $\{(p_x, \rho_x)\}$ can be arbitrarily chosen, but the channel must be the one associated to the optimal Jamiołkowski matrix for $\mathcal{K}(\rho, \sigma)$.

For the next proposition, we consider bipartite initial and final spaces, and for simplicity we assume them to be the same: $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}_1 \otimes \mathcal{H}_2$. In this setting we want to study how the cost for product states relates to related costs in $\mathcal{H}_1$ and $\mathcal{H}_2$ for different configurations of the joint cost matrix.

Proposition 4.20. *Let $\mathcal{H}_i$ be Hilbert spaces and $\rho_i, \sigma_i \in \mathcal{S}(\mathcal{H}_i)$ with $i = 1, 2$. Let K_{12} be a cost matrix in $\mathcal{B}((\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes 2})$, and K_i be cost matrices in $\mathcal{B}(\mathcal{H}_i^{\otimes 2})$. Then the optimal transport cost of $\mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$ fulfils the following:*

i) If $K_{12} = K_1 \otimes K_2$, then $\mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) \leq \mathcal{K}(\rho_1, \sigma_1)\mathcal{K}(\rho_2, \sigma_2)$.

ii) If $K_{12} = K_1 \otimes \mathbb{1}_2 + \mathbb{1}_1 \otimes K_2$, then $\mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) \leq \mathcal{K}(\rho_1, \sigma_1) + \mathcal{K}(\rho_2, \sigma_2)$.

Proof. Objects in different subsystems commute and $J_1 \otimes J_2$ is admissible for $(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2)$ if J_i is admissible for ρ_i, σ_i , $i = 1, 2$. To show the first inequality, consider two optimal Jamiołkowski matrices J_1^o, J_2^o that optimise the costs between ρ_i, σ_i with cost matrix K_i , $i = 1, 2$. Then,

$$\begin{aligned} \mathcal{K}(\rho_1, \sigma_1) \cdot \mathcal{K}(\rho_2, \sigma_2) &= \text{Tr}[K_1(\rho_1 \star J_1^o)] \cdot \text{Tr}[K_2(\rho_2 \star J_2^o)] \\ &= \text{Tr}[(K_1(\rho_1 \star J_1^o)) \otimes (K_2(\rho_2 \star J_2^o))] = \text{Tr}[(K_1 \otimes K_2)((\rho_1 \otimes \rho_2) \star (J_1^o \otimes J_2^o))] \\ &\geq \mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2). \end{aligned} \quad (4.72)$$

For the second inequality, consider the same Jamiołkowski matrices as before. Then

$$\begin{aligned} \mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) &= \text{Tr}[K_{12}(\rho_1 \otimes \rho_2) \star J_{12}^o] \\ &\leq \text{Tr}[(K_1 \otimes \mathbb{1}_2 + \mathbb{1}_1 \otimes K_2)(\rho_1 \otimes \rho_2) \star (J_1^o \otimes J_2^o)] \\ &= \text{Tr}[K_1 \rho_1 \star J_1^o] \text{Tr}[\rho_2 \star J_2^o] + \text{Tr}[\rho_1 \star J_1^o] \text{Tr}[K_2 \rho_2 \star J_2^o] \\ &= \mathcal{K}(\rho_1, \sigma_1) + \mathcal{K}(\rho_2, \sigma_2). \end{aligned} \quad (4.73)$$

□

5 Unitary invariant cost

Let $\mathcal{H}$ be a finite dimensional Hilbert space of dimension d . Here we consider cost matrices which yield unitary invariant quantum optimal costs, that is

$$\mathcal{K}(\rho, \sigma) = \mathcal{K}(U\rho U^*, U\sigma U^*) \quad \forall U \in U(d), \rho, \sigma \in \mathcal{S}(\mathcal{H}). \quad (5.1)$$

This follows automatically if the cost matrix is invariant under simultaneous unitary transformations of both systems:

$$K = (U \otimes U)K(U^* \otimes U^*) \quad \forall U \in U(d). \quad (5.2)$$

In this section, we consider the case where the cost is unitarily invariant (UI), meaning it remains unchanged under a common unitary transformation of both the input and output states:

$$\kappa(\rho, \mathcal{E}) = \kappa(U\rho U^*, \mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^{-1}) \quad \forall U \in U(d), \rho, \sigma \in \mathcal{S}(\mathcal{H}). \quad (5.3)$$

where we have defined the unitary channel $\mathcal{U}(X) = UXU^*$, and we denote by $U(d)$ the group of unitary operators acting on the Hilbert space, of dimension d . We show the relation between the maps and states in the following commutative diagram:

$$\begin{array}{ccc} \rho & \xrightarrow{\mathcal{E}} & \sigma \\ \mathcal{U} \downarrow & & \downarrow \mathcal{U} \\ U\rho U^* & \xrightarrow{\mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^{-1}} & U\sigma U^* \end{array}$$

Throughout this section, we show that unitary invariance fully determines the cost matrix up to a single parameter. Later, we compute the cost for several relevant quantum channels, and derive a simplified optimisation problem to compute the cost $\mathcal{K}(\rho, \sigma)$ valid for arbitrary states within the UI setting. In Proposition 5.10, we obtain analytic expressions for the optimal cost in the special case of commuting states. Section 5.2 explores the striking modification of the cost induced by embedding

the problem into a higher-dimensional Hilbert space, and study its asymptotic behaviour. Finally, in Example 5.5, we present an example that illustrates several noteworthy features.

The following proposition shows that there is a single cost matrix (up to positive scaling) with the UI property:

Proposition 5.1. *The only cost matrices that belong to the dual to the cone of states over time, assign cost 0 to the identity channel according to Proposition 4.1 and commute with unitaries of the form $U \otimes U$ are positive multiples of*

$$K_0 = d\mathbb{1} - \mathcal{S}. \quad (5.4)$$

Proof. First, let us prove that if J is the Jamiołkowski matrix whose channel takes ρ to $\sigma = \mathcal{E}(\rho)$, then the Jamiołkowski matrix corresponding to the associated channel $\mathcal{E}'_U = \mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^{-1}$ which maps $U\rho U^*$ to $U\sigma U^*$, is given by $J'_U = (U \otimes U)J(U^* \otimes U^*)$.

For this purpose we start from the definition in Eq. (1.4)

$$\begin{aligned} J'_U &= (\text{id} \otimes \mathcal{E}'_U)(\mathcal{S}) = (\text{id} \otimes \mathcal{U}) \circ (\text{id} \otimes \mathcal{E}) \circ (\text{id} \otimes \mathcal{U}^{-1})(\mathcal{S}) \\ &= (\mathbb{1} \otimes U)(\text{id} \otimes \mathcal{E})(\mathbb{1} \otimes U^*)\mathcal{S}(\mathbb{1} \otimes U)(\mathbb{1} \otimes U)^* \end{aligned} \quad (5.5)$$

Recall that the swap operator satisfies $\mathcal{S}(X \otimes Y)\mathcal{S} = Y \otimes X$, i.e. $(I \otimes V)\mathcal{S} = \mathcal{S}(V \otimes I)$. Using this relation twice we can write $(\mathbb{1} \otimes U^*)\mathcal{S}(\mathbb{1} \otimes U) = \mathcal{S}(U^* \otimes \mathbb{1})(\mathbb{1} \otimes U) = \mathcal{S}(\mathbb{1} \otimes U)(U^* \otimes \mathbb{1}) = (U \otimes \mathbb{1})\mathcal{S}(U^* \otimes \mathbb{1})$, and thereby pull the unitaries out of the action of the map $\mathcal{E}$, i.e.

$$\begin{aligned} J'_U &= (\mathbb{1} \otimes U)(U \otimes \mathbb{1})(\text{id} \otimes \mathcal{E})(\mathcal{S})(U^* \otimes \mathbb{1})(\mathbb{1} \otimes U^*) \\ &= (U \otimes U)J(U^* \otimes U^*). \end{aligned} \quad (5.6)$$

Now we can impose the UI of the cost function, $\kappa(\rho, \mathcal{E}) = \kappa(U\rho U^*, \mathcal{E}'_U)$, to obtain

$$\begin{aligned} \text{Tr}[K_0(\rho \star J)] &= \text{Tr}[K_0((U\rho U^*) \star ((U \otimes U)J(U^* \otimes U^*)))] \\ &= \text{Tr}[(U \otimes U)K_0(U^* \otimes U^*)(\rho \star J)]. \end{aligned} \quad (5.7)$$

We require this equality to hold for all choices of input states and channels, i.e. for all elements of the set of states over time $Q = \rho \star J$. Since this set is spanning (Proposition 4.10), the scalar equality Eq. (5.7) translates into the operator equality $K_0 = (U \otimes U)K_0(U^* \otimes U^*)$ or equivalently $[U \otimes U, K_0] = 0$. This must hold for all unitaries.

From the representation theory of $GL(d)$, and because the set of unitaries generates the whole of $GL(d)$ as an algebra, whose operations leave the commutator invariant, the only elements with this property are the symmetric and antisymmetric projectors. The vector space generated by these two projectors also has $\{\mathbb{1}, \mathcal{S}\}$ as a basis [45]. Therefore $K_0 = a(b\mathbb{1} - \mathcal{S})$ with real a and b , to preserve Hermiticity.

We can now impose the second condition, Proposition 4.1:

$$0 = \text{Tr}_B[\mathcal{S} \star K_0] = a \text{Tr}_B[\mathcal{S} \star (b\mathbb{1} - \mathcal{S})] = a \text{Tr}_B[b\mathcal{S} - \mathbb{1}] = a(b - d)\mathbb{1}. \quad (5.8)$$

We find that $b = d$. Finally, we see that the positivity of the cost requires a to be positive:

$$\begin{aligned} \text{Tr}[K_0(\rho \star J)] &= a \text{Tr}[(d\mathbb{1} - \mathcal{S})(\rho \star J)] \\ &= ad \text{Tr}[\rho \star J] - a \text{Tr}[\mathcal{S}(\rho \star J)] \\ &= a[d - \text{Tr}[\rho(\mathcal{S} \star J)]] \end{aligned} \quad (5.9)$$

The last term can be upper bounded by switching to the Choi matrix representation of the problem using Lemma 4.7. Recall that $\mathcal{S}^{TA} = |\Phi_+\rangle\langle\Phi_+|$, where $|\Phi_+\rangle = \sum_i |ii\rangle$ is the unnormalised maximally entangled state and is such that $\text{Tr}[|\Phi_+\rangle\langle\Phi_+|] = d$. Then

$$\begin{aligned} \text{Tr}[\rho(\mathcal{S} \star J)] &= \text{Tr}[\mathcal{S}^{TA}(\rho \star J)] = \text{Tr}[|\Phi_+\rangle\langle\Phi_+|(\rho \star J)^{TA}] = \langle\Phi_+|\rho^T \star J^{TA}|\Phi_+\rangle \\ &= \langle\Phi_+|\rho^T \star C|\Phi_+\rangle \leq d \text{Tr}[\rho^T \star C] = d. \end{aligned} \quad (5.10)$$

Thus we have that a times a positive constant has to be positive, therefore a is positive. $\square$

We will also use the normalised version of K_0 : $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$, so that the maximum achievable cost is 1. With this specific cost matrix, there are some simple ways to write the cost associated to a channel depending on which representation of the channel we take. These forms will be useful later.

Remark 5.2. Let ρ be a state in a finite dimensional Hilbert space $\mathcal{H}$ and the cost matrix $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$. Consider a channel $\mathcal{E}$ with associated Jamiołkowski and Choi matrices J , C , respectively and Kraus representation $\{E_k\}$. Moreover, let $\rho = \sum_i p_i |i\rangle\langle i|$ for some basis $\{|i\rangle\}$, $|\rho\rangle = \sum_i p_i |ii\rangle$ its vectorized form and $|\Phi_+\rangle = \sum_i |ii\rangle$ is the unnormalised maximally entangled state. Then

$$\kappa(\rho, \mathcal{E}) = \text{Tr} [\tilde{K}_0(\rho \star J)] = 1 - \frac{1}{d} \langle \Phi_+ | \rho^T \star C | \Phi_+ \rangle \quad (5.11)$$

$$= 1 - \frac{1}{d} \Re(\langle \rho | C | \Phi_+ \rangle) \quad (5.12)$$

$$= 1 - \frac{1}{d} \sum_{ij} \frac{p_i + p_j}{2} \langle i | \mathcal{E}(|i\rangle\langle j|) | j \rangle \quad (5.13)$$

$$= 1 - \frac{1}{d} \sum_i p_i \sum_j \Re(\langle i | \mathcal{E}(|i\rangle\langle j|) | j \rangle) \quad (5.14)$$

$$= 1 - \frac{1}{d} \sum_k \Re(\text{Tr}[E_k^*] \text{Tr}[E_k \rho]). \quad (5.15)$$

Proof. The term 1 in every equation comes from the trace of the states over time with the identity, which is always one because the partial trace of a state over time is a state. The other part is associated to $\text{Tr}[\mathcal{S}(\rho \star J)]$, and we will focus on that.

Eq. (5.11) and Eq. (5.12) are a direct consequence of Lemma 4.7, recalling that $\mathcal{S}^{TA} = |\Phi_+\rangle\langle\Phi_+|$.

Eq. (5.13) comes from the definition of the Jamiołkowski matrix, $J = \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|j\rangle\langle i|)$ and Theorem 2.3, which shows that in the product basis of the diagonal basis of ρ , $\rho \star J = \sum_{ij} \frac{p_i + p_j}{2} |i\rangle\langle j| \otimes \mathcal{E}(|j\rangle\langle i|)$. Then we define the swap operator in this product basis, $\mathcal{S} = \sum_{i'j'} |i'\rangle\langle j'| \otimes |j'\rangle\langle i'|$ and calculate $\text{Tr}[\mathcal{S}(\rho \star J)]$:

$$\begin{aligned} \text{Tr}[\mathcal{S}(\rho \star J)] &= \text{Tr} \left[\sum_{ij} \frac{p_i + p_j}{2} \sum_{i'j'} (|i'\rangle\langle j'| \otimes |j'\rangle\langle i'|) (|i\rangle\langle j| \otimes \mathcal{E}(|j\rangle\langle i|)) \right] \\ &= \sum_{ij} \frac{p_i + p_j}{2} \sum_{i'j'} \delta_{ij'} \delta_{ji'} \langle i' | \mathcal{E}(|j\rangle\langle i|) | j' \rangle = \sum_{ij} \frac{p_i + p_j}{2} \langle i | \mathcal{E}(|i\rangle\langle j|) | j \rangle. \end{aligned} \quad (5.16)$$

Finally, for Eq. (5.15) consider J written as a function of the Kraus operators:

$$J = (\text{id} \otimes \mathcal{E})(\mathcal{S}) = \sum_k (\mathbb{1} \otimes E_k) \mathcal{S} (\mathbb{1} \otimes E_k^*) = \sum_k (E_k^* \otimes E_k) \mathcal{S}. \quad (5.17)$$

Then we add ρ and the $\mathcal{S}$ from the cost:

$$\begin{aligned} \text{Tr}[\mathcal{S}(\rho \star J)] &= \sum_k \text{Tr}[\mathcal{S}(\rho \star (E_k^* \otimes E_k) \mathcal{S})] \\ &= \frac{1}{2} \sum_k (\text{Tr}[\mathcal{S}(\rho(E_k^* \otimes E_k) \mathcal{S})] + \text{Tr}[\mathcal{S}((E_k^* \otimes E_k) \mathcal{S} \rho)]) \\ &= \frac{1}{2} \sum_k (\text{Tr}[\rho E_k^* \otimes E_k] + \text{Tr}[E_k \rho \otimes E_k^*]) = \frac{1}{2} \sum_k (\text{Tr}[\rho E_k^*] \text{Tr}[E_k] + h.c.) \\ &= \sum_k \Re(\text{Tr}[E_k^*] \text{Tr}[\rho E_k]). \end{aligned} \quad (5.18)$$

This concludes the proof. $\square$

Remark 5.3. It is well known that a single channel can have multiple Kraus representations [35, Theorem 8.2]. By Eq. (5.15), for every Kraus representation of a channel

$$1 - \frac{1}{d} \sum_k \Re(\text{Tr}[E_k^*] \text{Tr}[\rho E_k]) = \text{Tr}[S(\rho \star J)], \quad (5.19)$$

and the Jamiołkowski matrix is unique, therefore different Kraus representations of the same channel have the same associated cost.

We can also show this explicitly. If two Kraus representations $\{E_i\}$, $\{F_j\}$ give rise to the same quantum channel, then there exists a unitary $U = (U_{ij})$ such that $E_i = \sum_j U_{ij} F_j$ [35]. Then

$$\begin{aligned} \sum_i \Re(\text{Tr}[E_i^*] \text{Tr}[\rho E_i]) &= \sum_i \Re\left(\text{Tr}\left[\sum_j \bar{U}_{ij} F_j^*\right] \text{Tr}\left[\rho \sum_{j'} U_{ij'} F_{j'}\right]\right) \\ &= \sum_{jj'} \left(\sum_i U_{ij'} \bar{U}_{ij}\right) \Re(\text{Tr}[F_j^*] \text{Tr}[\rho F_{j'}]) \\ &= \sum_{jj'} (U^T (U^T)^*)_{j'j} \Re(\text{Tr}[F_j^*] \text{Tr}[\rho F_{j'}]) \\ &= \sum_{jj'} \delta_{jj'} \Re(\text{Tr}[F_j^*] \text{Tr}[\rho F_{j'}]) = \sum_j \Re(\text{Tr}[F_j^*] \text{Tr}[\rho F_j]). \end{aligned} \quad (5.20)$$

In the following examples we calculate the cost associated to two important channels using the unitary invariant cost matrix. Example 5.5 in particular allows for some observations on the properties of the unitary invariant cost.

Example 5.4 (The replacement channel). Consider $\mathcal{E}(x) = (\text{Tr } x)\sigma$. Then the associated Jamiołkowski matrix is $\mathbb{1} \otimes \sigma = \sigma_B$ and

$$\kappa(\rho, \mathcal{E}_R) = \text{Tr}[(d\mathbb{1} - \mathcal{S})\rho_A \star \sigma_B] = d - \text{Tr}[\rho\sigma] \geq d - 1, \quad (5.21)$$

where the last inequality can be seen, for example, using the trace and operator norms of ρ and σ : $\text{Tr}[\rho\sigma] \leq \|\rho\|_{tr} \|\sigma\|_{op} \leq 1$.

Example 5.5 (Unitary channels). Consider now the class of unitary channels: $\mathcal{E}_U(x) = UxU^*$, where U is a unitary operator. The associated Jamiołkowski matrix is $J_U = U_B \mathcal{S} U_B^*$, where $U_B = \mathbb{1} \otimes U$. We can use Eq. (5.15) in Remark 5.2 to calculate the cost, since the Kraus operators associated to a Unitary channel are just $\{U\}$. This cost is then

$$\kappa(\rho, \mathcal{E}_U) = \text{Tr}[(\rho_A \star (U_B \mathcal{S} U_B^*))(d\mathbb{1} - \mathcal{S})] = d - \Re(\text{Tr}[U^*] \text{Tr}[\rho U]) \quad (5.22)$$

We can find an expression for the optimal U when ρ and σ are pure states. Due to the unitary invariance, w.l.o.g. $\rho = |0\rangle\langle 0|$ and $\sigma = |\varphi\rangle\langle \varphi|$ where $|\varphi\rangle = \alpha|0\rangle + \sqrt{1-\alpha^2}|1\rangle$ with $\alpha \in \mathbb{R}_+$. The optimal unitary (in terms of maximising its trace) will leave $\{|0\rangle, |1\rangle\}$ invariant and have 1 in the diagonal elements outside this subspace. Therefore the optimal (i.e. largest) value is

$$\Re(\text{Tr}[U^*] \text{Tr}[\rho U]) = \alpha(d - 2 + 2\alpha) \quad (5.23)$$

with associated cost

$$\mathcal{K}_U(\rho, \sigma) = d - \alpha(d - 2 + 2\alpha) = d(1 - \alpha) + 2\alpha(1 - \alpha) = (1 - \alpha)(d + 2\alpha). \quad (5.24)$$

Note that this optimum value is influenced by the action of the unitary on an invariant subspace orthogonal to the subspace where our state evolves; in particular it depends on its dimension. We will later address this further and show how we can remove this dependency in the limit when $d \rightarrow \infty$.

If we now further restrict the problem to $d = 2$, then this becomes $2(1 - \alpha)(1 + \alpha) = 2(1 - \alpha^2) = 2(1 - |\langle 0|\varphi\rangle|^2) = 2T(|0\rangle\langle 0|, |\varphi\rangle\langle \varphi|)^2$, where T is the trace distance. Since it is the square of a distance, it cannot be a distance, as it violates the triangle inequality.

In the limit of high $d \rightarrow \infty$ this quantity approximately becomes $d(1 - \alpha)$, which, for small angles is $d(1 - \cos \theta) \approx \frac{d}{2}\theta^2$, which is again the square of a distance.

Importantly, the second example computes the optimal unitary cost, which can then be compared to the optimal cost over all channels. In Fig. 3, we numerically observe that for dimension 4, the cost and the unitary cost are equal when the input states are pure but differ when the states are mixed. We analytically prove the case for general pure states in Proposition 5.8.

For the proof of Proposition 5.8 we will first show Lemma 5.6, where we consider the case where the joint support of ρ and σ , by which we mean $\mathcal{H}_S = \text{supp } \rho + \text{supp } \sigma$, is strictly contained in the overall Hilbert space $\mathcal{H} = \mathcal{H}_S \oplus \mathcal{H}_\perp$. This setting will appear again later in Section 5.2.

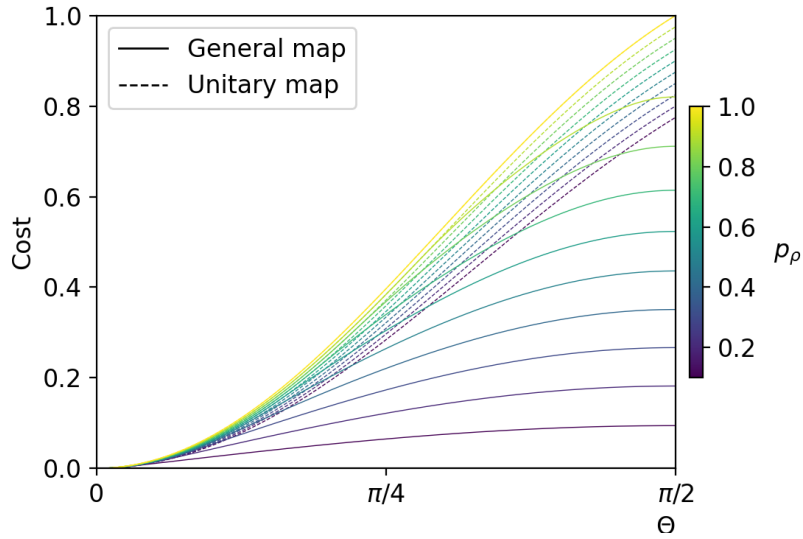


Figure 3: Plot comparing the global optimal cost with the result of optimising only over unitaries for various values of p_ρ , where the states are $\rho = p_\rho |0\rangle\langle 0| + (1 - p_\rho)\frac{1}{d}\mathbb{1}$ and $\sigma = p_\rho |\varphi\rangle\langle \varphi| + (1 - p_\rho)\frac{1}{d}\mathbb{1}$.

Lemma 5.6. *Let ρ, σ have joint support $\mathcal{H}_S \subseteq \mathcal{H} = \mathcal{H}_S \oplus \mathcal{H}_\perp$. Then there exists an optimal channel of the unitary invariant quantum optimal transport $\mathcal{K}(\rho, \sigma)$ such that its associated Kraus operators are of the form*

$$E = E_S \oplus c\Pi_\perp, \quad (5.25)$$

where $\Pi_\perp$ is the projector on the orthogonal, or embedding Hilbert space, $\mathcal{H}_\perp$. Therefore, the optimal channel acts as the identity channel on the embedding Hilbert space.

Proof. We start by noting that our optimization problem (3.3) remains invariant under unitaries that act non-trivially only outside the joint support of the input and output states, i.e. they are of the form $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$, where Π_S is the projector onto $\mathcal{H}_S$ and $U_\perp$ are unitaries on $\mathcal{H}_\perp$ (see [46] for a general discussion of SDP under symmetries). We show that the cost and constraints of the problem remain invariant under unitaries of this form: Given a valid Jamiołkowski matrix J , we can construct a new Jamiołkowski matrix $J_{U_\perp} = (\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp) J (\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*)$ that has the same cost and satisfies the same constraints as J from (3.3):

$$\begin{aligned} \text{Tr}[(\tilde{K}_0 \star \rho) J_{U_\perp}] &= \text{Tr}[(\tilde{K}_0 \star \rho)((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J ((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))] \\ &= \text{Tr}[(\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*)(\tilde{K}_0 \star \rho)((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J] \\ &= \text{Tr}[(\tilde{K}_0 \star \rho) J], \end{aligned} \quad (5.26)$$

where we have used the unitary invariance of the cost $\tilde{K}_0$ and the fact that ρ and σ commute with $U_\perp$. Similarly, the constraints of the problem are also invariant:

$$\begin{aligned}\text{Tr}_A [\rho J_{U_\perp}] &= \text{Tr}_A [\rho ((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))] \\ &= \text{Tr}_A [((\Pi_S \oplus U_\perp^*) \otimes \mathbb{1}) \rho ((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J(\mathbb{1} \otimes (\Pi_S \oplus U_\perp^*))] \\ &= (\Pi_S \oplus U_\perp) \text{Tr}_A [\rho J] (\Pi_S \oplus U_\perp^*) = (\Pi_S \oplus U_\perp) \sigma (\Pi_S \oplus U_\perp^*) = \sigma,\end{aligned}\quad (5.27)$$

$$\begin{aligned}\text{Tr}_B [J_{U_\perp}] &= \text{Tr}_B [((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))] \\ &= \text{Tr}_B [(\mathbb{1} \otimes (\Pi_S \oplus U_\perp^*)) ((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)) J((\Pi_S \oplus U_\perp^*) \otimes \mathbb{1})] \\ &= (\Pi_S \oplus U_\perp) \text{Tr}_B [J] (\Pi_S \oplus U_\perp^*) = (\Pi_S \oplus U_\perp) \mathbb{1} (\Pi_S \oplus U_\perp^*) = \mathbb{1}.\end{aligned}\quad (5.28)$$

Due to the linearity of the cost and constraints, it follows that a twirled matrix $J' = \int dU_\perp J_{U_\perp}$ is also a valid Jamiołkowski matrix, as it represents a convex combination of valid Jamiołkowski matrices. That is, for any J we can construct a twirled version J' that retains the same cost. Hence, without loss of generality, we can optimize our cost over the set of Jamiołkowski matrices satisfying the symmetry condition: $[(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp^*), J] = 0$.

Since we wish to identify optimal Kraus operators we will work with the Choi matrix C : the canonical Kraus representation can be readily obtained from $C = \sum_E |E\rangle\langle E|$, where the (unnormalized) eigenstates are $|E\rangle = \sum_{rs} E_{rs} |rs\rangle$ are the vectorised form of the Kraus operator $\sum_{rs} E_{rs} |r\rangle\langle s|$. In addition, we can use a basis for the partial transpose such that the Choi matrices inherit the symmetry of the problem. In particular, we consider a basis such that Π_S is diagonal which will preserve the direct sum structure of $\Pi_S \oplus U_\perp$. In such a basis, the Choi matrix inherits the symmetry of the problem as follows:

$$\begin{aligned}C' &= J'^{TA} = \int dU_\perp J_{U_\perp}^{TA} = \int dU_\perp ((\Pi_S \oplus \bar{U}_\perp) \otimes (\Pi_S \oplus U_\perp)) J^{TA} ((\Pi_S \oplus U_\perp^T) \otimes (\Pi_S \oplus U_\perp^*)) \\ &= \int dU_\perp ((\Pi_S \oplus \bar{U}_\perp) \otimes (\Pi_S \oplus U_\perp)) C((\Pi_S \oplus U_\perp^T) \otimes (\Pi_S \oplus U_\perp^*))\end{aligned}\quad (5.29)$$

where in the second equality we have used the properties of the partial trace given in Lemma 4.7. In order to ease the notation we will use Latin letters to label the n elements of the basis of $\mathcal{H}_S$ and Greek letters to label the $d_\perp = d - n$ elements of the basis of $\mathcal{H}_\perp$. We can expand the 16 terms appearing from the double direct sum in each side of the twirl in Eq. (5.29). All terms where each factor $(U_\perp)_{\alpha\beta}$ cannot be matched with a factor $(\bar{U}_\perp)_{\alpha\beta}$ will be zero. This follows from the invariance of the Haar measure $dU = dU'$ taking $U = WU'Z$ with $W = \sum_\alpha e^{i\theta_\alpha} |\alpha\rangle\langle\alpha|$ and $Z = \sum_\alpha e^{i\varphi_\alpha} |\alpha\rangle\langle\alpha|$. Any unmatched factor $U_{\alpha\beta}$ picks up a phase $e^{i(\theta_\alpha - \varphi_\beta)}$. Since the result must be independent of the value of these phases, it must vanish—evident, for instance, by integrating over uniformly distributed phases in $[0, 2\pi]$. For example, $\int dU_\perp (U_\perp)_{\alpha\beta} = \int dU_\perp (U_\perp)_{\alpha\beta} (U_\perp)_{\gamma\delta} = 0$.

The only non-zero terms can be written as

$$C_S := (\Pi_S \otimes \Pi_S) C (\Pi_S \otimes \Pi_S) \quad (5.30)$$

$$V := \int dU_\perp (\Pi_S \otimes \Pi_S) C (U_\perp^T \otimes U_\perp^*) = |v\rangle\langle\Phi_\perp| \quad \text{with } |v\rangle := \sum_{ij,\beta} C_{\beta\beta;ij}^* |ij\rangle \quad (5.31)$$

$$V^\dagger := \int dU_\perp (\bar{U}_\perp \otimes U_\perp) C (\Pi_S \otimes \Pi_S) = |\Phi_\perp\rangle\langle v| \quad (5.32)$$

$$A := \int dU_\perp (\Pi_S \otimes U_\perp) C (\Pi_S \otimes U_\perp^*) = A_S \otimes \Pi_\perp \quad \text{where } A_S = \sum_{ij} \left(\sum_\alpha C_{i\alpha;j\alpha} \right) |i\rangle\langle j| \quad (5.33)$$

$$B := \int dU_\perp (\bar{U}_\perp \otimes \Pi_S) C (U_\perp^T \otimes \Pi_S) = \Pi_\perp \otimes B_S \quad \text{where } B_S = \sum_{ij} \left(\sum_\alpha C_{\alpha i;\alpha j} \right) |i\rangle\langle j| \quad (5.34)$$

$$D := \int dU_\perp (\bar{U}_\perp \otimes U_\perp) C (U_\perp^T \otimes U_\perp^*) = a \Pi_\perp \otimes \Pi_\perp + b |\Phi_\perp\rangle\langle\Phi_\perp|. \quad (5.35)$$

where $\Pi_\perp = \sum |\alpha\rangle\langle\alpha|$ is the projector onto $\mathcal{H}_\perp$ and $|\Phi_\perp\rangle = \sum_{\alpha=1}^{d_\perp} |\alpha\alpha\rangle$ is the maximally entangled state in $\mathcal{H}_\perp \otimes \mathcal{H}_\perp$. In Eqs. (5.30-5.35) we used the $U(d)$ group integral [47] $\int dU U_{\alpha\beta} \bar{U}_{\nu\mu} = \frac{1}{d} \delta_{\alpha\nu} \delta_{\beta\mu}$ and the higher order contraction

$$\int dU U_{\alpha\beta} U_{\gamma\epsilon} \bar{U}_{\tau\xi} \bar{U}_{\nu\mu} = \frac{\delta_{\alpha\tau} \delta_{\gamma\nu} \delta_{\beta\xi} \delta_{\epsilon\mu} + \delta_{\alpha\nu} \delta_{\gamma\tau} \delta_{\beta\mu} \delta_{\epsilon\xi}}{d^2 - 1} - \frac{\delta_{\alpha\tau} \delta_{\gamma\nu} \delta_{\beta\mu} \delta_{\epsilon\xi} + \delta_{\alpha\nu} \delta_{\gamma\tau} \delta_{\beta\xi} \delta_{\epsilon\mu}}{d(d^2 - 1)}. \quad (5.36)$$

Notice that the symmetry under $U \otimes \bar{U}$ singles out the invariants $|\Phi^+\rangle\langle\Phi^+|$ and the identity, i.e. the so-called isotropic states [45]. If we define the projector $\mathbb{1}_R = \Pi_\perp \otimes \Pi_\perp - \frac{1}{d_\perp} |\Phi_\perp\rangle\langle\Phi_\perp|$, we can further decompose the block D in (5.35) in two diagonal blocks and write the twirled Choi matrix in the basis

$$C = \begin{pmatrix} \boxed{C_S} & |v\rangle\langle\Phi_\perp| & & & \\ & \boxed{a' |\Phi_\perp\rangle\langle\Phi_\perp|} & & & \\ & & \boxed{b' \mathbb{1}_R} & & \\ & \mathbf{0} & & \boxed{A} & \\ & & & & \boxed{B} \end{pmatrix} \quad (5.37)$$

Now we recall that a canonical set of Kraus operators can be obtained from the eigenstates of C . In particular, the eigenstates corresponding block $A = A'_S \otimes \Pi_\perp$ can be written as $|\psi_S\rangle |\varphi_\perp\rangle$ with $|\psi_S\rangle \in \mathcal{H}_S$ and $|\varphi_\perp\rangle \in \mathcal{H}_\perp$, which correspond to Kraus operators of the form $E_A = |\psi_S\rangle\langle\varphi_\perp|$. Similarly the Kraus operator corresponding to the last block, B , are of the form $E_B = |\varphi'_\perp\rangle\langle\psi'_S|$, those the central block $E_R = \sum R_{\alpha\beta} |\alpha\rangle\langle\beta|$, and, finally, the first block's eigenstates must to be of the form $|E_o\rangle = \sum_{ij} c_{ij} |ij\rangle + c |\Phi_\perp\rangle$ corresponding to a Kraus operator of the form $E_o = E_S \oplus c \Pi_\perp$.

To conclude the proof, we need to show that we can restrict to Kraus operators of the latter form. We first note that $\text{Tr}(E_A) = \text{Tr}(E_B) = 0$ and that $\text{Tr}(E_R \rho) = 0$, and hence, E_A , E_B and E_R do not contribute to the non-trivial terms in the cost, $\sum_E \Re(\text{Tr}[E^*] \text{Tr}[E \rho])$, as given in Eq. (5.15). Moreover, these terms do not contribute to ensuring $\mathcal{E}(\rho) = \sum_E E \rho E^* = \sigma$ as either the range or the image lies in the orthogonal subspace $\mathcal{H}_\perp$. Notably, since each term is positive semidefinite, any leakage from the support induced by one term cannot be canceled by another. Finally, the completeness relation $\sum_E E^* E = \mathbb{1}$ can be satisfied solely with Kraus operators of the form E_o since $E_o^* E_o = E_S^* E_S \oplus c^2 \Pi_\perp$, without adversely affecting the cost. Note that this amounts to a choice of Choi matrix that lives entirely in first block of Eq. (5.37), with $a' = 1$ so that $\text{Tr}_B C = \mathbb{1}$. $\square$

We next show that the optimization of the cost when the input and output are embedded in a larger Hilbert space can be written in terms of a quantum map acting only on the joint support of input and output.

Theorem 5.7. *Let $\rho = \sum_i p_i |i\rangle\langle i|$ and σ have joint support $\mathcal{H}_S \subseteq \mathcal{H} = \mathcal{H}_S \oplus \mathcal{H}_\perp$; with orthonormal basis $B_S = \{|i\rangle\}_{i=1}^n$ of $\mathcal{H}_S$ and $B_\perp = \{|\alpha\rangle\}_{\alpha=1}^{d_\perp}$ of $\mathcal{H}_\perp$, and $d = n + d_\perp$. The optimal unitarily invariant cost is*

$$\mathcal{K}(\rho, \sigma) = 1 - \frac{1}{d} \max_{\{E_k\}} \left(\Re \left(\sum_k \text{Tr}[E_k \rho] \text{Tr}[E_k^*] \right) + d_\perp \sqrt{\sum_k |\text{Tr}(E_k \rho)|^2} \right) \quad (5.38)$$

$$= 1 - \frac{1}{d} \max_{\mathcal{E}} \left(\Re \left(\sum_{ij} p_i \langle i | \mathcal{E}(|i\rangle\langle j|) | j \rangle \right) + d_\perp \sqrt{\sum_{ij} p_i p_j \langle i | \mathcal{E}(|i\rangle\langle j|) | j \rangle} \right), \quad (5.39)$$

where the maximisation is over CPTP maps $\mathcal{E}(\bullet) = \sum_k E_k \bullet E_k^*$ on $\mathcal{B}(\mathcal{H}_S)$ s.t. $\mathcal{E}(\rho) = \sigma$. Equivalently, we can write

$$\mathcal{K}(\rho, \sigma) = 1 - \frac{1}{d} \max_{C_S} \left(\Re(\langle \rho | C_S | \Phi_S \rangle) + d_\perp \sqrt{\langle \rho | C_S | \rho \rangle} \right) \quad (5.40)$$

s.t. $C_S \geq 0$, $\text{Tr}_B C_S = \mathbb{1}_S$, $\text{Tr}[\rho^T C_S] = \sigma$,

where $|\Phi_S\rangle = \sum_{i=1}^n |ii\rangle$, $|\Phi_\perp\rangle = \sum_{\alpha=1}^{d_\perp} |\alpha\alpha\rangle$, and the input is written in vectorized form $|\rho\rangle = \sum_{ii} p_i |ii\rangle$.

Proof. From Lemma 5.6, we can take Kraus operators to be of the form $E_k = E_{kS} \oplus c_k \Pi_\perp$, such that $\{E_{kS}\}$ is a set of Kraus operators restricted to the support $\mathcal{H}_S$ and c_k form a unit vector. The non-trivial part of the cost, starting from Eq. (5.15), then is

$$\Re \left(\sum_k \text{Tr} [\rho E_k] \text{Tr} [E_k^*] \right) = \Re \left(\sum_k \text{Tr} [\rho E_{kS}] \text{Tr} [E_{kS}^*] + d_\perp \sum_k c_k^* \text{Tr} [\rho E_{kS}] \right). \quad (5.41)$$

We can remove the real part on the second term due to the phase freedom of each c_k with respect to E_{kS} . This freedom will allow us to tune the phase of c_k in each Kraus operator such that $c_k^* \text{Tr} [\rho E_{kS}] = |c_k| |\text{Tr} [\rho E_{kS}]|$, which is the maximum achievable real part.

Now fix a set of Kraus operators on the support $\{E_{kS}\}$. Consider the vectors $\mathbf{u} = (c_k)$ and $\mathbf{v} = (\text{Tr} [\rho E_{kS}])$. With this notation we are maximising the inner product between $\mathbf{v}$ and a unit vector $\mathbf{u}$. The Cauchy-Schwartz inequality states that $|\langle \mathbf{u} | \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| = \|\mathbf{v}\|$ and that the inequality is tight if and only if $\mathbf{v}$ and $\mathbf{u}$ are linearly dependent. Therefore the non-trivial part of the cost is

$$\Re \left(\sum_k \text{Tr} [\rho E_{kS}] \text{Tr} [E_{kS}^*] + d_\perp \sqrt{\sum_k |\text{Tr} [\rho E_{kS}]|^2} \right), \quad (5.42)$$

where $\sqrt{\sum_k |\text{Tr} [\rho E_{kS}]|^2} = \sqrt{\langle \mathbf{v} | \mathbf{v} \rangle}$ is the norm of $\mathbf{v}$ and $c_k = \frac{\text{Tr} [\rho E_{kS}]}{\sqrt{\sum_{k'} |\text{Tr} [\rho E_{k'S}]|^2}}$. Finally, we can take the maximum over all admissible channels to obtain the optimal channel. The equivalent equations Eq. (5.39) and Eq. (5.40) follow immediately from the general expressions Eq. (5.14) and Eq. (5.12), respectively. $\square$

Proposition 5.8. Consider two pure states in a d -dimensional Hilbert space $\mathcal{H}$, w.l.o.g. $\rho = |0\rangle\langle 0|$ and $\sigma = |\psi\rangle\langle \psi|$ where $|\psi\rangle = \alpha |0\rangle + \sqrt{1-\alpha^2} |1\rangle$ with $\alpha \geq 0$. Let $\hat{K}_0$ be the cost matrix. Then

$$\mathcal{K}(\rho, \sigma) = (1-\alpha)(d+2\alpha) \quad (5.43)$$

and the optimal channel is given by conjugation with the unitary

$$U = \begin{bmatrix} \alpha & -\sqrt{1-\alpha^2} \\ \sqrt{1-\alpha^2} & \alpha \end{bmatrix} \oplus \mathbb{1}_{d-2}. \quad (5.44)$$

Proof. Let $\{E_k\}$ be an admissible set of Kraus operators for ρ and σ . Because σ is pure, $E_k \rho E_k^* = p_k \sigma$, with p_k a probability distribution. From Lemma 5.6 and Theorem 5.7, we can write E_k as follows

$$E_k = \sqrt{p_k} \left[\begin{array}{cc|c} \alpha & \gamma_k & 0 \\ \beta & \eta_k & 0 \\ \hline 0 & & \Pi_\perp \end{array} \right], \quad (5.45)$$

for some $\gamma_k, \eta_k \in \mathbb{C}$ and where we defined $\beta = \sqrt{1-\alpha^2} \geq 0$. The trace-preserving condition $\sum_k E_k^* E_k = \mathbb{I}$ gives, from the lower-right entry of the 2×2 block,

$$\sum_k p_k (|\gamma_k|^2 + |\eta_k|^2) = 1, \quad (5.46)$$

and, from the off-diagonal entry,

$$\sum_k p_k (\alpha \gamma_k + \sqrt{1-\alpha^2} \eta_k) = 0. \quad (5.47)$$

Taking real parts, we define

$$\gamma_R := \sum_k p_k \Re(\gamma_k), \quad \eta_R := \sum_k p_k \Re(\eta_k). \quad (5.48)$$

So that

$$\alpha\gamma_R + \beta\eta_R = 0. \quad (5.49)$$

For the cost we want to maximise, from Theorem 5.7,

$$\begin{aligned} \Re \left(\sum_k \text{Tr} [E_{kS} \rho] \text{Tr} [E_{kS}^*] \right) + d_\perp \sqrt{\sum_k |\text{Tr}(E_{kS} \rho)|^2} &= \Re \left(\sum_k p_k \alpha (\alpha + \bar{\eta}_k) + d_\perp \sqrt{\sum_k p_k \alpha^2} \right) \\ &= f(\alpha, d_\perp) + \alpha\eta_R, \end{aligned} \quad (5.50)$$

where $f(\alpha, d_\perp) = \alpha(\alpha + d_\perp)$ is a function of α and $d_\perp$ only, and therefore is a constant in the optimisation. Since α is positive, thus the optimisation reduces to maximising η_R defined in (5.48) subject to (5.46) and (5.47). Since $x \mapsto x^2$ is convex, Jensen's inequality yields

$$\begin{aligned} \gamma_R^2 + \eta_R^2 &= \left(\sum_k p_k \Re(\gamma_k) \right)^2 + \left(\sum_k p_k \Re(\eta_k) \right)^2 \\ &\leq \sum_k p_k \left[(\Re(\gamma_k))^2 + (\Re(\eta_k))^2 \right] \leq \sum_k p_k (|\gamma_k|^2 + |\eta_k|^2) = 1, \end{aligned} \quad (5.51)$$

where the last equality follows from the trace-preserving condition (5.46). Combining this with (5.47), which for $\alpha > 0$ reads,

$$\gamma_R = -\frac{\beta}{\alpha}\eta_R, \quad (5.52)$$

we get

$$\left(1 + \frac{\beta^2}{\alpha^2} \right) \eta_R^2 = \frac{1}{\alpha^2} \eta_R^2 \leq 1 \Rightarrow \eta_R \leq \alpha. \quad (5.53)$$

For $\alpha = 0$, (5.47) gives $\eta_R = 0$, so the bound holds in all cases. Therefore,

$$f(\alpha, d_\perp) + \alpha\eta_R \leq f(\alpha, d_\perp) + \alpha^2 = \alpha(2\alpha + d_\perp) \quad (5.54)$$

Hence,

$$\mathcal{K}(\rho, \sigma) = d - (f(\alpha, d_\perp) + \alpha\eta_R) \leq d - \alpha(2\alpha + d - 2) = (1 - \alpha)(d + 2\alpha) \quad (5.55)$$

As seen in Example 5.5, this bound is achieved by the unitary

$$U = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} \oplus \Pi_\perp, \quad (5.56)$$

for which $U|0\rangle = |\psi\rangle$ and $\eta_R = \alpha$, completing the proof. $\square$

5.1 Commuting density matrices

In this section we continue the study of the unitary invariant cost problem by adding an extra constraint, that is that the states commute: $[\rho, \sigma] = 0$. First we show that the cost between arbitrary states can be bounded by the cost between the first states and the pinching of the second in the basis of the first, which will yield commuting states. Then we see that the optimal quantum transport cost between commuting states can be analytically calculated and that the optimal map is purely quantum. We also bound the cost between commuting states given by classical maps and show it is much larger than the general quantum cost in general.

Proposition 5.9. Let $\mathcal{H}$ be a finite dimensional Hilbert space, $\rho, \sigma \in S(\mathcal{H})$ with ρ diagonal in the basis $\{|i\rangle\}$ and $\mathcal{E}_\rho$ the pinching map in this basis, $\mathcal{E}_\rho(x) = \sum_i \langle i|x|i\rangle |i\rangle\langle i|$. Let $\mathcal{K}$ be the quantum optimal transport cost associated to the unitary invariant cost matrix K_0 . Then

$$\mathcal{K}(\rho, \sigma) \geq \mathcal{K}(\rho, \mathcal{E}_\rho(\sigma)). \quad (5.57)$$

Proof. Let $\rho = \sum_i p_i |i\rangle\langle i|$ and consider the Choi matrix C associated to a channel $\mathcal{E}$ such that $\mathcal{E}(\rho) = \sigma$. This matrix will be $C = \left(\sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)\right)$. The diagonal elements of this matrix, which need to be positive, are $\langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle$ and fulfil

$$\sum_j \langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle = \text{Tr} [\mathbb{1}\mathcal{E}(|i\rangle\langle i|)] = 1. \quad (5.58)$$

Thus these form a classical stochastic map $p(j|i) = \langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle$. Also note that because $\mathcal{E}(\rho) = \sigma$ the Choi matrix must fulfil $\sigma = \text{Tr}_A [\rho C] = \sum_i p_i \mathcal{E}(|i\rangle\langle i|)$. We can apply the pinching $\mathcal{E}_\rho$ to this equation to obtain

$$\mathcal{E}_\rho(\sigma) = \sum_i p_i \mathcal{E}_\rho(\mathcal{E}(|i\rangle\langle i|)) = \sum_{ij} p_i \langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle |j\rangle\langle j| = \sum_{ij} p_i p(j|i) |j\rangle\langle j|, \quad (5.59)$$

which will be useful later.

We can now bound the cost associated to each channel with an expression of the associated classical stochastic map. Note that $\langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle$ are the off diagonal elements of C that complete a 2×2 minor with $\langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle$ and $\langle j|\mathcal{E}(|j\rangle\langle j|)|j\rangle$. Therefore,

$$\Re(\langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle) \leq \sqrt{\langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle \langle j|\mathcal{E}(|j\rangle\langle j|)|j\rangle} = \sqrt{p(i|i)p(j|j)}. \quad (5.60)$$

Finally, with Eq. (5.60) we can calculate the bound on the non-trivial part of the cost associated to an admissible Choi matrix. Let $|\Phi_+\rangle = \sum_i |ii\rangle$ again be the unnormalised maximally mixed state, then:

$$\begin{aligned} \text{Tr} [|\Phi_+\rangle\langle\Phi_+| (\rho \star C)] &= \sum_{i'j'ij} \langle i'i'| \frac{p_i + p_j}{2} (|i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)) |j'j'\rangle \\ &= \sum_{ij} \frac{p_i + p_j}{2} \langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle = \sum_i p_i \langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle + \sum_{i \neq j} \frac{p_i + p_j}{2} \langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle \\ &= \sum_i p_i \langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle + \sum_{i < j} (p_i + p_j) \Re(\langle i|\mathcal{E}(|i\rangle\langle j|)|j\rangle) \\ &\leq \sum_i p_i p(i|i) + \sum_{i < j} (p_i + p_j) \sqrt{p(i|i)p(j|j)} = \sum_i p_i p(i|i) + \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)}. \end{aligned} \quad (5.61)$$

Fix an admissible channel between ρ and σ and its associated classical stochastic map $p(j|i)$. Let $C_p = |\phi\rangle\langle\phi| + \sum_{i \neq j} p(j|i) |ij\rangle\langle ij|$ with $|\phi\rangle = \sum_i \sqrt{p(i|i)} |ii\rangle$. This is clearly positive and

$$\text{Tr}_B [C_p] = \text{Tr}_B \left[|\phi\rangle\langle\phi| + \sum_{i \neq j} p(j|i) |ij\rangle\langle ij| \right] = \sum_i p(i|i) |i\rangle\langle i| + \sum_{i \neq j} p(j|i) |i\rangle\langle i| = \mathbb{1}, \quad (5.62)$$

$$\text{Tr}_A [\rho^T C_p] = \text{Tr}_A \left[\sum_k p_k (|k\rangle\langle k| \otimes \mathbb{1}) \left(|\phi\rangle\langle\phi| + \sum_{i \neq j} p(j|i) |ij\rangle\langle ij| \right) \right] \quad (5.63)$$

$$= \sum_i p_i p(i|i) |i\rangle\langle i| + \sum_{i \neq j} p_i p(j|i) |j\rangle\langle j| = \sum_{ij} p(j|i) p_i |j\rangle\langle j| = \mathcal{E}_\rho(\sigma), \quad (5.64)$$

where the last equality was seen in Eq. (5.59). Therefore C_p is a Choi matrix with associated channel such that $\mathcal{E}_{C_p}(\rho) = \mathcal{E}_\rho(\sigma)$. The elements $\langle i|\mathcal{E}_{C_p}(|i\rangle\langle j|)|j\rangle$ are

$$\langle i|\mathcal{E}_{C_p}(|i\rangle\langle j|)|j\rangle = \langle i|\text{Tr}_A \left[(|j\rangle\langle i| \otimes \mathbb{1}) \left(|\phi\rangle\langle\phi| + \sum_{i' \neq j'} p(j|i') |i'j'\rangle\langle i'j'| \right) \right] |j\rangle \quad (5.65)$$

$$= \langle i|\text{Tr}_A [(|j\rangle\langle i| \otimes \mathbb{1}) |\phi\rangle\langle\phi|] |j\rangle = \sqrt{p(i|i)p(j|j)}, \quad (5.66)$$

which is the tight version of Eq. (5.60). This means the bound (5.61) can be made tight for every admissible classical stochastic map between ρ and σ in the problem between ρ and $\mathcal{E}_\rho(\sigma)$ by choosing the adequate channel C_p . In particular, we can tighten this bound in the problem between ρ and $\mathcal{E}_\rho(\sigma)$ for a classical stochastic map associated to an optimal channel between ρ and σ , thus yielding

$$\mathcal{K}(\rho, \sigma) \geq d - \sum_i p_i p(i|i) + \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)} = \text{Tr} \left[K_0(\rho \star C_p^{TA}) \right] \geq \mathcal{K}(\rho, \mathcal{E}_\rho(\sigma)), \quad (5.67)$$

finalising the proof. $\square$

Proposition 5.10. *Let ρ and σ commute. In a common diagonal basis they can be written as $\rho = \sum_i p_i |i\rangle\langle i|$, $\sigma = \sum_i q_i |i\rangle\langle i|$. Let $\tilde{K}_0$ be the cost matrix. Then*

$$\mathcal{K}(\rho, \sigma) = \frac{1}{d} \left(d - \sum_{ij} p_i \sqrt{\min\{1, \frac{q_i}{p_i}\} \min\{1, \frac{q_j}{p_j}\}} \right) \quad (5.68)$$

Proof. We have seen in the proof of Proposition 5.9 that for every admissible channel there is an associated stochastic map and that for each stochastic map with an associated channel there is a channel that makes Eq. (5.61) tight. Therefore the problem is equivalent to the following optimisation over classical stochastic maps:

$$\min_{p(j|i)} d - \sum_i p_i p(i|i) - \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)}, \quad (5.69)$$

such that $q_j = \sum_i p(j|i)p_i$. Note that only the diagonal terms of the classical stochastic map contribute to the cost and we want to maximise them. This is equivalent to the well known problem of writing the total variation distance as a classical optimal transport problem (see [4, Section 6], also [48, Example 4.14] for a detailed explanation of the total variation distance as a classical transport problem). We succinctly give the solution to this problem for completion.

The maximum value of $0 \leq p(i|i) \leq 1$ subject to $q_i = \sum_j p(i|j)p_j \geq p(i|i)p_i$, is $p(i|i) = \frac{q_i}{p_i} \leq 1$ if $p_i \geq q_i$ and $p(i|i) = 1$ if $p_i < q_i$, or more succinctly

$$p(i|i) = \min\{1, \frac{q_i}{p_i}\}. \quad (5.70)$$

This will maximise the amount of weight the map leaves in place. If $p_i = 0$ then q_i/p_i is not well defined. In this case we take the maximum value $p(i|i) = 1$, as this value appears in the cost and we want to have as large as possible for the optimisation, but it does not appear in the constraint (it is multiplied by $p_i = 0$).

To finalise we need to find the values of the terms $p(j|i)$ for $i \neq j$. If $p(i|i) = 1$, then $p(j|i) = 0$ for $i \neq j$, as the sum of these positive terms is 1. When $p(i|i) < 1$, given that the terms $p(j|i)$, $i \neq j$, do not contribute to the cost, the optimal transport plan $p(j|i)$ can take a variety of values as long as the admissibility of the overall coupling, given by the condition $q_j = \sum_i p(j|i)p_i$, is preserved. $\square$

Remark 5.11. The optimal channel associated to the unitary invariant quantum optimal transport problem between commuting states with common basis $\{|i\rangle\}$ is, as we have seen, in Choi matrix form:

$$C = |\phi\rangle\langle\phi| + \sum_{i \neq j} p(j|i) |ij\rangle\langle ij|, \quad (5.71)$$

with $|\phi\rangle = \sum_i \sqrt{p(i|i)} |ii\rangle$ and $p(i|i)$ as previously defined in Eq. (5.70), which has rank at most $d^2 - d + 1$.

We can further study the structure of these maps by looking at their Kraus matrices. We have the unnormalised eigenvectors of the Choi matrix: $\{|E_k\rangle\} = \{\sum_i \sqrt{p(i|i)} |ii\rangle; \sqrt{p(j|i)} |ij\rangle, i \neq j\}$, such that $C = \sum_k |E_k\rangle\langle E_k|$. The associated Kraus matrices are the un-vectorised elements:

$$\{E_k\} = \left\{ \sum_i \sqrt{p(i|i)} |i\rangle\langle i|; \sqrt{p(j|i)} |j\rangle\langle i|, i \neq j \right\}. \quad (5.72)$$

These Kraus matrices have the characteristic of not being able to generate coherence, but, in the case of $\sum_i \sqrt{p(i|i)} |i\rangle\langle i|$, not completely destroy it. In the context of the theory of quantum coherence as a resource, these matrices are incoherent operations (IO), but not strictly incoherent (SIO) [49].

Remark 5.12. If ρ and σ commute, we can consider the case where we restrict our channels to classical channels to see how it relates to known classical distances and whether classical maps are optimal in the quantum setting. A quantum map will be classical if its Jamiołkowski matrix is diagonal in a product basis. W.l.o.g. ρ and σ are diagonal in the computational basis and the Jamiołkowski matrix be diagonal in the product of canonical basis. As seen in Remark 2.5 a state over time in this case will be of the form $Q = \sum_{ij} p_{ij} |ij\rangle\langle ij|$, with p_{ij} a joint probability distribution, that is a classical coupling. It is immediate to see that the associated cost with cost matrix $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$ is related to the total variation distance as follows: $\text{Tr} \left[\left(\mathbb{1} - \frac{1}{d}\mathcal{S} \right) Q \right] = \sum_{ij} p_{ij} - \frac{1}{d} \sum_i p_{ii} = 1 - \frac{1}{d} \sum_i p_{ii} = 1 - \frac{1}{d} + \frac{1}{d} \frac{1}{2} |\rho - \sigma| \geq 1 - \frac{1}{d}$.⁷ We obtain the cost in Proposition 5.10 without the term $-\frac{1}{d} \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)}$. The fact that this cost is larger than $1 - \frac{1}{d}$ shows that a large gap can exist between the cost associated to classical channels and the optimal quantum cost, which can go to zero by definition.

5.2 Limit $d \rightarrow \infty$

With the analytical formula of the particular case of commuting states for the normalised cost matrix $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$, we can consider what happens if we embed our finite dimensional states into a larger d dimension system and then take the limit $d \rightarrow \infty$. First, let us rewrite Eq. (5.68). We can split the sum into a sum over i and a sum over j as

$$\frac{1}{d} \left(d - \sum_i p_i p(i|i) - \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)} \right) = \frac{1}{d} \left(d - \left(\sum_i p_i \sqrt{p(i|i)} \right) \left(\sum_j \sqrt{p(j|j)} \right) \right). \quad (5.73)$$

We should address what happens to $p(i|i) = \min\{1, \frac{q_i}{p_i}\}$ when $p_i = 0$. Because $p_i = 0$, $p(i|i)$ does not affect the outcome of applying the map to the relevant state. As the value of $p(i|i)$ does not matter, we take the choice $p(i|i) = 1$.

With this we can calculate the limit. Let ρ, σ be commuting states in a finite dimensional Hilbert space $\mathcal{H}_S$ of dimension n , such that the joint support of ρ, σ is $\mathcal{H}_S$. For a dimension $d \geq n$ we have a finite dimensional Hilbert space and a natural embedding that allows us to consider ρ and σ in this space. We can take a basis in which ρ and σ are diagonal and compute the cost. $\sum_i p_i \sqrt{p(i|i)}$ is fixed regardless of dimension. $\sum_j \sqrt{p(j|j)}$ has a fixed part, the sum of $p(j|j)$ in the support of ρ , and $d - n$ times $p(j|j) = 1$ for the part not in the support, which add up to $d - n$. We call these fixed parts N and M , respectively, and calculate the cost:

$$\begin{aligned} \mathcal{K}_d(\rho, \sigma) &= \frac{1}{d} (d - N (M + d - n)) = 1 - \frac{NM}{d} - N + \frac{Nn}{d} \\ &\xrightarrow{d \rightarrow \infty} 1 - N = 1 - \sum_i p_i \sqrt{p(i|i)}. \end{aligned} \quad (5.74)$$

We can further develop this expression to write it as a function of p_i and q_i , the diagonal elements of ρ and σ , only:

$$\begin{aligned} \mathcal{K}_\infty(\rho, \sigma) &= 1 - \sum_i p_i \sqrt{p(i|i)} = \sum_i p_i (1 - \sqrt{p(i|i)}) = \sum_i \sqrt{p_i} (\sqrt{p_i} - \sqrt{p(i|i)p_i}) \\ &= \sum_{q_i < p_i} \sqrt{p_i} (\sqrt{p_i} - \sqrt{q_i}), \end{aligned} \quad (5.75)$$

where the last equality comes from the definition of the optimal $p(i|i) = \min\{1, \frac{q_i}{p_i}\}$ in Eq. (5.68).

Using Theorem 5.7 the general case is immediate:

⁷We slightly abuse notation here by denoting the classical probability distribution associated to the diagonal of states ρ, σ in the canonical basis by ρ, σ .

Theorem 5.13. Let ρ, σ be states in a finite dimensional Hilbert space $\mathcal{H}_S$ of dimension n , such that the joint support of ρ, σ is $\mathcal{H}_S$. Let $d \geq n$ and $\mathcal{H}_d = \mathcal{H}_S \oplus \mathcal{H}_\perp$ be a finite dimensional Hilbert space of dimension d . In $\mathcal{H}_d$, consider the cost matrix $\tilde{K}_d = \mathbb{1}_d - \frac{1}{d}\mathcal{S}_d$. We denote the cost associated to the embedded ρ, σ in a larger Hilbert space with cost matrix $\tilde{K}_d$ as $\mathcal{K}_d(\rho, \sigma)$. Then,

$$\mathcal{K}_\infty(\rho, \sigma) = \lim_{d \rightarrow \infty} \mathcal{K}_d(\rho, \sigma) = 1 - \max_{\{E_k\}} \sqrt{\sum_k |\text{Tr}[\rho E_k]|^2} = 1 - \max_{\mathcal{E}: \mathcal{E}(\rho) = \sigma} F(\rho, \mathcal{E}), \quad (5.76)$$

where the maximisation is over all admissible Kraus representations $\{E_k\}$ of channels $\mathcal{E}$ acting on $\mathcal{H}_S$, and where $F(\rho, \mathcal{E}) := \sqrt{\sum_k |\text{Tr}(\rho E_k)|^2}$ denotes the entanglement fidelity (also called channel fidelity) [50, 51] of the state ρ under the channel $\mathcal{E}$.

The entanglement fidelity measures how well the channel $\mathcal{E}$ preserves not only the state ρ , but also its entanglement with a reference system.

Before we give the proof, note that we now have two optimisation problems in the support Hilbert space $\mathcal{H}_S$, the original problem, with cost function equivalent to

$$\sum_k \text{Tr}[E_k \rho] \text{Tr}[E_k^*] \quad (5.77)$$

and the limit problem, with cost function

$$\sqrt{\sum_k |\text{Tr}[\rho E_k]|^2}. \quad (5.78)$$

Both are optimised over the same set of quantum channel. As the costs are non-equivalent, these two problems are different they in general should have different optimal costs and channels. That said, numerical evidence suggests that these channels are equal or at least very close.

Proof. Using Eq. (5.40) in Theorem 5.7 we can immediately obtain the result:

$$\begin{aligned} \mathcal{K}_\infty(\rho, \sigma) &= \lim_{d \rightarrow \infty} 1 - \frac{1}{d} \max_{\{E_k\}} \left(\Re \left(\sum_k \text{Tr}[E_k \rho] \text{Tr}[E_k^*] \right) + (d - n) \sqrt{\sum_k |\text{Tr}[\rho E_k]|^2} \right) \\ &= 1 - \max_{\{E_k\}} \sqrt{\sum_k |\text{Tr}[\rho E_k]|^2}, \end{aligned} \quad (5.79)$$

as claimed. $\square$

As the set of Kraus operators that optimises Eq. (5.76) is *a priori* different to the optimal set of Kraus operators on any finite dimension, it is unclear if $\mathcal{K}_\infty(\rho, \sigma)$ is efficiently computable. In the following Remark we see that we can efficiently compute $\mathcal{K}_\infty(\rho, \sigma)$, as well as the optimal channel, through an SDP on the joint support of ρ, σ .

Remark 5.14. The cost in the limit, $\mathcal{K}_\infty(\rho, \sigma)$, can be efficiently calculated from the SDP

$$\begin{aligned} \max_J \quad & \text{Tr}[\rho \mathcal{S} \rho J] \\ \text{s.t.} \quad & \begin{cases} \text{Tr}_A[\rho J] = \sigma \\ \text{Tr}_B J = \mathbb{1} \\ J^{T_A} \geq 0 \end{cases}, \end{aligned} \quad (5.80)$$

where $S, J \in \mathcal{B}(\mathcal{H}_S \otimes \mathcal{H}_S)$, $\mathcal{H}_S = \text{supp } \rho \cup \text{supp } \sigma$, and we simplified $(\rho \otimes \mathbb{1})$ to ρ .

Before the proof, note that the constraints are the same as in Eq. (3.3), they denote that J is a Jamiołkowski operator such that $\mathcal{E}_J(\rho) = \sigma$.

Proof. We only need to write the objective function in Eq. (5.76) as a function of the Jamiołkowski matrix instead of the Kraus operators. Recall that $J = \sum_k (\mathbb{1} \otimes E_k) \mathcal{S}(\mathbb{1} \otimes E_k^*)$ and that the optimisation in Eq. (5.76) is over Kraus operators $E_k : \mathcal{H}_S \rightarrow \mathcal{H}_S$. Let $\mathcal{S}, J \in \mathcal{B}(\mathcal{H}_S \otimes \mathcal{H}_S)$, then:

$$\begin{aligned} \text{Tr}[\rho \mathcal{S} \rho J] &= \text{Tr} \left[\rho \mathcal{S} \rho \sum_k (\mathbb{1} \otimes E_k) \mathcal{S}(\mathbb{1} \otimes E_k^*) \right] = \sum_k \text{Tr} [\mathcal{S}(\rho \otimes \rho)(E_k^* \otimes E_k) \mathcal{S}] \\ &= \sum_k \text{Tr} [\rho E_k^*] \text{Tr} [\rho E_k] = \sum_k |\text{Tr} [\rho E_k]|^2. \end{aligned} \quad (5.81)$$

Because the square root is monotonic, maximising this quantity is equivalent to maximising the square root, allowing us to efficiently compute $\mathcal{K}_\infty(\rho, \sigma)$ as

$$\mathcal{K}_\infty(\rho, \sigma) = 1 - \sqrt{\max_J \text{Tr} [\rho \mathcal{S} \rho J]}, \quad (5.82)$$

where the optimisation is over Jamiołkowski matrices of channels mapping ρ to σ . $\square$

Remark 5.15. We can see that the general formula for the limit reduces to the commuting case correctly. If $[\rho, \sigma] = 0$, the Kraus operators for the optimal channel are of the form $\{E_k\} = \left\{ \sum_i \sqrt{p(i|i)} |i\rangle\langle i|; \sqrt{p(j|i)} |j\rangle\langle i|, i \neq j \right\}$, as seen in Eq. (5.72). Furthermore, $\text{Tr} [\rho \sqrt{p(j|i)} |j\rangle\langle i|] = 0$ for all i, j because ρ is diagonal and

$$\text{Tr} \left[\rho \sum_i \sqrt{p(i|i)} |i\rangle\langle i| \right] = \sum_i p_i \sqrt{p(i|i)}. \quad (5.83)$$

If we input this single nonzero value into the equation we obtain

$$\mathcal{K}_\infty(\rho, \sigma) = 1 - \sqrt{\left(\sum_i p_i \sqrt{p(i|i)} \right)^2} = 1 - \sum_i p_i \sqrt{p(i|i)}, \quad (5.84)$$

equal to Eq. (5.74).

5.3 Asymmetry and discontinuity of the cost function

We will consider a similar setting as we had, with the symmetric cost matrix $\tilde{K}_0 = \mathbb{1} - \frac{1}{d} \mathcal{S}$, $\rho = |0\rangle\langle 0|$ and $\sigma = (1-p)\rho + p \mathbb{1}/d$. We now consider the symmetry gap: $\mathcal{K}(\rho, \sigma) - \mathcal{K}(\sigma, \rho)$. The cost will be symmetric when this is zero. Fig. 4 shows that in the proposed example, this gap is nonzero for all $\sigma \neq \rho$.

The symmetry gap also shows a discontinuity when σ goes to ρ . This is due to the following: if a quantum channel takes any non pure state to a pure state, this channel must be the replacement channel due to the continuity of quantum channels. Therefore, in our example, $\mathcal{Q}(\sigma, \rho) = \{\sigma \otimes \rho\}$. This is not true anymore for $\sigma = \rho$, since the states over time associated to all the unitary channels that send ρ to itself (including the identity channel) are now feasible. This discontinuity in the feasible set causes a discontinuity in $\mathcal{K}(\sigma, \rho)$, which translates to the symmetry gap. We can see that analytically with an example.

Let $\rho = (1-\varepsilon)|0\rangle\langle 0| + \varepsilon \mathbb{1}/d$ and $\sigma = |+\rangle\langle +|$. If $\varepsilon > 0$, there is a single admissible state over time: $\mathcal{Q} = \rho \otimes \sigma$. The associated cost is

$$\begin{aligned} \mathcal{K}(\rho(\varepsilon), \sigma) &= 1 - \frac{1}{d} \text{Tr} [\mathcal{S}(\rho \otimes \sigma)] = 1 - \text{Tr} [\rho \sigma] \\ &= 1 - \frac{1}{d} \left(\frac{1}{2}(1-\varepsilon) + \varepsilon \frac{1}{d} \right) \xrightarrow{\varepsilon \rightarrow 0} 1 - \frac{1}{2d}. \end{aligned} \quad (5.85)$$

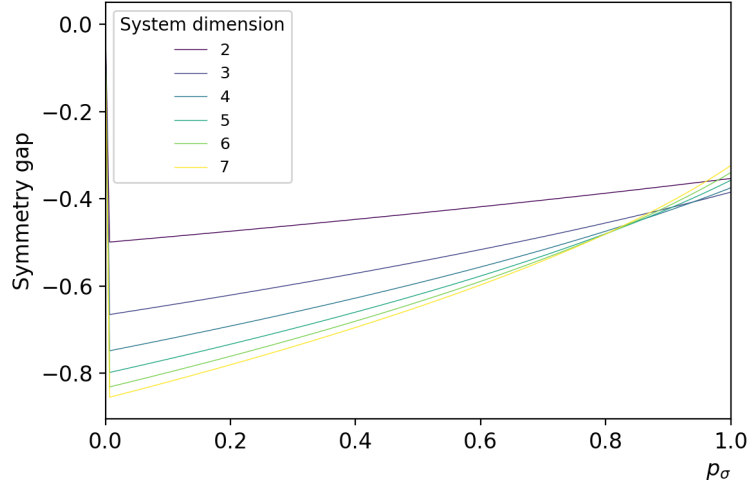


Figure 4: Symmetry gap between for $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$, $\rho = |0\rangle\langle 0|$ and $\sigma = (1 - p_\sigma)\rho + p_\sigma \mathbb{1}/d$.

If we consider $\varepsilon = 0$, ρ and σ are pure and we can use Proposition 5.8 to obtain

$$\mathcal{K}(\rho(0), \sigma) = \frac{1}{d} \left(1 - \frac{1}{\sqrt{2}}\right) \left(d + 2\frac{1}{\sqrt{2}}\right) = 1 - \frac{d + 2 - 2\sqrt{2}}{2d}. \quad (5.86)$$

As $d + 2 - 2\sqrt{2} > 1$ for all $d \geq 2$, we get the strict inequality

$$\mathcal{K}(\rho(0), \sigma) < \lim_{\varepsilon \searrow 0} \mathcal{K}(\rho(\varepsilon), \sigma). \quad (5.87)$$

6 Conclusions and open problems

We have introduced a formulation of optimal transport cost for quantum states as an application of the formalism of states over time (*stotes*), in an attempt to base it on a notion of cost bilinear in the initial quantum state (mass distribution) and quantum channel (transport plan). This formalism was introduced to expand on our current understanding of spatial quantum correlations, expressed in joint density matrices, to incorporate temporal correlations induced by a given time evolution. In it, stotes are Jordan products of density matrices with Jamiolkowski matrices of quantum channels. This has allowed us to define a formalism of optimal transport with a straightforward physical interpretation for couplings, albeit outside the realm of density matrices. Despite the clear interpretation of the couplings, and the mathematical properties observed by the stote and state-channel cost function κ , a physical meaning of the optimal cost remains a crucial open question in this setting.

After introducing the necessary notions, we set out to explore the new definition of cost, in particular in view of the possibility of obtaining interesting metrics on the set of quantum states. The biggest open problem we, and in fact the stote formalism as a whole, face is that there is currently no concise characterisation of the convex hull of the set of stotes, nor of the convex cone generated by it, nor the dual cone. The latter cone encodes all information required in the selection of a suitable cost operator: it should be in the dual cone of stotes, and the same dual cone plays an important role in deciding the triangle inequality of a given cost. Our original motivation was to be able to design cost matrices that can be interpreted in physical terms, such as showing energy differences for a given Hamiltonian. Currently this is work in progress. The stote cone itself enters in each attempt of calculating the optimal cost for a given cost operator. However, at least fixing cost operator as well as initial and final density matrix, this optimisation is an SDP.

As a case study and because of its distinguished symmetry, we have investigated in detail the unitary invariant cost, which is analogous to the trivial metric in the classical case. We have calculated this

cost for commuting states and pure states. These examples have allowed us to observe some properties and facts regarding our formalism.

A surprising fact can be observed from the second item in Example 5.5. One of our main motivations for this formalism was the linearity of the state over time with respect to both the initial state and the channel. Other approaches to quantum transport [8, 9] led to the observation that to save the triangle inequality, a square root of the cost had to be taken. Likewise, from the example one notices that our cost behaves like the square of a distance, indicating that the square root would be necessary here, too, to preserve the triangle inequality. Nothing similar has been observed in the classical case, where the roots only appear when taking powers of distances as cost functions (as seen in the Wasserstein distances [52, 53]). This contrast motivates us to conjecture that the appearance of a square root to preserve the triangle inequality could be a quantum feature of optimal transport costs.

Other features of our optimal quantum transport cost are that even when the cost operator is exchange symmetric (as the unitary invariant is), the resulting optimal cost is not necessarily symmetric, adding to doubts that this approach can yield meaningful metrics on states. On top of that, the examples of asymmetry exhibit even instances of discontinuity of the optimal cost in the two states.

The second item of Example 5.5 allows for another observation. In every dimension, we considered the initial/final states $\rho = |0\rangle\langle 0|$, $\sigma = (\alpha |0\rangle + \sqrt{1 - \alpha^2} |1\rangle)(\alpha \langle 0| + \sqrt{1 - \alpha^2} \langle 1|)$. The optimal channels turn out to be the conjugation by unitaries having a block structure: a direct sum of a 2×2 -unitary and an identity of rank $d - 2$. Only the first summand is relevant to joint support of ρ and σ , but the cost is a function of the dimension nonetheless, as seen in Eq. (5.24). The calculation of the optimal unitary also shows that the cost for a given channel is sensitive to the behaviour of the channel outside the space spanned by the relevant states. In fact, this is a general feature for arbitrary states. This contrasts classical optimal transport, where the behaviour of the channel on regions where the input probability is zero has no effect on the cost. This feature is reminiscent of the Aharonov-Bohm effect [54], a purely quantum effect where the magnetic field far away from a charged particle can affect interference fringes of its wave function. Motivated by this, in Section 5.2 we considered the limit of larger and larger ambient Hilbert spaces, for a given pair of states. This leads to a certain renormalisation of the cost (always in the unitary invariant case), in particular in the limit we obtain a formula for the optimal transport cost that manifestly “feels” only the supports of the two states. It remains for future investigation to determine whether this leads to well-behaved metrics with interesting properties.

Acknowledgments

The first author thanks Marco Fanizza, Giulio Gasbarri, Niklas Galke, Carlo Marconi and Santiago Llorens for helpful discussions. The authors collectively thank Sheila Folsom for listening to our problems about states over time and for showing us how to keep an open mind. The authors are supported by the Spanish MICIN (project PID2022-141283NB-I00) with the support of FEDER funds, by the Spanish MICIN with funding from European Union NextGenerationEU (project PRTR-C17.I1) and the Generalitat de Catalunya, and by the Spanish MTDFP through the QUANTUM ENIA project: Quantum Spain, funded by the European Union NextGenerationEU within the framework of the “Digital Spain 2026 Agenda”. MHR furthermore acknowledges support by the Generalitat de Catalunya through a FI-AGAUR scholarship. JCC furthermore acknowledges support from an ICREA Academia Award. AW furthermore acknowledges support by the European Commission QuantERA grant ExTRaQT (Spanish MICIN project PCI2022-132965), by the Alexander von Humboldt Foundation, and by the Institute for Advanced Study of the Technical University Munich.

The numerical calculations were performed with Python, using the package *cvxpy* [55] and the solver *mosek* [56]. The package *qutip* [57, 58] has been used for manipulation of quantum objects.

Author contributions

All authors contributed to the research and writing of the manuscript. No generative AI has been used in the research or writing of this paper.

References

- [1] Gaspard Monge. “Mémoire sur la théorie des déblais et des remblais”. De l’Imprimerie Royale. Paris (1781).
- [2] Leonid V. Kantorovich. “On a problem of Monge”. *Uspekhi Mat. Nauk.* **3**, 225–226 (1948).
- [3] Leonid V. Kantorovich. “On the Translocation of Masses”. *Management Science* **5**, 1–4 (1958).
- [4] Cédric Villani. “Optimal transport: Old and new”. *Volume 338 of Grundlehren der mathematischen Wissenschaften*. Springer Verlag. Berlin, Paris, Dakar (2008).
- [5] Karol Życzkowski and Wojciech Słomczyński. “The Monge distance between quantum states”. *Journal of Physics A: Mathematical and General* **31**, 9095–9104 (1998).
- [6] Karol Życzkowski and Wojciech Słomczyński. “The Monge metric on the sphere and geometry of quantum states”. *Journal of Physics A: Mathematical and General* **34**, 6689–6722 (2001).
- [7] Matt Hoogsteder-Riera. “A transport approach to distances in quantum systems”. Bachelor’s thesis. Universitat Autònoma de Barcelona, Department of Physics. (2018).
- [8] Giacomo De Palma and Dario Trevisan. “Quantum Optimal Transport with Quantum Channels”. *Annales Henri Poincaré* **22**, 3199–3234 (2021).
- [9] Shmuel Friedland, Michał Eckstein, Sam Cole, and Karol Życzkowski. “Quantum Monge-Kantorovich Problem and Transport Distance between Density Matrices”. *Physical Review Letters* **129**, 110402 (2022).
- [10] Rafał Bistoń, Michał Eckstein, and Karol Życzkowski. “Monotonicity of a quantum 2-Wasserstein distance”. *Journal of Physics A: Mathematical and Theoretical* **56**, 095301 (2023).
- [11] Gergely Bunth, József Pitrik, Tamás Titkos, and Dániel Virosztek. “Wasserstein distances and divergences of order p by quantum channels”. [arXiv\[math-ph\]:2501.08066](https://arxiv.org/abs/2501.08066) (2025).
- [12] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. “The Quantum Wasserstein Distance of Order 1”. *IEEE Transactions on Information Theory* **67**, 6627–6643 (2021).
- [13] Giacomo De Palma and Dario Trevisan. “The Wasserstein Distance of Order 1 for Quantum Spin Systems on Infinite Lattices”. *Annales Henri Poincaré* **24**, 4237–4282 (2023).
- [14] Jean-David Benamou and Yann Brenier. “A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem”. *Numerische Mathematik* **84**, 375–393 (2000).
- [15] Eric A. Carlen and Jan Maas. “An Analog of the 2-Wasserstein Metric in Non-commutative Probability under which the Fermionic Fokker-Planck Equation is Gradient Flow for the Entropy”. *Communications in Mathematical Physics* **331**, 887–926 (2014).
- [16] Cambyse Rouzé and Nilanjana Datta. “Concentration of quantum states from quantum functional and transportation cost inequalities”. *Journal of Mathematical Physics* **60**, 012202 (2019).
- [17] Emanuele Caglioti, François Golse, and Thierry Paul. “Towards Optimal Transport for Quantum Densities”. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze* **24**, 1981–2045 (2023).
- [18] Andrzej Jamiołkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. *Reports on Mathematical Physics* **3**, 275–278 (1972).
- [19] Man-Duen Choi. “Completely positive linear maps on complex matrices”. *Linear Algebra and its Applications* **10**, 285–290 (1975).

- [20] Dominic Horsman, Chris Heunen, Matthew F. Pusey, Jonathan Barrett, and Robert W. Spekkens. “Can a quantum state over time resemble a quantum state at a single time?”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **473**, 20170395 (2017).
- [21] Matt S. Leifer and Robert W. Spekkens. “Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference”. *Physical Review A* **88**, 052130 (2013).
- [22] Joseph F. Fitzsimons, Jonathan A. Jones, and Vlatko Vedral. “Quantum correlations which imply causation”. *Scientific Reports* **5**, 18281 (2015).
- [23] James Fullwood and Arthur J. Parzygnat. “On quantum states over time”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **478**, 20220104 (2022).
- [24] James Fullwood. “Quantum dynamics as a pseudo-density matrix”. *Quantum* **9**, 1719 (2025).
- [25] James Fullwood and Arthur J. Parzygnat. “On Dynamical Measures of Quantum Information”. *Entropy* **27**, 331 (2025).
- [26] Seok Hyung Lie and Nelly H. Y. Ng. “Quantum state over time is unique”. *Physical Review Research* **6**, 033144 (2024).
- [27] Pascual Jordan. “Über eine Klasse nichtassoziativer hyperkomplexer Algebren”. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse* **1932**, 569–575 (1932).
- [28] Pascual Jordan. “Über die Multiplikation quantenmechanischer Größen”. *Zeitschrift für Physik* **80**, 285–291 (1933).
- [29] Pascual Jordan. “Über die Multiplikation quantenmechanischer Größen. II”. *Zeitschrift für Physik* **87**, 505–512 (1934).
- [30] Pascual Jordan, John von Neumann, and Eugene Wigner. “On an Algebraic Generalization of the Quantum Mechanical Formalism”. *Annals of Mathematics* **35**, 29–64 (1934).
- [31] Lieven Vandenberghes and Stephen Boyd. “Semidefinite Programming”. *SIAM Review* **38**, 49–95 (1996).
- [32] John Watrous. “Lecture Notes “Theory of Quantum Information”. Lecture 7: Semidefinite Programming” (2011). Accessed March 2025.
- [33] Matteo Scandi, Paolo Abiuso, Jacopo Surace, and Dario De Santis. “Quantum Fisher information and its dynamical nature”. *Reports on Progress in Physics* **88**, 076001 (2025). [arXiv:2304.14984](#).
- [34] Karl Kraus. “States and effects”. *Pages 1–12*. Springer Berlin Heidelberg. Berlin, Heidelberg (1983).
- [35] Michael A. Nielsen and Isaac L. Chuang. “Quantum computation and quantum information”. *Cambridge University Press*. Cambridge (2010).
- [36] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti. “Theoretical framework for quantum networks”. *Physical Review A* **80**, 022339 (2009).
- [37] Stephen Boyd and Lieven Vandenberghes. “Convex Optimization”. *Cambridge University Press*. Cambridge (2004).
- [38] Tomohiro Ogawa and Hiroshi Nagaoka. “Strong converse and Stein’s lemma in quantum hypothesis testing”. *IEEE Transactions on Information Theory* **46**, 2428–2433 (2000).
- [39] Marco Tomamichel. “Quantum information processing with finite resources”. *Springer International Publishing*. (2016).
- [40] Lukasz Skowronek, Erling Størmer, and Karol Życzkowski. “Cones of positive maps and their duality relations”. *Journal of Mathematical Physics* **50**, 062106 (2009).
- [41] Mihály Bakonyi and Hugo J. Woerdeman. “Matrix completions, moments, and sums of hermitian squares”. *Princeton University Press*. Princeton NJ (2011).
- [42] Lawrence Narici and Edward Beckenstein. “Topological vector spaces”. *Chapman and Hall/CRC*. (2010).

- [43] Jacob Biamonte. “Lectures on Quantum Tensor Networks”. arXiv[quant-ph]:1912.10049v2 (2020).
- [44] Eric Beutner. “On the closedness of the sum of closed convex cones in reflexive banach spaces”. *Journal of Convex Analysis* **14**, 99–102 (2007). url: https://www.heldermann-verlag.de/jca/jca14/jca0560_b.pdf.
- [45] Reinhard F. Werner. “Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model”. *Physical Review A* **40**, 4277–4281 (1989).
- [46] Frank Vallentin. “Symmetry in semidefinite programs”. *Linear Algebra and its Applications* **430**, 360–369 (2009).
- [47] Zbigniew Puchała and Jarosław A. Miszczak. “Symbolic integration with respect to the Haar measure on the unitary groups”. *Bulletin of the Polish Academy of Sciences Technical Sciences* **65**, 21–27 (2017).
- [48] Ramon van Handel. “Probability in high dimensions” (2016). url: <https://web.math.princeton.edu/~rvan/APC550.pdf>.
- [49] Alexander Streltsov, Gerardo Adesso, and Martin B. Plenio. “Colloquium: Quantum coherence as a resource”. *Reviews in Modern Physics* **89**, 041003 (2017).
- [50] Benjamin Schumacher. “Sending entanglement through noisy quantum channels”. *Phys. Rev. A* **54**, 2614–2628 (1996).
- [51] John Watrous. “The theory of quantum information”. *Cambridge University Press*. (2018).
- [52] Leonid V. Kantorovich. “Mathematical Methods of Organizing and Planning Production”. *Management Science* **6**, 366–422 (1960).
- [53] Leonid N. Vaserstein. “Markov processes over denumerable products of spaces, describing large systems of automata”. *Problemy Peredachi Informatsii (Probl. Inf. Transm.)* **5**, 64–72 (1969). url: <https://www.mathnet.ru/eng/ppi/v5/i3/p64>.
- [54] Yakir Aharonov and David Bohm. “Significance of Electromagnetic Potentials in the Quantum Theory”. *Physical Review* **115**, 485–491 (1959).
- [55] Steven Diamond and Stephen Boyd. “CVXPY: A Python-embedded modeling language for convex optimization”. *Journal of Machine Learning Research* **17**, 1–5 (2016). url: <http://jmlr.org/papers/v17/15-408.html>.
- [56] MOSEK ApS. “MOSEK Optimizer API for Python 10.2.5”. (2024). url: <https://docs.mosek.com/latest/pythonapi/index.html>.
- [57] Jesper R. Johansson, Paul D. Nation, and Franco Nori. “QuTiP: An open-source Python framework for the dynamics of open quantum systems”. *Computer Physics Communications* **183**, 1760–1772 (2012).
- [58] Jesper R. Johansson, Paul D. Nation, and Franco Nori. “QuTiP 2: A Python framework for the dynamics of open quantum systems”. *Computer Physics Communications* **184**, 1234–1240 (2013).