UAB
Universitat Autònoma de Barcelona

States over time for quantum optimal transport and relative entropy chain rules

Matt Hoogsteder Riera

States over time for quantum optimal transport and relative entropy chain rules

by
Matt Hoogsteder Riera

This page is intentionally left blank

Universitat Autònoma de Barcelona

States over time for quantum optimal transport and relative entropy chain rules

by

Matt Hoogsteder Riera

under the supervision of

Prof. John Calsamiglia
Prof. Andreas Winter

A thesis submitted in partial fulfillment for the degree of
Doctor of Philosophy in Physics

Física Teòrica: Informació i Fenòmens Quàntics
Departament de Física, Facultat de Ciències

Bellaterra, October 11th, 2025

The struggle itself towards the heights is enough to fill a man's heart. One must imagine Sisyphus happy.

— Albert Camus, The Myth of Sisyphus

This page is intentionally left blank

Resum

Aquesta tesi està dividida asimètricament en dues parts: la primera desenvolupa una teoria per transport òptim quàntic, la segona estableix una regla de la cadena per l'entropia relativa quàntica. Aquests dos temes, aparentment independents, es connecten en aquesta tesi mitjançant la noció d'estats en el temps ("states over time"). Aquest concepte té l'objectiu de caracteritzar les propietats de l'evolució de sistemes quàntics, en contrast amb el concepte de "estats en l'espai", més familiar, que descriu les correlacions entre estats espacialment separats. Descriure correlacions quàntiques en el temps resulta ser considerablement més subtil i intricat, cosa que requereix noves estructures matemàtiques per capturar la composició i transformació d'aquests sistemes.

En la primera part, dedicada a *transport òptim quàntic* ("quantum optimal transport"), busquem un anàleg quàntic a la teoria clàssica mitjançant la identificació la codificació adequada dels elements del problema—en particular, l'estat inicial (una matriu densitat) i els plans de transport admissibles (canals quàntics)—en un sol objecte matemàtic anomenat acoblament ("coupling"). Comencem amb una formulació ingènua, inspirada per la formulació clàssica, basada en l'ús d'estats quàntics conjunts com a acoblament. Desenvolupem aquest acoblament i identifiquem les seves peculiars (indesitjables) propietats per motivar la nostra cerca amb un enfocament diferent.

En el nou enfocament, que constitueix el tema principal d'aquest treball, busquem un acoblament bilineal que permeti la composició natural de canals, per tal d'obtenir una noció consistent d'estat en el temps. Amb aquesta restricció, el producte de Jordan emergeix com a noció físicament motivada d'acoblament. Partint d'això, definim el cost òptim de transport quàntic com la minimització d'un funcional de cost sobre tots els canals admissibles que connecten els dos estats i estudiem les

seves propietats físiques i matemàtiques. Parem especial atenció a costos invariants sota accions unitàries, cosa que permet simplificar i resoldre analíticament el problema per estats que commuten o amb altres estructures. Aquests exemples revelen profundes diferències estructurals entre les nocions clàssiques i quàntiques de transport òptim.

En la segona part, dedicada a regles de la cadena per l'entropia relativa quàntica, revisitem una de les mesures de distingibilitat més fonamentals en teoria de la informació quàntica, que destaca per la seva significança operacional en tests d'hipòtesi. En teoria de probabilitat clàssica, l'entropia relativa compleix una regla de la cadena que descompon el canvi total en entropia relativa sota l'acció de dos processos estocàstics en una mitjana divergències de les distribucions puntuals sota l'acció dels processos. Tanmateix, en el cas quàntic només es coneixien versions asimptòtiques, regularitzades d'aquesta regla de la cadena.

En aquesta obra, derivem dues regles de la cadena quàntiques basades en un sol sistema. La primera generalitza la noció clàssica de distribucions puntuals a un conjunt d'estats quàntics construïts per l'acció d'operadors de mesura sobre estats en el temps. La segona manté projectors ortonormals de rang u, però imposa una condició suficient que restringeix la generalitat de l'equació. Junts, aquests resultats mostren que és possible obtenir regles de la cadena significatives en el règim d'un sol sistema, però també ressalten els obstacles fonamentals a una generalització total i la necessitat per desigualtats més ajustades.

Abstract

This thesis is divided asymmetrically into two main parts: the first develops a framework for quantum optimal transport, and the second establishes chain rules for the quantum relative entropy. These two apparently unrelated topics are connected in this thesis by the unifying notion of *states over time*. This concept aims at characterising the properties of quantum systems as they evolve, in contrast to the more familiar"states over space" that describe correlations between spatially separated subsystems. Describing quantum correlations across time proves to be considerably more subtle and intricate, requiring new mathematical structures to capture their composition and transformation.

In the first part, devoted to *quantum optimal transport*, we seek a quantum analogue of the classical theory by identifying an appropriate encoding of the problem's elements—namely, the initial state (a density matrix) and the admissible transport plans (quantum channels)—into a single mathematical object called a coupling. We begin with a naive, classically inspired formulation based on joint quantum states as a coupling. We develop this coupling and note its peculiar (undesirable) issues to motivate our search for a different approach.

In the new approach, which constitutes the main topic of this work, we seek a bilinear coupling that allows for the natural composition of channels, leading to a consistent notion of a state over time. From this requirement, the Jordan product emerges as a physically motivated notion of coupling. On this basis, we define a quantum optimal transport cost as the minimisation of a cost functional over all admissible channels connecting two quantum states and study its mathematical and physical properties. Special attention is given to unitarily invariant costs, which allow for simplifications and fully analytical solutions for commuting or otherwise structured states. These examples reveal deep structural differences between the

classical and quantum notions of optimal transport.

In the second part, devoted to chain rules for the quantum relative entropy, we revisit one of the most fundamental measures of distinguishability in quantum information theory, notable for its operational significance in hypothesis testing. In classical probability theory, the relative entropy satisfies a chain rule that decomposes the total decrease in relative entropy under a pair of stochastic maps into an average of pointwise divergences between mapped distributions. In the quantum setting, however, only asymptotic, regularised versions of this rule were previously known.

In this work, we derive two single-letter quantum chain rule inequalities. The first generalises the classical notion of pointwise distributions to an ensemble of quantum states constructed from the action of measurement operators on states over time. The second retains orthonormal rank-one projectors but imposes a sufficient condition that restricts its full generality. Together, these results show that meaningful single-letter quantum chain rule inequalities are possible, while also highlighting both the fundamental obstructions that prevent a full generalisation of the classical case and the need for tighter inequalities.

Resumen

Esta tesis está dividida asimétricamente en dos partes: la primera desarrolla una teoría para el transporte óptimo cuántico, la segunda establece una regla de la cadena para la entropía relativa cuántica. Estos dos temas, aparentemente independientes, se conectan en esta tesis por la noción de estados en el tiempo ("states over time"). Este concepto tiene el objetivo de caracterizar las propiedades de la evolución de sistemas cuánticos, en contraste con el concepto de "estados en el espacio", más familiar, que describe las correlaciones entre estados espacialmente separados. Describir correlaciones cuánticas en el tiempo resulta ser considerablemente más sutil e intrincado, lo que requiere nuevas estructuras matemáticas para capturar la composición y transformación de estos sistemas.

En la primera parte, dedicada al transporte óptimo cuántico ("quantum optimal transport"), buscamos un análogo cuántico a la teoría clásica mediante la identificación de la codificación adecuada de los elementos del problema — en particular, el estado inicial (una matriz densidad) y los planes de transporte admisibles (canales cuánticos) — en un solo objeto matemático llamado acoplamiento ("coupling"). Comenzamos con una formulación ingenua, inspirada en la formulación clásica, basada en el uso de estados cuánticos conjuntos como acoplamiento. Desarrollamos este acoplamiento e identificamos sus peculiares (indeseables) propiedades para motivar nuestra búsqueda con un enfoque diferente.

En el nuevo enfoque, que constituye el tema principal de este trabajo, buscamos un acoplamiento bilineal que permita la composición natural de canales, para obtener una noción consistente de estado en el tiempo. Con esta restricción, el producto de Jordan emerge como una noción físicamente motivada de acoplamiento. Partiendo de esto, definimos el costo óptimo de transporte cuántico como la minimización de un funcional de costo sobre todos los canales admisibles que conectan

los dos estados y estudiamos sus propiedades físicas y matemáticas. Prestamos especial atención a costos invariantes bajo acciones unitarias, lo que permite simplificar y resolver analíticamente el problema para estados que conmuten o con otras estructuras. Estos ejemplos revelan profundas diferencias estructurales entre las nociones clásicas y cuánticas de transporte óptimo.

En la segunda parte, dedicada a reglas de la cadena para la entropía relativa cuántica, revisamos una de las medidas de distinguibilidad más fundamentales en teoría de la información cuántica, que destaca por su significancia operacional en tests de hipótesis. En teoría de probabilidad clásica, la entropía relativa cumple una regla de la cadena que descompone el cambio total en entropía relativa bajo la acción de dos procesos estocásticos en un promedio de divergencias de las distribuciones puntuales bajo la acción de los procesos. Sin embargo, en el caso cuántico solo se conocían versiones asintóticas, regularizadas de esta regla de la cadena.

En esta obra, derivamos dos reglas de la cadena cuánticas basadas en un solo sistema. La primera generaliza la noción clásica de distribuciones puntuales a un conjunto de estados cuánticos construidos por la acción de operadores de medida sobre estados en el tiempo. La segunda mantiene proyectores ortonormales de rango uno, pero impone una condición suficiente que restringe la generalidad de la ecuación. Juntos, estos resultados muestran que es posible obtener reglas de la cadena significativas en el régimen de un solo sistema, pero también resaltan los obstáculos fundamentales a una generalización total y la necesidad de desigualdades más ajustadas.

This page is intentionally left blank

Declaration

I declare that this thesis has been written by myself and that this works has not been submitted to any other qualification or degree. The work submitted is my own in collaboration with others, which is duly noted in the text. Credit has been given within this work when using other author's work.

List of Publications

[HCW25] Hoogsteder-Riera, M., Calsamiglia, J., and Winter, A., "Approach to optimal quantum transport via states over time", arXiv:2504.04856 (2025).

[HCW25] is currently under review in *Quantum*.

In preparation:

Gasbarri, G., and Hoogsteder-Riera, M., "Chain Inequalities for Quantum Relative Entropy in the Single-Copy Regime".

This page is intentionally left blank

Acknowledgments

I would like to thank my directors, John and Andreas, for their insight and patience during the research and writing for this thesis. I would also like to thank Giulio for his help with my dumb problems and pushing me into the world of entropies. Thank you also to Ilaria for her support during the writing. Thanks to the rest of the senior members of the group, Anna, Emili and Ramon, for helping create a good environment. Thanks to everyone else at the quantum information group in Barcelona: Albert, Arnau, Carlo, Elisabet, Gael, Hari, Jennifer, Joe, Jofre, Jordi, Mani, Marco, Matias, Michalis, Naga, Niklas, Phillip, Toni, Yago and everyone else who I'm forgetting and has been there over the years for their help and friendship.

Gràcies també als meus amics i familiars: en especial als meus pares i germà, i també a Adri, Andreu, Julián, Marta i Neus, pel seu suport incondicional.

Schließlich danke ich den Menschen in Köln, die mich während der Schreibmonate unterstützt haben.

Tyty :)

This page is intentionally left blank

Contents

1

Introduction

Consider the relation

$$p(x, y) = p(y|x)p(x), \tag{1.1}$$

connecting the joint distribution of two variables with the marginal and conditional distributions. This classical equation, usually seen as a step to the proof of Bayes' theorem, a more common form of the same result, provides a significant insight into classical probabilities. Namely, consider an initial probability distribution $p(x)$ and a stochastic map $p(y|x)$. Can we possibly obtain a single object that encodes this? Of course, we could just use the Cartesian product $(p(x), p(y|x))$, which is somewhat a single object. Eq. (1.1) not only answers this question, it gives meaning to this object: the product yields a joint probability distribution on the joint space.

But this is not the only interpretation of $p(y|x)$ and Eq. (1.1). Given a joint probability distribution $p(x, y)$ we can then recover $p(y|x)$ and $p(x)$ from Eq. (1.1). In this case $p(y|x)$ is not a stochastic map, but rather a way to update our knowledge of the probability distribution of y when obtaining information about the state of x.

For a simple example, consider someone we meet regularly wears socks with two properties: colour and length. We know that someone has three pairs of socks that they randomly wear with equal probability: red and long, red and short and blue and long. We now know the joint distribution. Moreover if we can observe that they are wearing blue socks one day we can then guarantee that the socks will be long, $p(\text{long}|\text{blue}) = 1$. There is no map here, not physically that is, just a way for an observer to process information and make better predictions upon obtaining new knowledge.

Conversely, consider a sock factory. We have a machine that is provided with yarn of two possible colours, red and blue, and then selects whether to make a long or short sock with it with an input dependent stochastic maps. With the correct choices of inputs and maps, we can recover through Eq. (1.1) the distribution from the previous example.

This two examples use the same objects in fundamentally different ways, but Eq. (1.1) stays true regardless. It is known that this is not true in quantum mechanics. To be more precise, a joint state ω will have a marginal $\mathrm{Tr}_B\left[\omega\right] = \rho$, but it is now unclear how we can recover a map from this pair of objects. Similarly, if we start with a state ρ and a map $\mathcal{E}$, it is unclear how or if we can obtain a joint state.

These questions have been the subject of study over the last 20 years in works such as [Lei06; LS13; HHP+17; FP22]. We will refer to the quantum generalisations of $p(y|x)p(x)$, as it is done in [HHP+17], as *states over time*. The main contribution of this thesis has been to take these results on states over time and apply them to two different topics: optimal transport and entropy inequalities.

Generalising the theory of classical optimal transport to quantum states was the initial motivation of this work. Classical optimal transport provides a powerful toolbox for defining distinguishability measures between probability distributions that reflect the underlying metric structure of the space in which they are defined. Extending this idea to quantum systems is particularly appealing, as it enables the construction of distance measures between states that go beyond the Hilbert space structure. In this way, one can incorporate physically meaningful notions of separation, such as energy, or informational measures like the Hamming distance between bit strings [Ham50]. For instance, in Hilbert space, two spatial wave packets with disjoint supports are considered maximally distant, regardless of how close or far their supports are in real space. By contrast, an optimal-transport-based notion of distance would naturally reflect their physical spatial separation.

The question of how to represent quantum states over time arises naturally from the generalization of couplings in classical optimal transport. In the classical setting, the notion of a coupling between two distributions $p(x)$ and $p(y)$—a joint distribution $p(x, y)$ whose marginals are $p(x)$ and $p(y)$—plays a central role, as it underlies the definition of transport costs and distances. In the quantum setting, however, we need an object that generalises this idea. This challenge quickly emerged as a central concept in this thesis.

In the context of entropy inequalities, we generalised a known classical formula, the chain rule for the relative entropy, to quantum states. We used two different methods to find three quantum chain rules.

i) First, we used a semiclassical approach and states over time to find a new chain rule bound. This result is interesting for two reasons: first, it improves on the known bounds, given by Fang *et al.* [FFR+20] in some regions and is easier to calculate; second, the methods used in the proof are interesting in their own right. For the proof, we used classical objects derived from quantum processes on certain states over time and then removed the classicality to obtain purely quantum set of relative entropies.

ii) The second set of results relies on a generic entropy inequality derived from a generalisation of well-known recovery maps [Wil15; JRS+18; SBT16]. We then study concrete cases of this inequality to obtain our results. The main result of this part is a conditioned quantum chain rule. We also show the necessity of having a condition.

This thesis is structured in the following way: we will first introduce general mathematical and physical theory in Chapter 2, with an emphasis on presenting quantum theory as an extension of classical probability theory. In Chapter 3 we introduce classical optimal transport, as well as our own and others' initial attempts at a quantum generalisation. In Chapter 4 we will introduce the theory of state over time. In Chapter 5 we lay out our theory of quantum optimal transport. In Chapter 6 we solve the particular case of a unitary invariant costs within within the framework of Chapter 5, and present some preliminary results toward constructing a quantum optimal transport that encodes the energy profile given by a Hamiltonian. Chapter 5 and Chapter 6 are based on [HCW25]. Finally, in Chapter 7 we present our results on chain rules for the quantum relative entropy. Chapter 7 is based on currently unpublished work [GH25].

2

Preliminaries

In this chapter we introduce well-known mathematical and physical concepts that are nonetheless important for this thesis. The aim is to make the thesis as self-contained as possible.

In Section 2.1 we introduce purely mathematical definitions and results; in Section 2.2 we introduce the postulates of quantum mechanics, as well as the basic objects we use throughout this thesis; in Section 2.3 we introduce quantum channels, as well as their multiple formulations, and in Section 2.4 we introduce basic notions of quantum information theory.

2.1 Mathematical concepts and notation

The overall framework of this thesis is finite dimensional quantum mechanics, which is mathematically expressed as finite dimensional Hilbert spaces and algebras of operators over these Hilbert spaces. In this section, we introduce these spaces and present some of the key definitions and results. We also clarify how certain definitions are used throughout the thesis, including the notational conventions and assumptions adopted.

The material in this section is based on [Arv76; Jon15; Hia20].

2.1.1 Hilbert spaces and operator algebras

The fundamental structure for quantum mechanics is the Hilbert space, defined as follows [Jon15]:

Definition 2.1
A Hilbert space $\mathcal{H}$ is a $\mathbb{C}$-vector space with an inner product $\langle\cdot,\cdot\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ such that:

i) $\langle\cdot,\cdot\rangle$ *is linear in the first variable.*

ii) $\forall\, \phi, \psi \in \mathcal{H}$, $\langle\psi, \phi\rangle = \overline{\langle\phi, \psi\rangle}$.

iii) If $0 \neq x \in \mathcal{H}$ *then* $\langle x, x\rangle > 0$.

iv) $\mathcal{H}$ *is closed under the norm* $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$.

We will use mostly standard physics notation, where elements of a Hilbert space are denoted by kets $|\varphi\rangle$ and the inner product is represented as $\langle\varphi|\varphi\rangle$. Additionally, we are only concerned with finite dimensional Hilbert spaces, thus we present results simplified for this case. We also generally ignore the closure condition on a lot of these definitions and results, since it is trivial in the finite dimensional case.

We denote by $\mathcal{B}(\mathcal{H})$ the $*$-algebra[1] of all bounded linear operators on a Hilbert space $\mathcal{H}$.

Any subalgebra $\mathscr{A} \subseteq \mathcal{B}(\mathcal{H})$ that is closed under the adjoint operation, which we denote by *, is a concrete operator algebra. Given a subset $A \subseteq \mathcal{B}(H)$, we denote by $C^*(A)$ the smallest operator algebra that contains A, in the sense that any other $*$-algebra containing A must also contain $C^*(A)$.

Given an operator algebra on a Hilbert space $\mathcal{B}(\mathcal{H})$, we consider the following definitions:

Definition 2.2 *i) The identity operator, denoted by* $\mathbb{1} \in \mathcal{B}(\mathcal{H})$ *is the map such that* $\mathbb{1}\,|\varphi\rangle = |\varphi\rangle$ *for all* $|\varphi\rangle \in \mathcal{H}$.

ii) A subalgebra of $\mathcal{B}(\mathcal{H})$ *is unital if it contains* $\mathbb{1}$.

iii) And element $x \in \mathcal{B}(\mathcal{H})$ *is Hermitian or self-adjoint if* $x = x^*$*. We denote the set of self-adjoint elements as* $\operatorname{Herm}\mathcal{H}$.

iv) An element $\Pi \in \mathcal{B}(\mathcal{H})$ *is an orthogonal projection if* $\Pi^2 = \Pi = \Pi^*$.

[1] A vector space equipped with multiplication and the adjoint operation, denoted by *, which is closed under these operations.

v) *An element* $\omega \in \mathcal{B}(\mathcal{H})$ *is called positive semi-definite (psd) if* $\langle\varphi|\omega|\varphi\rangle \geq 0$ $\forall\,|\varphi\rangle \in \mathcal{H}$*. We denote the set of psd elements as* $\mathcal{B}_+(\mathcal{H})$*.*

vi) *An element* $U \in \mathcal{B}(\mathcal{H})$ *is a unitary if* $UU^* = U^*U = \mathbb{1}$*.*

vii) *The dual space* $\mathcal{H}^*$ *of a Hilbert space* $\mathcal{H}$ *is the set of all continuous linear maps* $f : \mathcal{H} \to \mathbb{C}$*. The Riesz representation theorem guarantees that for every* $f \in \mathcal{H}^*$ *there exists a unique vector* $|\varphi\rangle \in \mathcal{H}$ *such that* $f(|\psi\rangle) = \langle\varphi|\psi\rangle$ *for all* $|\psi\rangle \in \mathcal{H}$*. Hence, in the bra–ket notation, the functional* f *is denoted by* $\langle\varphi|$*.*

viii) *The commutator of two elements* $x, y \in \mathcal{B}(\mathcal{H})$ *is* $[x, y] = xy - yx$*.*

Finally, we write the finite dimensional, hermitian version of the well-known spectral theorem.

Theorem 2.3 (Spectral theorem)
Let $\mathcal{H}$ *be finite-dimensional Hilbert space and* $A \in \mathrm{Herm}\,\mathcal{H}$*. Then there exists an orthonormal basis of* $\mathcal{H}$ *consisting of eigenvectors of* A*, and the corresponding eigenvalues are real.*

Given two Hilbert spaces $\mathcal{H}_0$, $\mathcal{H}_1$ their tensor product is the linear completion of the map $(|\psi\rangle, |\varphi\rangle) \in \mathcal{H}_0 \times \mathcal{H}_1 \to |\psi\rangle \otimes |\varphi\rangle \in \mathcal{H}_0 \otimes \mathcal{H}_1$ with the inner product defined as

$$\langle\psi \otimes \varphi|\psi' \otimes \varphi'\rangle = \langle\psi|\psi'\rangle \langle\varphi|\varphi'\rangle \,. \tag{2.1}$$

Given two linear operators $A \in \mathcal{B}(\mathcal{H}_0)$, $B \in \mathcal{B}(\mathcal{H}_1)$ there exists a bounded operator $A \otimes B \in \mathcal{B}(\mathcal{H}_0 \otimes \mathcal{H}_1)$ defined as $(A \otimes B)(x \otimes y) = Ax \otimes By$.

2.1.2 von Neumann algebras

In this section we introduce von Neumann algebras and some properties that we use later.

Definition 2.4
A von Neumann algebra is a subalgebra of an operator algebra on a Hilbert space $\mathcal{B}(\mathcal{H})$ *that is unital and closed.*

We need von Neumann's bicommutant theorem, which we write for the case of finite dimensional operator algebras [Jon15; Hia20], also throrem 1.2.1 in [Arv76]. The commutant of a set $A \subseteq \mathcal{B}(\mathcal{H})$ is denoted by A' and defined as

$$A' = \{x \in \mathcal{B}(\mathcal{H}) \,|\, [x, a] = 0 \,\forall\, a \in A\} \,. \tag{2.2}$$

Similarly we write the commutant of the commutant, known as bicommutant, as $A'' = (A')'$. Von Neumann's bicommutant theorem makes it so that we never need to go further that the bicommutant. This specific formulation of the theorem is taken from [Jon15].

Theorem 2.5 (von Neumann's bicommutant theorem)
Let $\mathscr{M}$ be a unital, self-adjoint subalgebra of $\mathcal{B}(\mathcal{H})$ with $\dim \mathcal{H} = n \leq \infty$. Then $\mathscr{M} = \mathscr{M}''$.

To close the section, we include two technical lemmas and a known theorem for the tensor product of commutators.

Lemma 2.6
Let V and $\mathscr{A}$ be a vector space and algebra, respectively, both subsets in the same operator algebra. Let $\mathcal{B}$ be the basis of a V and let $\mathscr{A}$ be generated by $\operatorname{gen}\mathscr{A}$. Then, $V = \mathscr{A}$ if and only if

i) $\mathcal{B} \subseteq \mathscr{A}$.

ii) $\mathcal{B} \cdot \mathcal{B} \subseteq V$.

iii) $\operatorname{gen}\mathscr{A} \subseteq V$.

Proof. The 'only if' direction is trivial, so we focus on the if direction. Because V is embedded in a larger operator algebra, it makes sense to consider the product between its elements. Assume the conditions are true. Then, from the first one we see that $V \subseteq \mathscr{A}$, since $\mathscr{A}$ is an algebra and therefore closed under addition and product by scalar. From ii), we see that V has to be an algebra, since if $v, w \in V$, then

$$v = \sum_{b \in \mathcal{B}} v_b b, \quad w = \sum_{b \in \mathcal{B}} w_b b \tag{2.3}$$

and

$$vw = \sum_{b, b' \in \mathcal{B}} v_b w_{b'} b b' \in V, \tag{2.4}$$

if $bb' \in V$ for all $b, b' \in \mathcal{B}$.

Finally, the combination of V being an algebra and the generators of $\mathscr{A}$ being in V imply that $\mathscr{A} \subseteq V$, finishing the proof. □

Lemma 2.7
Let $\mathscr{A}_1, \mathscr{A}_2$ be finite dimensional von Neumann algebras. Then

$$(\mathscr{A}_1 \cap \mathscr{A}_2)' = (\mathscr{A}_1' \cup \mathscr{A}_2')''. \tag{2.5}$$

Proof. We start from the following equality [Arg20]:

$$(\mathscr{M}_1 \cup \mathscr{M}_2)' = \mathscr{M}_1' \cap \mathscr{M}_2', \tag{2.6}$$

where $\mathscr{M}_1$, $\mathscr{M}_2$ are von Neumann algebras. We first prove this equality by double inclusion. For $L \subseteq R$, consider T in the commutator of the union. That means that T commutes with the elements of $\mathscr{M}_1$ and $\mathscr{M}_2$, respectively. Thus, $T \in \mathscr{M}_1', \mathscr{M}_2'$. Therefore $T \in \mathscr{M}_1' \cap \mathscr{M}_2'$. Conversely for $L \supseteq R$, let T be in the intersection of commutants. Then T commutes with the elements of $\mathscr{M}_1$ and the elements of $\mathscr{M}$. Thus, T commutes with the elements of $\mathscr{M}_1 \cup \mathscr{M}_2$. Therefore, $T \in (\mathscr{M}_1 \cup \mathscr{M}_2)'$.

With this equality, we can write $\mathscr{M}_i' = \mathscr{A}_i$. Because $\mathscr{M}_i$ are von Neumann algebras, they fulfil $\mathscr{M}_i = \mathscr{M}_i''$, due to von Neumann's bicommutant theorem. To find the result, we take the previous equation and add a commutant to each side. Then we write $\mathscr{M}_i$ as $\mathscr{A}_i'$ and $\mathscr{M}_i'$ as $\mathscr{A}_i$:

$$\begin{aligned} (\mathscr{M}_1 \cup \mathscr{M}_2)' = \mathscr{M}_1' \cap \mathscr{M}_2' \quad &\Rightarrow \quad (\mathscr{M}_1 \cup \mathscr{M}_2)'' = (\mathscr{M}_1' \cap \mathscr{M}_2')' \\ &\Rightarrow \quad (\mathscr{A}_1' \cup \mathscr{A}_2')'' = (\mathscr{A}_1 \cap \mathscr{A}_2)'. \end{aligned} \tag{2.7}$$

Note that $(\mathscr{A}_1' \cup \mathscr{A}_2')'' = \mathscr{A}_1' \cup \mathscr{A}_2'$ is false in general because the union of algebras is not an algebra. ☐

Finally, we write the commutation theorem for tensor products of von Neumann algebras, taken from [Arv76].

Theorem 2.8 (Commutation theorem)
Let $\mathscr{A}_1, \mathscr{A}_2$ be von Neumann operator algebras. Then

$$(\mathscr{A}_1 \otimes \mathscr{A}_2)' = \mathscr{A}_1' \otimes \mathscr{A}_2'. \tag{2.8}$$

2.1.3 Convex cones

Definition 2.9
A subset S of a complex vector space V is convex if $\forall x, y \in S$ and $t \in [0,1]$

$$tx + (1-t)y \in S. \tag{2.9}$$

Definition 2.10
A subset C of a complex vector space V is a cone if $tC \subseteq C$ for $t > 0$[2].

Definition 2.11 *i) A cone C is pointed if $C \cap (-C) = \{0\}$.*

ii) Given a set S the cone generated by S is the minimal convex cone that contains S and it is denoted by $\mathrm{cone}\,(S)$.

iii) If V is equipped with an inner product and $S \subseteq V$, the dual cone to S is $\{v \in V \mid \langle v, s\rangle \geq 0 \,\forall s \in S\}$. We denote the dual as S^*. The dual cone is always a convex cone, even if S is not.

iv) Generally we consider $V = \mathcal{B}(\mathcal{H})$, equipped with the Hilbert-Schmidt inner product: $\langle x, y\rangle_{HS} = \mathrm{Tr}\,[x^* y]$.

The rest of the section contains three technical lemmas on cones.

Lemma 2.12
Let I be an index set and $\{\mathcal{C}_i\}_{i\in I}$ a set of pointed cones. Then

$$\bigcap_{i\in I} \mathcal{C}_i^* = \left(\sum_{i\in I} \mathcal{C}_i\right)^* . \tag{2.10}$$

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

$$\begin{aligned} x \in \mathcal{C}_i^* \;\forall i \in I \quad &\Leftrightarrow \quad \langle x|c_i\rangle \geq 0 \;\forall c_i \in \mathcal{C}_i \;\forall i \in I \quad \Leftrightarrow \quad \sum_{i\in I} \langle x|c_i\rangle \geq 0 \;\forall c_i \in \mathcal{C}_i \\ &\Leftrightarrow \quad \left\langle x \middle| \sum_{i\in I} c_i \right\rangle \geq 0 \;\forall c_i \in \mathcal{C}_i \quad \Leftrightarrow \quad x \in \left(\sum_{i\in I} \mathcal{C}_i\right)^* . \end{aligned} \tag{2.11}$$

□

Lemma 2.13
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 $x \in C$.

[2] Sometimes the definition includes $t = 0$, or asks that the cone is pointed. In this thesis we always work with cones that include 0 but some of them are not pointed.

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, there exists a linear map such that $f(x) > 0$ for all $x \in C\setminus\{0\}$. □

Lemma 2.14
Let K be a convex cone and A an invertible linear map. Then,

$$A(K)^* = (A^*)^{-1}(K^*). \tag{2.12}$$

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

$$\begin{aligned} x \in A(K)^* \quad &\Leftrightarrow \quad \langle x, A(y)\rangle \geq 0 \;\; \forall y \in K \quad \Leftrightarrow \quad \langle A^*(x), y\rangle \geq 0 \;\; \forall y \in K \\ &\Leftrightarrow \quad A^*(x) \in K^* \quad \Leftrightarrow \quad x \in (A^*)^{-1}(K^*). \end{aligned} \tag{2.13}$$

□

2.2 Quantum mechanics

The setting of this thesis is quantum mechanics. In this section we introduce quantum mechanics as a mathematical theory from the ground up: we start with the postulates; then we introduce the quantum state; we show how quantum states can be manipulated, and finally we introduce a key result that connects quantum maps to quantum states: the Choi and Jamiołkowski isomorphisms.

Most of the general information in this chapter has been taken from [NC10], although any other general quantum mechanics book with any basics of quantum information would serve.

2.2.1 Postulates

As far as this work is concerned, quantum mechanics is a mathematical theory that can be used to describe and, probabilistically, predict reality. We are not particularly concerned with the second part of that statement, and will look at quantum theory as mostly a mathematical theory. In this section we introduce the postulates of quantum mechanics. These postulates serve both as the connection between physical reality and the mathematical theory, and as the foundation of the mathematical theory itself.

Postulate 1. An isolated physical system can be described with a complex Hilbert space $\mathcal{H}$, known as the state space of the system. The state of the system will then be represented by an element of this Hilbert space with norm one:

$$|\varphi\rangle \in \mathcal{H} \quad \text{such that} \quad \langle\varphi|\varphi\rangle = 1. \tag{2.14}$$

This first postulate is the main link between physical reality and mathematical theory, it describes how reality can be mathematically described for the purpose of quantum theory.

Postulate 2. A closed quantum system evolves under a unitary operator, with its Hamiltonian H acting as the generator of this evolution. The corresponding time-evolution operator is $U_t = e^{-\frac{\mathrm{i}tH}{\hbar}}$. Hence, the state at time t is

$$|\varphi(t)\rangle = U_t\,|\varphi(0)\rangle = e^{-\frac{\mathrm{i}tH}{\hbar}}\,|\varphi(0)\rangle\,. \tag{2.15}$$

This postulate describes the evolution of isolated quantum systems, with isolated being a key qualifier. Although most systems are open, a system can be viewed as a subsystem (see Postulate 4) of a larger isolated system. This perspective leads to the Stinespring dilation theorem [Sti55], introduced later in Theorem 2.26, which shows that any non-unitary evolution of a system can be realized as a unitary evolution of the system plus a suitable environment.

Postulate 3. A quantum measurement is given by a set of positive operators $\{A_i\}_{i\in\mathcal{I}}$, where $\mathcal{I}$ is some index set, such that

$$\sum_i A_i^* A_i = \mathbb{1}. \tag{2.16}$$

The probability of obtaining each measurement result $i \in \mathcal{I}$ when measuring on a state $|\varphi\rangle$ is

$$P^{\mathbf{M}}_{|\varphi\rangle}(i) = \langle\varphi|A_i^* A_i|\varphi\rangle\,, \tag{2.17}$$

and the post-measurement state for measurement result i is

$$\frac{A_i\,|\varphi\rangle}{\sqrt{\langle\varphi|A_i^* A_i|\varphi\rangle}}. \tag{2.18}$$

This postulate expresses how we can interact with our quantum system to extract information from it. It famously includes two of the key features of quantum mechanics: the results are probabilistic in general and measuring the state of the system changes the state of the system.

Postulate 4. The state space of a composite physical system is described as the tensor product of the component systems. If we have a composite system with n subsystems and each subsystem is denoted by $\mathcal{H}_i$, the state space of the composite system $\mathcal{H}$ is

$$\mathcal{H} = \bigotimes_i \mathcal{H}_i. \tag{2.19}$$

This postulate describes how to describe systems made made of smaller subsystems. Note that, by Postulate 1, any properly normalized linear combination of states is a valid state, reflecting the superposition principle. The interplay between tensor-product composition and linearity gives rise to the richness of quantum correlations and entanglement. Such correlations can, for instance, arise from initially uncorrelated states through interactions between the subsystems.

Postulates 3 and 4 are both very relevant to this thesis: composite systems, particularly bipartite ($n = 2$) systems, play a central role in the Choi and Jamiołkowski isomorphisms, while measurements are essential for connecting quantum and classical relative entropies.

2.2.2 Classical uncertainty and the quantum state

We have seen on Postulate 3 that quantum states have an inherent uncertainty. Even if we perfectly know the state of the system its outcome under measurement is not deterministic. On top of that we can add some classical uncertainty, that is some lack of knowledge of the system on the observer's part. We can consider a setting where the observer knows that a state $|\varphi_i\rangle$ is prepared with probability p_i, but does not know which particular instance has been realized. This lack of knowledge is described mathematically by a convex combination, as shown in the following definition:

Definition 2.15
Let $\mathcal{H}$ be a Hilbert space, and let $\{p_i, |\varphi_i\rangle\}_{i\in\mathcal{I}}$, be a collection of states $\mathcal{H}$, each associated to a probability. We call this an ensemble, and the associated ensemble state is

$$\rho = \sum_i p_i \, |\varphi_i\rangle\!\langle\varphi_i| \in \mathcal{B}(\mathcal{H}). \tag{2.20}$$

We can note two properties of this ensemble state. Firstly, the normalisation of the states $\{|\varphi_i\rangle\}_{i\in\mathcal{I}}$ and the probability distribution imply that $\mathrm{Tr}\,[\rho] = 1$. Secondly, the structure of $|\varphi_i\rangle\!\langle\varphi_i|$ and the positivity of each p_i imply that ρ is a positive operator. We can turn this properties around into a definition.

Definition 2.16
Let $\mathcal{H}$ be a Hilbert space. We call $\mathcal{S}(\mathcal{H}) \subseteq \mathcal{B}(\mathcal{H})$ the set of quantum states. *A state $\rho \in \mathcal{S}(\mathcal{H})$ if and only if*

$$\rho \geq 0, \quad \mathrm{Tr}\,[\rho] = 1. \tag{2.21}$$

These states are also sometimes called density matrices. *We will use the terms state, quantum state, density matrix and density operator interchangeably.*

Definition 2.16 is the definition of state we work with for the entirety of he thesis, leaving the more interpretable Definition 2.15 behind. The ensemble state gives us a view on how the quantum state relates to classical probability theory. In classical probability theory we have deterministic states of the system and a lack of complete knowledge of which state the system is in, which adds uncertainty. In quantum theory we have non-deterministic states and, on top of that, a lack of complete knowledge of the state of the system, leading to even more uncertainty.

The states $|\varphi\rangle \in \mathcal{H}$ are represented as density matrices as $|\varphi\rangle\langle\varphi|$. If a state can be written as $|\varphi\rangle\langle\varphi|$, that it if it has a unique ensemble with a single non-zero probability, it is a pure state. If a state is not pure it is a mixed state. If a state is pure, we sometimes write the Hilbert space element $|\varphi\rangle$ instead of the density matrix $|\varphi\rangle\langle\varphi|$.

We have shown two definitions and claimed they are equivalent. We have also shown that Definition 2.15 implies Definition 2.16. We want to show the converse. Let $\rho \in \mathcal{S}(\mathcal{H})$. ρ is in particular Hermitian, so by the spectral theorem Theorem 2.3 there exist orthonormal pure states $|\varphi_i\rangle\langle\varphi_i|$ and positive eigenvalues a_i such that

$$\rho = \sum_i a_i |\varphi_i\rangle\langle\varphi_i| . \tag{2.22}$$

Because ρ is positive, its eigenvalues a_i are also positive. Moreover, because ρ and each $|\varphi_i\rangle\langle\varphi_i|$ have trace 1, the sum of eigenvalues has to be 1 as well. With this we obtain an ensemble associated to ρ: $\{a_i, |\varphi_i\rangle\}$, obtaining that both definitions are equivalent.

Note that we mentioned we obtained *an* ensemble, not *the* ensemble. That is because different ensembles can lead to the same state. Moreover the pure states in the ensemble need not be orthogonal, and there can be more than the dimension of $\mathcal{H}$. We show this in the following example:

Example 2.17
Let $\mathcal{H} = \mathbb{C}^2$. Consider $\rho = \frac{1}{4}|0\rangle\langle 0| + \frac{3}{4}|1\rangle\langle 1|$. From the definition of ρ we see that $\{(\frac{1}{4}, |0\rangle), (\frac{3}{4}, |1\rangle)\}$ is an ensemble for ρ. We can now write ρ in matrix form in the basis $\{|0\rangle, |1\rangle\}$ and split it into three pure states:

$$\begin{aligned}\rho =& \frac{1}{4}\begin{pmatrix}1 & 0\\ 0 & 3\end{pmatrix} = \frac{3}{8}\left[\frac{1}{3}\begin{pmatrix}1 & \sqrt{2}e^{\mathrm{i}\phi}\\ \sqrt{2}e^{-\mathrm{i}\phi} & 2\end{pmatrix}\right] + \frac{3}{8}\left[\frac{1}{3}\begin{pmatrix}1 & -\sqrt{2}e^{\mathrm{i}\phi}\\ -\sqrt{2}e^{-\mathrm{i}\phi} & 2\end{pmatrix}\right]\\ &+ \frac{1}{4}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} = \frac{3}{8}|\tilde{+}\rangle\langle\tilde{+}| + \frac{3}{8}|\tilde{-}\rangle\langle\tilde{-}| + \frac{1}{4}|1\rangle\langle 1| ,\end{aligned} \tag{2.23}$$

with $|\tilde{\pm}\rangle = \frac{1}{\sqrt{3}}\left(|0\rangle \pm \sqrt{2}e^{\mathrm{i}\phi}|1\rangle\right)$. This yields a different ensemble $\{(\frac{3}{8}, |\tilde{\pm}\rangle), (\frac{1}{4}, |1\rangle)\}$ for ρ. Moreover, we see that there are more states in the ensemble than the dimension and the states in the ensemble are not mutually orthogonal.

It is important to emphasize that the density operator is the unique and most complete characterization of the state of knowledge about a quantum system. It captures all statistical information available when the system is in a probabilistic mixture of pure states as well as when it is a subsystem of a larger entangled system. In the latter case, the density operator of the subsystem is obtained by taking the partial trace over the degrees of freedom of the environment/remaining subsystems, as defined in Eq. (2.39) below.

The postulates in the state formalism

In Section 2.2.1 we stated the postulates of quantum mechanics for observers with complete knowledge of the state. In the following we show the same postulates reformulated for the density matrix formalism. The comments on the postulates from Section 2.2.1 still apply, the only difference is how the concepts are expressed mathematically.

Postulate 1. An isolated physical system can be described with a complex Hilbert space $\mathcal{H}$, known as the state space of the system. The state of the system will then be represented by its associated density matrix, an operator $\rho \in \mathcal{B}(\mathcal{H})$ such that

$$\rho \geq 0, \quad \mathrm{Tr}\left[\rho\right] = 1. \tag{2.24}$$

Postulate 2. A closed quantum system will evolve under a unitary operator. If the system is under a Hamiltonian H, this unitary operator is $U_t = e^{-\frac{itH}{\hbar}}$. Mathematically:

$$\rho(t) = U_t \rho(0) U_t^* = e^{-\frac{itH}{\hbar}} \rho(0) e^{\frac{itH}{\hbar}}. \tag{2.25}$$

Postulate 3. A quantum measurement is given by a set of operators $\{A_i\}_{i\in\mathcal{I}}$, where $\mathcal{I}$ is some index set, such that

$$\sum_i A_i^* A_i = \mathbb{1}. \tag{2.26}$$

The probability of obtaining each measurement result $i \in \mathcal{I}$ of the measurement when measuring on a state ρ is given by the Born rule

$$P_\rho^{\mathbf{A}}(i) = \mathrm{Tr}\left[A_i^* A_i \rho\right], \tag{2.27}$$

and the post-measurement state for measurement result i is

$$\frac{A_i \rho A_i^*}{\mathrm{Tr}\left[A_i^* A_i \rho\right]}. \tag{2.28}$$

Typically we are not concerned with the post measurement state. In this case we can define a POVM:

Definition 2.18
A POVM is given by a set of operators positive $\mathrm{M} = \{M_i\}_{i\in\mathcal{I}} \subset \mathcal{B}_+(\mathcal{H})$*, where* $\mathcal{I}$ *is some index set, such that*

$$\sum_i M_i = \mathbb{1}. \tag{2.29}$$

The probability of obtaining each measurement result $i \in \mathcal{I}$ *of the measurement when measuring on a state* ρ *is given by the Born rule*

$$P_\rho^{\mathbf{M}}(i) = \mathrm{Tr}\left[M_i \rho\right]. \tag{2.30}$$

Postulate 4. The state space of a composite physical system is described as the tensor product of the component systems. If we have a composite system with n subsystems and each subsystem Hilbert space is denoted by $\mathcal{H}_i$, then the operator space to the composite Hilbert space $\mathcal{H}$ is

$$\mathcal{B}\left(\mathcal{H}\right) = \mathcal{B}\left(\bigotimes_i \mathcal{H}_i\right). \tag{2.31}$$

Note that for a composite system with n independently prepared subsystems in states ρ_i, the state of the composite system is

$$\rho = \bigotimes_{i=1}^{n} \rho_i. \tag{2.32}$$

As in the pure state case, the state of a composite system need not be a simple product. In the most general case, correlations between subsystems can be present, including quantum entanglement. A composite state ρ acting on the tensor product Hilbert space of all subsystems is called *separable* if it can be written as a convex combination of product states,

$$\rho = \sum_k p_k\, \rho_1^{(k)} \otimes \rho_2^{(k)} \otimes \cdots \otimes \rho_n^{(k)}, \quad p_k \geq 0,\ \sum_k p_k = 1, \tag{2.33}$$

and *entangled* if it cannot be written in this form. Separable states describe correlations that can be understood as classical mixtures of product states, whereas entangled states exhibit intrinsically quantum correlations that have no classical analogue [Wer89].

2.2.3 The Bloch sphere

The Bloch sphere is a three dimensional real representation of the 2 dimensional Hilbert space $\mathbb{C}^2$. In this representation, pure states are elements of the unit sphere, mixed states are the interior of the unit ball and unitary operators turn into rotations. Generally, we work with the canonical basis $\{|0\rangle, |1\rangle\}$, which we associate to the z axis, such that

$$\vec{n}_{|0\rangle} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{n}_{|1\rangle} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. \tag{2.34}$$

Note that the orthogonal quantum states $|0\rangle$ and $|1\rangle$ map to non-orthogonal $\mathbb{R}^3$ vectors. Instead, orthogonality in $\mathbb{C}^2$ corresponds to antipodal states on the Bloch representation. Similarly, the states $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ and $|\mathrm{i}\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm \mathrm{i}\,|1\rangle)$ are mapped to the x and y axis, respectively:

$$\vec{n}_{|\pm\rangle} = \begin{pmatrix} \pm 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{n}_{|\mathrm{i}\pm\rangle} = \begin{pmatrix} 0 \\ \pm 1 \\ 0 \end{pmatrix}. \tag{2.35}$$

A general mixed state maps to a $\mathbb{R}^3$ vector $\vec{n}_\rho$, such that $\|\vec{n}_\rho\| \leq 1$, as

$$\rho = \frac{1}{2}\mathbb{1} + \vec{n}_\rho \cdot \vec{\sigma}, \quad \vec{\sigma} = \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right). \tag{2.36}$$

From this equation, it is clear that $\vec{0} \in \mathbb{R}^3$ does not map to some zero density matrix in the quantum space, but instead to the maximally mixed state: $\mathbb{1}/2$. The pure states are sometimes also identified with standard spherical coordinates, that is (θ, ϕ); where θ is the angle between the projection of the state on the $x - y$ plane and the positive x axis, and ϕ is the angle between the state and the positive z axis.

Finally, if a rotation with angle ϕ around and axis $\vec{n}$ will map to the following unitary:

$$U_{\vec{n}}(\phi) = e^{-\mathrm{i}\frac{\phi}{2}\vec{n}\cdot\vec{\sigma}}. \tag{2.37}$$

For the purpose of this thesis, the Bloch sphere representation is mostly used as a visual aid in some examples and the representation of unitary maps is not relevant, but has been added for completeness.

2.2.4 Composite systems

As stated in Section 2.2.1, composite systems are an important part of this thesis due to their appearance in the Choi-and Jamiołkowski isomorphisms. In this sections we outline some properties and operations of composite systems.

From Postulate 4, the state space of a composite system is the tensor product of the state spaces of its components. Given n Hilbert spaces $\mathcal{H}_i$, $i \in [n]$, the space state of the composite system $\mathcal{H} = \mathcal{H}_0 \otimes \cdots \otimes \mathcal{H}_{n-1}$ is

$$\mathcal{S}(\mathcal{H}) = \{x \in \mathcal{B}_+(\mathcal{H}_0 \otimes \cdots \otimes \mathcal{H}_{n-1}) \mid \operatorname{Tr}[x] = 1\}, \tag{2.38}$$

For states in a composite system we use the notation ρ_i, where the subscript i indicates the state space in which ρ_i belongs. This notation allows us to easily denote the state on a subsystem. Let $\mathcal{I} \subseteq [n]$. We usually denote the state of the space $\bigotimes_{i\in\mathcal{I}} \mathcal{H}_i$ ρ. Often in this work we work with bipartite systems. That is two subsystems, $\mathcal{H}_A$ and $\mathcal{H}_B$ where we denote the states on each system by ρ_{AB}, $\rho_A = \operatorname{Tr}_B \rho_{AB}$ and $\rho_B = \operatorname{Tr}_A \rho_{AB}$.

We can obtain the state of a subset of subsystems, sometimes called the *reduced state*, with the partial trace:

Definition 2.19
Let $\mathcal{H}_A$, H_B be Hilbert spaces and $|a_{1,2}\rangle \in \mathcal{H}_A$, $|b_{1,2}\rangle \in \mathcal{H}_B$. The partial trace on subsystem B is the linear map $\operatorname{Tr}_B : B(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ such that

$$\operatorname{Tr}_B [|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|] = \langle b_2|b_1\rangle \, |a_1\rangle\langle a_2| \,. \tag{2.39}$$

The partial trace fulfils the following properties:

Proposition 2.20
Let $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$ be Hilbert spaces.

i) Let $M_A \in \mathcal{B}(\mathcal{H}_A)$ an operator and consider $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$. The following identity is true:

$$\operatorname{Tr}[\rho_{AB}(M_A \otimes \mathbb{1}_B)] = \operatorname{Tr}[\operatorname{Tr}_B[\rho_{AB}] M_A] = \operatorname{Tr}[\rho_A M_A] \,. \tag{2.40}$$

ii) Consider an operator $M_{ABC} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$. Then

$$\operatorname{Tr}_{AB}[M_{ABC}] = \operatorname{Tr}_A[\operatorname{Tr}_B[M_{ABC}]] = \operatorname{Tr}_B[\operatorname{Tr}_A[M_{ABC}]] \tag{2.41}$$

Physically, the partial trace corresponds to forgetting or ignoring the information contained in a system. The partial trace of a joint state corresponds to the accessible state by the party that has possession of the subsystem that is left.

Similarly to the partial trace, we can define the partial transpose.

Definition 2.21
Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces with basis $\mathcal{B}_A$, $\mathcal{B}_B$. The partial transpose on subsystem B with respect to basis $\mathcal{B}_B$ is the linear map ${}^{T_{\mathcal{B}_B}} : B(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that

$$(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|)^{T_B} = |a_1\rangle\langle a_2| \otimes |b_2\rangle\langle b_1| \quad \forall\, |b_{1,2}\rangle \in \mathcal{B}_B, \quad |a_{1,2}\rangle \in \mathcal{H}_A. \tag{2.42}$$

Often the partial transpose is taken with respect to the canonical basis and just denoted ^{T_B}.

Two important objects in bipartite systems for this work are the (canonical) maximally entangled state and the swap operator. Consider two copies of a d-dimensional Hilbert space $H = \mathcal{H}_A = \mathcal{H}_B$: $\mathcal{H}^2 = \mathcal{H}_A \otimes \mathcal{H}_B$. The unnormalized maximally entangled state $|\Phi_+\rangle$ and swap operator $\mathcal{S}$ on $\mathcal{H}^2$ are defined as follows:

$$|\Phi_+\rangle = \sum_{i\in[d]} |i\rangle \otimes |i\rangle = \sum_i |ii\rangle\,, \quad \mathcal{S} = \sum_{i,j\in[d]} |i\rangle\langle j| \otimes |j\rangle\langle i| = \sum_{ij} |ij\rangle\langle ji|\,. \tag{2.43}$$

Notably, they are related by the partial transpose:

$$|\Phi_+\rangle\langle\Phi_+|^{T_B} = \sum_{ij} |ii\rangle\langle jj|^{T_B} = \sum_{ij} |ij\rangle\langle ji| = \mathcal{S}. \tag{2.44}$$

Because the partial transpose is clearly self-inverse $|\Phi_+\rangle\langle\Phi_+| = \mathcal{S}^{T_B}$ is also true.

We have not talked about positivity of maps yet, but we mention the properties of the partial trace in this regard. For the definition of positivity see the next section. Note that the partial trace is a completely positive trace preserving map but the partial transpose is not. In fact, it is the generic example of map that is positive, but not 2-positive.

To end this section, we have the follow technical lemma on the partial transpose:

Lemma 2.22
The partial transpose map, denoted here by $T_A(\cdot)$, fulfils the following:

$$\begin{aligned} \mathrm{Tr}_A\left[T_A(K)C\right] &= \mathrm{Tr}_A\left[K\; T_A(C)\right] \\ T_A(\mathrm{Tr}_B\left[K\; C\right]) &= \mathrm{Tr}_B\left[T_A(C)\; T_A(K)\right] \quad \forall K, C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B),\ \rho \in \mathcal{B}(\mathcal{H}_A). \\ T_A((\rho \otimes \mathbb{1})C) &= T_A(C)(\rho^T \otimes \mathbb{1}) \end{aligned} \tag{2.45}$$

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 three 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 [Bia20] as shown in Fig. 2.1.

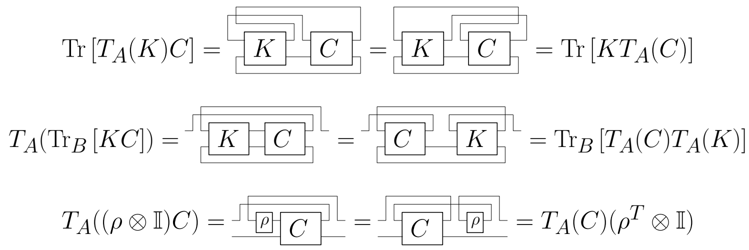

Figure 2.1: Proofs of the expressions in Lemma 2.22 using tensor network notation.

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:

$$\begin{aligned} \mathrm{Tr}_B\,\mathrm{Tr}_A\left[T_A(K^*)C\right] &= \mathrm{Tr}_B\,\mathrm{Tr}_A\left[K^*\,T_A(C)\right] \\ \Leftrightarrow \mathrm{Tr}\left[T_A(K^*)C\right] &= \mathrm{Tr}\left[K^*\,T_A(C)\right] \\ \Leftrightarrow \langle T_A(K^*), C\rangle_{HS} &= \langle K^*, T_A(C)\rangle_{HS}. \end{aligned} \tag{2.46}$$

□

2.3 Operations on quantum states

The most general operation that can be performed on a quantum state is described by a quantum channel. A quantum channel is a map that sends quantum states to quantum states—that is, it maps positive semidefinite, unit-trace operators to positive semidefinite, unit-trace operators. To ensure physical consistency, the map must be linear, so that ensembles of input states are mapped to the corresponding mixtures of output states, preserving causality. Moreover, it must be completely positive, meaning that the map preserves positivity even when the input is part of a larger, possibly entangled system that is unaffected by the channel. Finally, the map must be trace-preserving to guarantee that probabilities remain normalized. Quantum channels are therefore linear, Hermiticity-preserving, completely positive, and trace-preserving maps, commonly abbreviated as CPTP maps. In this sense, CPTP maps provide the most general framework for describing physically valid transformations of quantum states, encompassing unitary evolution, measurements, decoherence, and interactions with ancillary systems. We precisely define define positive, then n-positive and finally completely positive before continuing.

Definition 2.23
Let $\mathcal{E}$ be a linear operator $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$. $\mathcal{E}$ is positive if

$$\mathcal{E}(\mathcal{B}_+(\mathcal{H}_A)) \subseteq \mathcal{B}_+(\mathcal{H}_B) \tag{2.47}$$

Definition 2.24
Let $\mathcal{E}$ be a linear operator $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$. $\mathcal{E}$ is n-positive if

$$\mathrm{id} \otimes \mathcal{E} : \mathbb{C}^{n \times n} \otimes \mathcal{H}_A \to \mathbb{C}^{n \times n} \otimes \mathcal{H}_B \tag{2.48}$$

is positive.

Definition 2.25
Let $\mathcal{E}$ be a linear operator $\mathcal{E} : \mathcal{H}_A \to \mathcal{H}_B$. $\mathcal{E}$ is completely positive (CP) if it is n-positive for all n.

In this section we introduce some useful representations of CPTP maps as well as some useful properties. For this entire section, unless stated otherwise, we consider up to three Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$, and CPTP maps

$$\mathcal{E}_{AB} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \quad \mathcal{E}_{BC} : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_C). \tag{2.49}$$

2.3.1 The Stinespring dilation theorem

In Section 2.2.1 we saw that an isolated quantum system evolves under a unitary operator. A general evolution under a CPTP map is the evolution of a non-isolated quantum system. We can consider the system with which our system is interacting, thus obtaining an isolated system. This system is then evolving unitarily. This is the idea of the Stinespring dilation theorem, stated below.

Theorem 2.26 (Stinespring dilation [Sti55])
Let $\mathcal{H}_A, \mathcal{H}_B$ be Hilbert spaces and consider $\rho \in \mathcal{S}(\mathcal{H}_A)$ and a CPTP map $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$. Then there exist a Hilbert space $\mathcal{H}_R$ and a bounded map $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_R$ that fulfils $V^*V = \mathrm{id}_A$ such that

$$\mathcal{E}(\rho) = \mathrm{Tr}_R \left[V \rho V^* \right] . \tag{2.50}$$

In the case where $\mathcal{H}_A = \mathcal{H}_B$ this can be restated as:

Corollary 2.27
Let $\mathcal{H}$ be a Hilbert space and consider $\rho \in \mathcal{S}(\mathcal{H})$ and a CPTP map $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$. Then there exist a Hilbert space $\mathcal{H}_R$ and a unitary

$$\mathcal{E}(\rho) = \mathrm{Tr}_R \left[U \left(\rho \otimes |0\rangle\langle 0|_R \right) U^* \right], \tag{2.51}$$

with U a unitary over $\mathcal{H} \otimes \mathcal{H}_R$.

This corollary shows the connection between CPTP maps and unitaries in the larger space.

2.3.2 The Choi and Jamiołkowski isomorphisms

The Choi-Jamiołkowski is a well known result in quantum mechanics. Throughout this thesis we have referred to this result as the Choi and Jamiołkowski isomorphisms because these are, in fact, two different but equivalent results. In this section we introduce these isomorphisms, their uses and their differences.

The Choi and Jamiołkowski isomorphisms [Jam72; Cho75] 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 really important to this thesis because they allow us to treat maps as states, 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 Chapter 5 and Chapter 6.

Theorem 2.28 (Choi isomorphism [Cho75])
Let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be a linear map and $|\Phi_+\rangle \in \mathcal{S}(\mathcal{H}_A)$ the maximally entangled state in (2.43). Then associated operator $C_\mathcal{E} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, defined as

$$C_\mathcal{E} = \left(\mathrm{id}_A \otimes \mathcal{E} \right) \left(|\Phi_+\rangle\langle \Phi_+| \right), \tag{2.52}$$

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

$$\mathcal{E}(x) = \mathrm{Tr}_A \left[(x^T \otimes \mathbb{1}_B) C_\mathcal{E} \right]. \tag{2.53}$$

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 $\mathrm{Tr}_B \left[C_\mathcal{E} \right] = \mathbb{1}_A$.

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

Theorem 2.29 (Jamiołkowski isomorphism [Jam72])
Let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be a linear map and $\mathcal{S} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ the swap operator on the joint system AB. The associated operator $J_\mathcal{E} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, defined as

$$J_\mathcal{E} = (\mathrm{id}_A \otimes \mathcal{E})(\mathcal{S}), \tag{2.54}$$

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

$$\mathcal{E}(x) = \mathrm{Tr}_A \left[(x \otimes \mathbb{1}_B) J_\mathcal{E}\right]. \tag{2.55}$$

Moreover $\mathcal{E}$ is completely positive if and only if $J_\mathcal{E}^{T_A}$ is positive[3] *and $\mathcal{E}$ is trace preserving if and only if $\mathrm{Tr}_B [J_\mathcal{E}] = \mathbb{1}_A$.*

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

$$J = C^{T_A}. \tag{2.56}$$

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 you need to check positivity 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 thesis. 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.

2.3.3 The operator-sum representation

The operator-sum representation is also common way of representing CPTP maps in physics. It has the advantage of giving rise to clearly CP maps as well as being easy to manipulate and combine for more practical purposes.

[3] For simplicity we usually use positive instead of the more rigorous positive semi-definite.

Theorem 2.30 (Operator-sum or Kraus representation [Kra83])
Let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ *be a completely positive map. Then there exist bounded operators* $\{K_k : \mathcal{H}_A \to \mathcal{H}_B\}$ *such that*

$$\mathcal{E}(x) = \sum_k K_k x K_k^*. \tag{2.57}$$

Additionally, $\mathcal{E}$ *is trace preserving if and only if*

$$\sum_k K_k^* K_k = \mathbb{1}. \tag{2.58}$$

The Kraus representation of a CPTP map is in general not unique. We can relate the Choi representation of a channel with its Kraus representation as follows. Given a Choi matrix $C \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we want to write this matrix as a sum of pure elements: $C = \sum_k |\varphi\rangle\langle\varphi|$. Because the Choi matrix is positive we can use its diagonal decomposition to achieve one particular instance of this. There will then exist a collection of orthonormal pure states $\{|\tilde{\varphi}_k\rangle\}$ and associated positive eigenvalues $\{a_k\}$. We can absorb the eigenvalues into the states, defining a new collection of mutually orthogonal but unnormalised elements $\{|\varphi_k\rangle = \sqrt{a_k}\,|\tilde{\varphi}_k\rangle\}$. With these new elements the Choi matrix is

$$C = \sum_k a_k \,|\tilde{\varphi}_k\rangle\langle\tilde{\varphi}_k| = \sum_k |\varphi_k\rangle\langle\varphi_k|\,. \tag{2.59}$$

Now, we choose a product basis of $\mathcal{H}_A \otimes \mathcal{H}_B$ and in this basis write each $|\varphi_k\rangle = \sum_{kij} b_{kij}\,|\alpha_i\beta_j\rangle$. Finally, we can flip the second vector in the product[4] to obtain $K_k = \sum_{ij} b_{kij}\,|\beta_j\rangle\langle\alpha_i|$. The Kraus representation obtained in this manner is orthogonal, that is $\mathrm{Tr}\,[K_k K_{k'}^*] = \delta_{kk'}$, but only because we used the spectral decomposition which uses orthogonal pure elements. Given an arbitrary Kraus representation, we can recover the Choi matrix using this method, or by the definition in Eq. (2.52), but only an orthogonal Kraus representation yields an eigendecomposition of C.

2.4 Quantum information theory

In this section we will explain the basics of quantum information theory and some specific results relevant to this thesis. A lot of the concepts and results have a

[4] In practice using the Riesz representation theorem for the isomorphism $|\varphi\rangle \in \mathcal{H} \mapsto \langle\varphi| = \langle\varphi|\cdot\rangle \in \mathcal{H}^*$.

classical equivalent, sometimes with interesting different properties, and we make an effort of mentioning this comparison.

Compared to classical information theory, we now work with quantum states (instead of probability distributions) and CPTP maps (instead of stochastic maps).

Quantum entropy

Definition 2.31
Let $\mathcal{H}$ be a Hilbert space and $\rho \in \mathcal{S}(\mathcal{H})$ a state. The von Neumann (or quantum) entropy associated to ρ is

$$S(\rho) = -\operatorname{Tr}\left[\rho \log \rho\right]. \tag{2.60}$$

The von Neumann entropy has an immediate characterisation as a classical entropy. Consider an eigendecomposition of ρ: $\{(p_i, |\varphi_i\rangle)\}$, then the von Neumann entropy of ρ is equal to the Shannon entropy of the eigenvalues of ρ:

$$S(\rho) = H(p). \tag{2.61}$$

For completion, we include some of the fundamental properties of the Shannon entropy:

i) If $\dim \mathcal{H} = d$ and $\rho \in \mathcal{S}(\mathcal{H})$, then $\log n \geq S(\rho) \geq 0$.

ii) If $\rho = \bigotimes_i \rho_i$, then
$$S(\rho) = \sum_i S(\rho_i). \tag{2.62}$$

iii) If $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$.

In similar fashion to the Shannon entropy, we can generalise other kinds of entropies, where the characterisation as a classical entropy does not hold in general. Of great relevance to this work is the relative entropy [HP91; OH01].

Definition 2.32
Let $\mathcal{H}$ be a Hilbert space and $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be states on $\mathcal{H}$. The quantum relative entropy *between ρ and σ is*

$$D(\rho\|\sigma) = \begin{cases} \operatorname{Tr}[\rho(\log \rho - \log \sigma)], & \textit{if } \operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma), \\ +\infty, & \textit{otherwise.} \end{cases} \tag{2.63}$$

Note that the relative entropy is always positive. Similarly to the von Neumann entropy, the relative entropy is additive. Let $\rho_n = \rho_0 \otimes \cdots \otimes \rho_{n-1}$ and $\sigma_n = \sigma_0 \otimes \cdots \otimes \sigma_{n-1}$. Then

$$D(\rho_n \| \sigma_n) = \sum_i D(\rho_i \| \sigma_i). \tag{2.64}$$

The relative entropy is also jointly convex. Let p be a probability distribution and

$$\rho = \sum_i p_i \rho_i, \quad \sigma = \sum_i p_i \sigma_i. \tag{2.65}$$

Then $D(\rho\|\sigma) \leq \sum_i p_i D(\rho_i \| \sigma_i)$

Processing data

The relative entropy has the well known property of being non-increasing under quantum channels [Lin75; Uhl77].

Theorem 2.33 (Data processing inequality)
Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces and $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A)$ states. Let $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be a CPTP map. Then

$$D(\rho\|\sigma) \geq D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)). \tag{2.66}$$

The equality conditions for the DPI were characterized in [Pet86; Pet88], where the Petz recovery map was introduced:

$$\mathcal{R}_{\sigma,\mathcal{M}}(X) = \sigma^{1/2}\, \mathcal{M}^\dagger\left(\mathcal{M}(\sigma)^{-1/2}\, X\, \mathcal{M}(\sigma)^{-1/2}\right)\, \sigma^{1/2}. \tag{2.67}$$

It was shown in these works that equality in Theorem 2.33 holds if and only if $\rho = \mathcal{R}_{\sigma,\mathcal{M}}(\mathcal{M}(\rho))$. This result initiated the systematic study of recoverability. We expand on these follow-up results in Chapter 7, where we give a summary of the advances in this field in the last 15 years as well as our contributions.

The measured entropies

In a practical experimental setting it is hard to compute the entropy of a given state given the difficulty with knowing its eigenvalues. It is only possible to directly access information about the state through a POVM. This naturally leads to the definition of the measured entropy:

Definition 2.34
Let $\mathcal{H}$ be a Hilbert space, $\mathbf{M} \subset \mathcal{B}(\mathcal{H})$ a POVM and $\rho \in \mathcal{S}(\mathcal{H})$ a state. Let $P_\rho^{\mathbf{M}}(i) = \mathrm{Tr}\left[M_i \rho\right]$ The measured entropy *of ρ over $\mathbf{M}$ is*

$$S_M(\rho) = H(P_\rho^{\mathbf{M}}). \tag{2.68}$$

Similarly we can define the relative version of the measured entropy:

Definition 2.35
Let $\mathcal{H}$ be a Hilbert space, $\mathbf{M} \subset \mathcal{B}(\mathcal{H})$ a POVM and $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ states. Let $P_\rho^{\mathbf{M}}(i) = \operatorname{Tr}[M_i \rho]$, and equivalently $P_\sigma^{\mathbf{M}}(i) = \operatorname{Tr}[M_i \sigma]$. The measured relative entropy *of ρ, σ over $\mathbf{M}$ is*

$$D_M(\rho\|\sigma) = D(P_\rho^{\mathbf{M}} \| P_\sigma^{\mathbf{M}}). \tag{2.69}$$

Note that some authors refer to the measured relative entropy as the largest possible value when optimising over all POVMs. The measured relative entropies establish a connection between quantum and classical entropies. Moreover, they establish what can in practice be learned from the relative entropy between two states. The measured entropy clearly fulfils the following inequality [Uhl77]:

$$D_M(\rho\|\sigma) \leq D(\rho\|\sigma). \tag{2.70}$$

This equation is a particular case of Theorem 2.33, nonetheless it is much easier to obtain than data processing.

2.4.1 Distinguishability of quantum states

The relative entropy exists in the broader category of distinguishability functions for quantum states. The aim of such a function is to give a quantitative idea of how far apart two elements are, in some sense. to finish this chapter we give a small introduction into distinguishability functions on quantum spaces. For more information, check [NC10, Chapter 9] or [BŻ06].

We introduce two types of distinguishability functions: distances and divergences. We also give some examples, classical and quantum as well as their relation. We start with the definitions.

Definition 2.36
Given a set X a distance is a function $d : X \times X \to \mathbb{R}$ that fulfils the following nice properties:

$$d(x, y) \geq 0 \tag{2.71}$$
$$d(x, y) = 0 \Leftrightarrow x = y \tag{2.72}$$
$$d(x, y) = d(y, x) \tag{2.73}$$
$$d(x, y) \leq d(x, z) + d(z, y). \tag{2.74}$$

A divergence is a function that only fulfils the first two conditions. Distances and divergences are widely used in quantum information as a way to numerically differentiate states. We provide the definition of some of the most widely used distances in quantum information, together with an operational interpretation and their classical counterpart.

The fidelity between two quantum states is defined[5] as

$$F(\rho,\sigma) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right] = \min_{E_j}\sum_i \sqrt{\mathrm{Tr}\left[E_j\rho\right]\mathrm{Tr}\left[E_j\sigma\right]}, \tag{2.75}$$

where the optimisation is taken over all POVMs.

The fidelity is not a distance, but it can be used to define the Bures distance:

$$d_B(\rho,\sigma) = \sqrt{1-F(\rho,\sigma)}. \tag{2.76}$$

Another widely used distance for quantum states is the trace distance:

$$d_T(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1 = \mathrm{Tr}\left[\sqrt{(\rho-\sigma)^*(\rho-\sigma)}\right]. \tag{2.77}$$

The trace distance can also be defined operationally as an optimisation over measurements:

$$d_T(\rho,\sigma) = \frac{1}{2}\max_{E_j}\sum_j |\mathrm{Tr}\left[E_j\rho\right] - \mathrm{Tr}\left[E_j\sigma\right]|. \tag{2.78}$$

The values $\mathrm{Tr}\left[E_j\rho\right]$ and $\mathrm{Tr}\left[E_j\sigma\right]$ are the probabilities given by measuring E on the states. In this sense,

$$\sum_i \sqrt{\mathrm{Tr}\left[E_j\rho\right]\mathrm{Tr}\left[E_j\sigma\right]}, \quad \sum_j |\mathrm{Tr}\left[E_j\rho\right] - \mathrm{Tr}\left[E_j\sigma\right]| \tag{2.79}$$

are the classical and versions of the fidelity and trace distance. Then the optimisation is taken over all classical distributions that can be generated by ρ, σ. The measured entropy performs a similar function for the relative entropy but, unlike the fidelity and trace distance, the maximal measured entropy is not, in general, equal to the relative entropy.

Finally, note that the relative entropies from the previous section are divergences, since they are positive and zero when the input states are equal, but they are not symmetric and do not fulfil the triangle inequality.

[5]Sometimes it is defined as the square of the definition we give.

3

Toward quantum optimal transport

Quantum optimal transport is the main topic of this thesis. This chapter deals with the specific theory necessary for the topic: the introduction of classical optimal transport and the existing literature on quantum optimal transport. Additionally, we develop our first approach to the quantum generalisation, some results and the issues we faced with it.

In Section 3.1 we introduce the theory of classical optimal transport and some of its successes, as a motivation to why we are interested in expanding the theory to quantum spaces. In Section 3.2 we discuss the motivation for quantum optimal transport and the literature regarding the field. In Section 3.3 we define the quantum optimal transport with joint state couplings. In Section 3.4 we show our attempts at giving an operation interpretation to the joint state coupling. Finally, in Section 3.5 we reflect on the results and issues of this approach.

3.1 Classical optimal transport

The theory of classical optimal transport was introduced in 1781 by G. Monge in [Mon81], while working as a civil engineer and considering the problem of optimally transporting soil from excavation sites to construction locations. Since then, the theory has expanded significantly and found numerous important applications. In this thesis, we will only need its definition and a few key properties, which we aim to reproduce in the quantum setting. Our goal is to provide a concise introduction to classical optimal transport, highlight some applications, and point the interested reader to further references.

3.1.1 Formulations of classical optimal transport

The original formulation by Monge is as follows:

Definition 3.1
Let X be a convex metric space and μ_0, μ_1 measures over X with associated densities f_0, f_1 and such that $\mu_0(X) = \mu_1(X)$. In general we can take $\mu_0(X) = 1$. Consider the measurable map $T : X \to X$ that transports *μ_0 into μ_1, in the sense that*

$$T_{\#}\mu_0 = \mu_1. \tag{3.1}$$

Here $T_{\#}\mu_0$ denotes the pushforward measure *of μ_0 under T, defined by*

$$T_{\#}\mu_0(A) := \mu_0(T^{-1}(A)), \qquad \forall A \subseteq X \text{ measurable.} \tag{3.2}$$

Equivalently, $T_{\#}\mu_0$ assigns to each measurable set A the measure that μ_0 gives to its preimage under T. With these hypotheses the Monge problem is defined as:

$$D(\mu_0, \mu_1) = \inf_T \int_X |x - T(x)| f_0(x) dx. \tag{3.3}$$

This definition has some important elements that appear in Eq. (3.3) and condition the cost $D(\mu_0, \mu_1)$. First we start with a metric space X. This metric appears in the integral as $|\cdot|$ and it heavily affects the cost. Second, our inputs to the distinguishability function D are probability measures over X. Finally, we have a set of maps $T : X \to X$ that, in a sense, bring μ_0 to μ_1.

This elements exemplify what the classical transport cost proposed by Monge attempted. Given two measures of equal total mass, find the map that takes one to the other via the shortest path in the underlying metric of X.

This original definition is very similar to the one commonly used with a key difference: the map T only allows us to move all the mass at a certain point x to a single point y. We want a definition that allows us to split the mass, that is instead of a function we want a stochastic map. As mentioned in the introduction, through the conditional probability formula in Eq. (1.1), the definition proposed by Kantorovich [Kan48; Kan58; Kan60] employs joint probability distributions. In addition we change the metric for a generic cost function $c : X \times X \to \mathbb{R}$.

Definition 3.2
With the same hypotheses as in Definition 3.1, an admissible transport plan *is a measure γ on $X \times X$ whose marginals are μ_0 and μ_1, i.e., for all measurable sets $A \subseteq X$,*

$$\gamma(A \times X) = \mu_0(A), \qquad \gamma(X \times A) = \mu_1(A). \tag{3.4}$$

The set of all admissible transport plans is denoted $\Gamma(\mu_0, \mu_1)$. The classical Monge problem can then be equivalently formulated in this relaxed *form, known as the* Monge-Kantorovich problem*:*

$$D(\mu_0, \mu_1) = \inf_{\gamma \in \Gamma(\mu_0, \mu_1)} \int_{X \times X} c(x, y) \, d\gamma(x, y), \tag{3.5}$$

for a given cost function $c(x, y)$.

Note that any admissible plan $\gamma \in \Gamma(\mu_0, \mu_1)$ can be disintegrated as

$$\gamma(x, y) = \mu_0(x) \, t(y|x), \tag{3.6}$$

where $t(y|x)$ is a *conditional transport map* (or *Markov kernel*), i.e., for each fixed x, $t(\cdot|x)$ is a probability measure on X, and for each measurable set $B \subseteq X$, the map $x \mapsto t(B|x)$ is measurable. This object describes the transport mechanism independently of the input distribution: when combined with μ_0, it generates the full transport plan γ.

We can formulate this problem equivalently with the dual formulation proposed by Kantorovich and shown to generally have no gap with the primal problem by Kellerer [Kan48; Kel84]. The dual formulation is as follows:

Definition 3.3
With the same hypotheses as in definition 3.2, the dual formulation of the Monge problem is:

$$D(\mu_0, \mu_1) = \sup_{\phi(y) - \psi(x) \leq c(x,y)} \int_X \phi d\mu_1 - \int_X \psi d\mu_0, \tag{3.7}$$

where $\phi, \psi : X \to \mathbb{R}$.

A particular case of the Monge-Kantorovich problem are the Wasserstein distances [Vas69], a set of problems that appear when the cost function is $c(x, y) = |x - y|^p$. These are defined as

Definition 3.4
With the same hypotheses as Definition 3.2, the p-Wasserstein distance between μ_0 and μ_1 is

$$W_p(\mu_0,\mu_1)^p = \inf_{\gamma\in\Gamma(\mu_0,\mu_1)} \int_{X\times X} |x-y|^p d\gamma. \tag{3.8}$$

As states in the name, these objects are distances. W_2 in particular can also be formulated in a continuous manner, as shown in [BB00]:

Definition 3.5
With the same hypothesis as in definition Definition 3.2, consider a real interval $[0,T]$ and let $f(t,x)$ be a density field such that $f(0,\cdot) = f_0$ and $\mu(T,\cdot) = f_1$[1]. Consider also a velocity field $v(t,x)$ and the condition

$$\partial_t f + \nabla\cdot(fv) = 0. \tag{3.9}$$

Then

$$W_2(\mu_0,\mu_1) = \inf_{(\rho,v)} T\int_X\int_0^T \rho(t,x)|v(t,x)|^2 dxdt. \tag{3.10}$$

In this work we are mostly concerned with the primal formulation in Definition 3.2 and a bit less so the dual formulation in Definition 3.3. The other two definitions have been included for completion, but as seen Definition 3.5 is only applicable to a specific (but important) case.

3.2 Motivation for quantum optimal transport

Classical optimal transport has found myriad applications, culminating in a Nobel prize in Economics in 1975 to Kantorovich and Koopmans ”for their contributions to the theory of optimum allocation of resources” [The75].

Besides resource allocation and economic planning, it has also found application in image processing, fluid dynamics, population dynamics and more [Vil08; Bre15; San18]. These applications do not straightforwardly translate to quantum physics, so why do we want to expand this theory to quantum systems?

The basis to answer to this question can be found in [San18]: ”Wasserstein distances can be used for many other purposes. Essentially, every time that you have a distance on a set you obtain a distance on the probability measures that are defined on it” (p. 12). This strictly applies to Wasserstein distances, which

[1] Recall that to measure μ_i we associated the density f_i

have the property of translating distances in the cost function into distances in the probability space. We take this quotation a bit less literally, focusing on the fact that properties of the set can give rise to distinguishability functions on the probability distribution that express these properties.

In the context of quantum mechanics, we would use states instead of probability measures, and the set on which this act as probabilities can be, for example, an observable from which we could extract some classical probabilities. For the time being, we are going to focus on the labels assigned to pure states. While this labels are a priori arbitrary, in practice useful labels are used, often with numerical values. We consider two important values that can appear in these labels: energy and bitstrings.

Consider a Hamiltonian H on a Hilbert space $\mathcal{H}$ and three pure states: $|0\rangle$, $|1\rangle$, $|10\rangle$ such that $H|n\rangle = n|n\rangle$ for $n = 0,\ 1,\ 10$. The energy gap between these states is clear: $\Delta_{0,1} = 1$, $\Delta_{0,10} = 10$ and $\Delta_{1,10} = 9$. Despite this fact, conventional distinguishability measures on quantum objects such as the fidelity, trace distance, relative entropy... all would give the same value between $0, 1$ and $1, 10$. Taking the fidelity as an example: $F_{0,1} = 0 = F_{1,10}$. This is because these functions all rely exclusively on the vector-space structure of $\mathcal{H}$. The aim of optimal transport is to add another option to what we can use to distinguish quantum states. In the classical case this was given by the cost function $c(x, y)$, which we extend to a cost matrix, usually denoted by K in the quantum setting. This extra variable can be used to encode properties of the system that are not given by the vector-space structure of $\mathcal{H}$.

To give another example, this time more related to information theory, consider bitstrings of lenght n. A natural question in error correction is: given a target string s, how far an output string r is from s. In classical information theory the Hamming distance [Ham50] answers this question. It is defined a the number of different bits between s and r. For example the Hamming distance between 0101 and 0110 is 2, while the Hamming distance between 0110 and 0010 is 1. If we input these two strings into quantum states and then calculate any standard distinguishability function on quantum states we again obtain the same result in both mentioned examples.

This examples gave easy cases where we were given eigenstates of the problem. We would like to obtain a theory of quantum optimal transport that allows us to generalise functions of this type to any pair of quantum states, not only the eigenstates.

3.2.1 Quantum optimal transport in the literature

Our work on quantum optimal transport is not the first contribution to the field. In this section we introduce some other results on quantum optimal transport. We can classify this results into three categories, depending on which of the three classical formulations they are trying to generalise.

The fist quantum generalisation comes from [ŻS98; ŻS01], where they used the classical formulation of the Husimi distributions [Hus40] of quantum states.

Several works have generalized the primal formulation of classical optimal transport, each adopting a different perspective on the coupling [GMP16; Hoo18; DT21; Duv21; FEC+22; BEŻ23; CEF+23; BS25; BPT+25]. Among these, [Hoo18; FEC+22; BEŻ23; CEF+23] develop approaches based on joint state couplings, which we will explore further in the remainder of this section. On the other hand, [DMT+21; DT23] propose a framework grounded in the dual formulation of classical transport and, as a consequence, establish a continuity bound for the von Neumann entropy with respect to their definition.

Finally, for the continuous formulation of W_2 we have [CM14; Agr13; RD19; RD17; CM17; CM19; Wir21].

For further reference, the introduction to [DT23] contains an extensive literature review on the topic for up to 2023 and [Bea25] is a recent full review at the time of writing. It includes an introduction to classical optimal transport and applications as well as different generalisations and applications to quantum systems.

3.3 Quantum couplings in the primal problem

The most naive candidate for the coupling in the quantum case is the quantum analogue to the joint probability distribution: the joint state. In the remainder of this chapter we develop quantum optimal transport with this coupling. We comment on some of the problems that arise with this approach that steered us toward new directions. This work is an extension to our previous work in [Hoo18] shares some similarities [CEF+23; FEC+22; BEŻ23].

3.3.1 The joint quantum state as a coupling

Th definition of the quantum optimal transport with a joint state coupling is fairly straightforward. Consider two Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$, a hermitian cost matrix on the joint space $K \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and two states $\rho \in \mathcal{S}(\mathcal{H}_A)$, $\sigma \in \mathcal{S}(\mathcal{B})$. The set of couplings associated to these two states is

$$\Omega(\rho,\sigma) = \{\omega \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \mathrm{Tr}_A[\omega] = \sigma,\ \mathrm{Tr}_B[\omega] = \rho\}. \tag{3.11}$$

Then the quantum optimal transport cost is

$$\mathcal{C}(\rho,\sigma) = \inf_{\omega\in\Omega(\rho,\sigma)} \mathrm{Tr}[K\omega]. \tag{3.12}$$

A positive definite cost matrix would ensure a positive, cost due to the self duality of the positive cone [BW11]. If $\mathcal{H}_A = \mathcal{H}_B$, a cost symmetric under system swap ensures a symmetric cost. For the other two properties of a distance (Section 2.4.1) two facts were observed by both [Hoo18; CEF+23]: the cost matrix K must be in the antisymmetric subspace and the square root should be added to the resulting number. The fist change appears so that we can have the property that the cost is zero if and only if the states are equal Eq. (2.72), as we show in Proposition 3.7. The second change was shown in [CEF+23] to produce a weak distance[2], which was an actual distance for $d = 2$. With these changes the cost, $\sqrt{T_K^Q(\rho,\sigma)}$ in the notation of [CEF+23], is

$$\sqrt{T_K^Q(\rho,\sigma)} = \inf_{\omega\in\Omega(\rho,\sigma)} \sqrt{\mathrm{Tr}[P_a K P_a \omega]}, \tag{3.13}$$

where P_a denotes the projection on the antisymmetric subspace $P_a = \frac{1}{2}(\mathbb{1} - \mathcal{S})$, with $\mathcal{S}$ the swap operator.

Before showing the need for the antisymmetric projector we show the following proposition:

Proposition 3.6
If either ρ or σ are pure states, then

$$\Omega(\rho,\sigma) = \{\rho\otimes\sigma\}. \tag{3.14}$$

Proof. Let $\sigma = |\psi\rangle\langle\psi|$ and take a state $\omega \in \Omega(\rho,\sigma)$. We can write ω in its diagonal basis $\{|\phi_t\rangle\}$ as

$$\omega = \sum_t p_t |\phi_t\rangle\langle\phi_t|. \tag{3.15}$$

Taking the partial trace over the first subsystem:

$$|\psi\rangle\langle\psi| = \mathrm{Tr}_A[\omega] = \sum_t p_t \mathrm{Tr}_A[|\phi_t\rangle\langle\phi_t|] = \sum_t p_t\rho_t \tag{3.16}$$

[2] A weak distance is a quantity lower bounded by a distance.

Because the coefficients p_t are positive, then $\rho_t = |\psi\rangle\langle\psi|$ for all t. We now want to study the pure states that form ω, $|\phi_t\rangle$. We know that these states fulfil $\mathrm{Tr}_A\left[|\phi_t\rangle\langle\phi_t|\right] = |\psi\rangle\langle\psi|$. From the Schmidt decomposition [Sch07; Eve57; EK95], we can write $|\phi_t\rangle$ as

$$|\phi_t\rangle = \sum_i \lambda_i \left|e_i, f_i\right\rangle \tag{3.17}$$

for orthonormal basis $\{|e_i\rangle\}$ and $\{|f_i\rangle\}$. Its partial trace is

$$|\psi\rangle\langle\psi| = \mathrm{Tr}_A\left[|\phi_t\rangle\langle\phi_t|\right] = \sum_i \lambda_i^2 \left|f_i\rangle\langle f_i\right|. \tag{3.18}$$

Since $\{|f_i\rangle\}$ is an orthonormal basis there must exist, for each t, i_0^t such that $\left|f_{i_0^t}\right\rangle = |\psi\rangle$ and $\lambda_{i_0^t} = 1$, $\lambda_i = 0$ for $i \neq i_0^t$. Thus

$$|\phi_t\rangle\langle\phi_t| = \left|e_{i_0^t}\rangle\langle e_{i_0^t}\right| \otimes |\psi\rangle\langle\psi|. \tag{3.19}$$

Using the above expression to rewrite ω we obtain

$$\omega = \sum_t p_t \left|e_{i_0^t}\rangle\langle e_{i_0^t}\right| \otimes |\psi\rangle\langle\psi| = \left(\sum_t p_t \left|e_{i_0^t}\rangle\langle e_{i_0^t}\right|\right) \otimes |\psi\rangle\langle\psi| \tag{3.20}$$

and the partial trace over the second subsystem is trivial and has to be ρ, so

$$\sum_t p_t \left|e_{i_0^t}\rangle\langle e_{i_0^t}\right| = \rho \tag{3.21}$$

and thus

$$\omega = \rho \otimes |\psi\rangle\langle\psi|. \tag{3.22}$$

□

Proposition 3.7
Let K be a positive operator acting on a finite dimensional Hilbert space $\mathcal{H}$. The cost

$$\mathcal{C}(\rho,\sigma) = \inf_\omega \mathrm{Tr}\left[P_a K P_a \omega\right] \tag{3.23}$$

fulfils $\mathcal{C}(\rho,\rho) = 0$ for all $\rho \in \mathcal{S}(\mathcal{H})$. Moreover, P_a is the largest projector we can conjugate in this way that yields the property.

Proof. Let $\rho \in \mathcal{S}(\mathcal{H})$ and take an orthonormal basis $\{|\varphi_i\rangle\}$ that diagonalises ρ so that

$$\rho = \sum_i p_i |\varphi_i\rangle\langle\varphi_i| . \tag{3.24}$$

Consider the pure state in $\mathcal{H}^2$

$$|\xi\rangle = \sum_i \sqrt{p_i} |\varphi_i, \varphi_i\rangle \tag{3.25}$$

and its associated density matrix

$$\omega = |\xi\rangle\langle\xi| = \sum_{ij} \sqrt{p_i p_j} |\varphi_i, \varphi_i\rangle\langle\varphi_j, \varphi_j| . \tag{3.26}$$

The partial traces of ω are

$$\begin{aligned} \mathrm{Tr}_B[\omega] = \mathrm{Tr}_B\left[\sum_{ij} \sqrt{p_i p_j} |\varphi_i, \varphi_i\rangle\langle\varphi_j, \varphi_j|\right] &= \sum_{ij} \sqrt{p_i p_j} |\varphi_i\rangle\langle\varphi_j| \langle\varphi_j|\varphi_i\rangle \\ = \sum_{ij} \sqrt{p_i p_j} |\varphi_i\rangle\langle\varphi_j| \delta_{ij} = \sum_i p_i |\varphi_i\rangle\langle\varphi_i| = \rho \end{aligned} \tag{3.27}$$

and, similarly, $\mathrm{Tr}_A[\omega] = \rho$. Therefore $\omega \in \Omega(\rho, \rho)$. ω is symmetric by construction therefore its antisymmetric projection is 0. The cost associated to ω is also 0 and since

$$0 \leq \mathcal{C}(\rho, \rho) \leq \mathrm{Tr}[K P_a \omega P_a] = 0 \tag{3.28}$$

we have that $\mathcal{C}(\rho, \rho) = 0$.

To see that the antisymmetric projector is the largest projector that fulfils the proposition consider any pure state $|\phi\rangle \in \mathcal{H}$. By Proposition 3.6 the only admissible coupling is $|\phi\phi\rangle\langle\phi\phi|$. Let

$$0 = \mathcal{C}(|\phi\rangle\langle\phi|, |\phi\rangle\langle\phi|) = \mathrm{Tr}[K |\phi\rangle\langle\phi| \otimes |\phi\rangle\langle\phi|] . \tag{3.29}$$

If this equation is true for any state it is in particular true for the following integral:

$$N \,\mathrm{Tr}\left[K \int |\phi\phi\rangle\langle\phi\phi| \, d\phi\right] = 0, \tag{3.30}$$

where N is a normalisation constant. The object

$$\int |\phi\phi\rangle\langle\phi\phi| \, d\phi \tag{3.31}$$

is a definition of the symmetric projector, therefore K has to be in the orthogonal space to the symmetric projector. The largest projector that fulfils this is the antisymmetric projector. □

3.4 Operational interpretation of the joint state coupling

As stated in the previous section, we aim to find a way to compose joint states. This is achieved in Proposition 3.8. Note this section is similar to the ideas in [Lei06]. We expand on this connection later.

Proposition 3.8
Let ρ and σ be 2 full-rank quantum states on the same Hilbert space $\mathcal{H}$. Let

$$\Omega(\rho,\sigma) = \{\omega_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \text{ s.t. } \mathrm{Tr}_B[\omega_{AB}] = \rho,\ \mathrm{Tr}_A[\omega_{AB}] = \sigma\} \tag{3.32}$$

and

$$\mathcal{E}(\rho,\sigma) = \{\mathcal{E} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \text{ CPTP s.t. } \mathcal{E}(\rho) = \sigma\}. \tag{3.33}$$

Then $\exists$ a vector space isomorphism[3]

$$\begin{aligned} \Xi_\rho : \mathcal{E}(\rho,\sigma) &\longrightarrow \Omega(\rho,\sigma) \\ \mathcal{E} &\longmapsto (\mathrm{id}_A \otimes \mathcal{E})\,(\|\rho\rangle\!\rangle\langle\!\langle\rho\|), \end{aligned} \tag{3.34}$$

where $\|\rho\rangle\!\rangle$ is the canonical purification of $\rho = \sum_i p_i \left|\phi_i\right\rangle\!\left\langle\phi_i\right|$:

$$\|\rho\rangle\!\rangle = \sum_i \sqrt{p_i}\,|\phi_i,\phi_i\rangle\,. \tag{3.35}$$

Proof. The proof is a somewhat lengthy checklist of every property that the maps Ξ_ρ and its proposed inverse should have.

Fix $\rho,\sigma \in \mathcal{S}(\mathcal{H})$, with rho diagonal in the canonical basis without loss of generality. First, note that $\Omega(\rho,\sigma) \neq \emptyset \neq \mathcal{E}(\rho,\sigma)$ because $\rho\otimes\sigma$ and $x \mapsto \sigma$ are elements of each respective set. The map is clearly linear form the definition, as the tensor product is linear, therefore we only need to find an inverse map. Consider the following map from the space of states to the space of channels:

$$\omega_{AB} \mapsto \mathcal{E}_{\omega_{AB}}(X) = \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}} U_\rho X^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\omega_{AB}\right], \tag{3.36}$$

with $\rho^T = \sum_i p_i \left|\alpha_i\right\rangle\!\left\langle\alpha_i\right|$ and $U_\rho = \sum_i \left|i\right\rangle\!\left\langle\alpha_i\right|$. We will write maps generated through a state $\mathcal{E}_{\omega_{AB}}$ and states generated from a map $\omega_{\mathcal{E}}$.

[3] A lineal bijective map. Here, linearity is taken as inherited from the parent vector spaces. A more correct statement would be that we have a linear map between parent spaces that restricted is a bijection.

We first need to show that the map and its inverse candidate are well defined:

$$\begin{aligned} \mathcal{E}(\rho) = \sigma \quad &\Rightarrow \quad \mathrm{Tr}_B[\omega_\epsilon] = \rho,\ \mathrm{Tr}_A[\omega_\epsilon] = \sigma, \\ \mathrm{Tr}_B[\omega_{AB}] = \rho,\ \mathrm{Tr}_A[\omega_{AB}] = \sigma \quad &\Rightarrow \quad \mathcal{E}(\rho) = \sigma, \end{aligned} \tag{3.37}$$

and that they are indeed inverse of each other:

$$\begin{aligned} \mathcal{E}_{\omega_\mathcal{E}} &= \mathcal{E}, \\ \omega_{\mathcal{E}_{\omega_{AB}}} &= \omega_{AB}. \end{aligned} \tag{3.38}$$

We show first that Ξ_ρ is well defined. Let $\mathcal{E} \in \mathcal{E}(\rho, \sigma)$. Then

$$\begin{aligned} \mathrm{Tr}_A\left[\omega_\mathcal{E}\right] = \mathrm{Tr}_A\left[\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\|\rho\rangle\!\rangle\langle\!\langle\rho\|\right)\right] &= \mathrm{Tr}_A\left[\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\sum_{ij}\sqrt{\rho_i\rho_j}\,|i,i\rangle\,\langle j,j|\right)\right] \\ &= \mathrm{Tr}_A\left[\sum_{ij}\sqrt{\rho_i\rho_j}\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(|i\rangle\!\langle j| \otimes |i\rangle\!\langle j|\right)\right] \\ &= \mathrm{Tr}_A\left[\sum_{ij}\sqrt{\rho_i\rho_j}\,|i\rangle\!\langle j| \otimes \mathcal{E}\left(|i\rangle\!\langle j|\right)\right] \\ &= \sum_{ij}\sqrt{\rho_i\rho_j}\delta_{ij}\mathcal{E}\left(|i\rangle\!\langle j|\right) = \sum_i \rho_i \mathcal{E}\left(|i\rangle\!\langle i|\right) \\ &= \mathcal{E}\left(\sum_i \rho_i\,|i\rangle\!\langle i|\right) = \mathcal{E}\left(\rho\right) = \sigma. \end{aligned} \tag{3.39}$$

Similarly:

$$\begin{aligned}
\mathrm{Tr}_B\left[\omega_{\mathcal{E}}\right] &= \mathrm{Tr}_B\left[\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\|\rho\rangle\!\rangle\langle\!\langle\rho\|\right)\right] = \mathrm{Tr}_B\left[\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\sum_{ij}\sqrt{\rho_i\rho_j}\,|i,i\rangle\,\langle j,j|\right)\right] \\
&= \mathrm{Tr}_B\left[\sum_{ij}\sqrt{\rho_i\rho_j}\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(|i\rangle\!\langle j| \otimes |i\rangle\!\langle j|\right)\right] \\
&= \mathrm{Tr}_B\left[\sum_{ij}\sqrt{\rho_i\rho_j}\,|i\rangle\!\langle j| \otimes \mathcal{E}\left(|i\rangle\!\langle j|\right)\right] \\
&= \sum_{ij}\sqrt{\rho_i\rho_j}\,|i\rangle\!\langle j|\,\mathrm{Tr}\left[\mathcal{E}\left(|i\rangle\!\langle j|\right)\right] = \sum_{ij}\sqrt{\rho_i\rho_j}\,|i\rangle\!\langle j|\,\mathrm{Tr}\left[|i\rangle\!\langle j|\right] \\
&= \sum_{ij}\sqrt{\rho_i\rho_j}\,|i\rangle\!\langle j|\,\delta_{ij} = \sum_i \rho_i\,|i\rangle\!\langle i| = \rho
\end{aligned} \tag{3.40}$$

Next we show the proposed inverse is well defined. Let $\omega \in \Omega(\rho,\sigma)$. Then

$$\begin{aligned}
\mathcal{E}_\omega(\rho) &= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho\rho^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\omega\right] \\
&= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}\rho\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\omega\right] \\
&= \mathrm{Tr}_A\left[\left(\mathbb{1}_A \otimes \mathbb{1}_B\right)\omega\right] = \sigma.
\end{aligned} \tag{3.41}$$

Next we show $\mathcal{E}_{\omega_{\mathcal{E}}} = \mathcal{E}$. For this, write a general element $X = \sum_{lm} x_{lm}\,|l\rangle\!\langle m|$ in the (canonical) basis of ρ. Then

$$\begin{aligned}
\mathcal{E}_{\omega_{\mathcal{E}}}(X) &= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho X^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\omega_{\mathcal{E}}\right] \\
&= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho X^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\|\rho\rangle\!\rangle\langle\!\langle\rho\|\right)\right] \\
&= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho X^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\left(\mathbb{1}_A \otimes \mathcal{E}\right)\left(\sum_{kj}\sqrt{\rho_k\rho_j}\,|k\rangle\!\langle j| \otimes |k\rangle\!\langle j|\right)\right] \\
&= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho X^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\left(\sum_{kj}\sqrt{\rho_k\rho_j}\left(\mathbb{1}_A \otimes \mathcal{E}\right)|k\rangle\!\langle j| \otimes |k\rangle\!\langle j|\right)\right] \\
&= \mathrm{Tr}_A\left[\left(\rho^{-\frac{1}{2}}U_\rho X^T U_\rho^*\rho^{-\frac{1}{2}} \otimes \mathbb{1}_B\right)\left(\sum_{kj}\sqrt{\rho_k\rho_j}\,|k\rangle\!\langle j| \otimes \mathcal{E}\left(|k\rangle\!\langle j|\right)\right)\right]
\end{aligned} \tag{3.42}$$

$$
\begin{aligned}
&= \sum_{kj} \sqrt{\rho_k\rho_j}\, \mathrm{Tr}_A \left[\left(\rho^{-\frac{1}{2}} U_\rho X^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_B \right) |k\rangle\langle j| \otimes \mathcal{E}\left(|k\rangle\langle j|\right) \right] \\
&= \sum_{kj} \sqrt{\rho_k\rho_j}\, \langle j| \rho^{-\frac{1}{2}} U_\rho X^T U_\rho^* \rho^{-\frac{1}{2}} |k\rangle\, \mathcal{E}\left(|k\rangle\langle j|\right) \\
&= \sum_{kjml} \sqrt{\rho_k\rho_j}\, \langle j| \rho^{-\frac{1}{2}} U_\rho \left(x_{lm} |l\rangle\langle m|\right)^T U_\rho^* \rho^{-\frac{1}{2}} |k\rangle\, \mathcal{E}\left(|k\rangle\langle j|\right) \\
&= \sum_{kjml} x_{lm} \sqrt{\rho_k\rho_j}\, \langle j| \rho^{-\frac{1}{2}} U_\rho |\alpha_m\rangle\langle \alpha_l| U_\rho^* \rho^{-\frac{1}{2}} |k\rangle\, \mathcal{E}\left(|k\rangle\langle j|\right) \\
&= \sum_{kjml} x_{lm} \sqrt{\rho_k\rho_j}\, \langle j| \rho^{-\frac{1}{2}} |m\rangle\langle l| \rho^{-\frac{1}{2}} |k\rangle\, \mathcal{E}\left(|k\rangle\langle j|\right) \\
&= \sum_{kjml} x_{lm} \sqrt{\rho_k\rho_j} \rho_m^{-\frac{1}{2}} \delta_{jm} \rho_l^{-\frac{1}{2}} \delta_{lk} \mathcal{E}\left(|k\rangle\langle j|\right) = \sum_{ml} x_{lm} \mathcal{E}\left(|l\rangle\langle m|\right) \\
&= \mathcal{E}\left(\sum_{lm} x_{lm} |l\rangle\langle m| \right) = \mathcal{E}(X)
\end{aligned}
$$

Finally, we show $\omega_{\mathcal{E}_{\omega_{AB}}} = \omega_{AB}$. Let $\omega \in \Omega(\rho, \sigma)$ be written as

$$
\omega = \sum_{klmn} \omega_{klmn} |k, l\rangle\langle m, n| \,. \tag{3.43}
$$

Then

$$
\begin{aligned}
\omega_{\mathcal{E}_\omega} &= \left(\mathbb{1}_A \otimes \mathcal{E}_\omega\right) \left(\|\rho\rangle\!\rangle\langle\!\langle\rho\|\right) = \left(\mathbb{1}_A \otimes \mathcal{E}_\omega\right) \left(\sum_{ij} \sqrt{\rho_i\rho_j}\, |i\rangle\langle j| \otimes |i\rangle\langle j| \right) \\
&= \sum_{ij} \sqrt{\rho_i\rho_j}\, |i\rangle\langle j| \otimes \mathcal{E}\left(|i\rangle\langle j|\right) \\
&= \sum_{ij} \sqrt{\rho_i\rho_j}\, |i\rangle\langle j| \otimes \mathrm{Tr}_A \left[\left(\rho^{-\frac{1}{2}} U_\rho \left(|i\rangle\langle j|\right)^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_B \right) \omega \right] \\
&= \sum_{ij} |i\rangle\langle j| \otimes \mathrm{Tr}_A \left[\left(|j\rangle\langle i| \otimes \mathbb{1}_B\right) \omega \right] \qquad (3.44) \\
&= \sum_{ij} |i\rangle\langle j| \otimes \mathrm{Tr}_A \left[\left(|j\rangle\langle i| \otimes \mathbb{1}_B\right) \left(\sum_{klmn} \omega_{klmn} |k, l\rangle\langle m, n| \right) \right] \\
&= \sum_{ijklmn} \omega_{klmn} |i\rangle\langle j|\, \mathrm{Tr}_A \left[|j\rangle \langle i|k\rangle \langle m| \otimes |l\rangle\langle n| \right]
\end{aligned}
$$

$$= \sum_{ijklmn} \omega_{klmn} \delta_{ik} \delta_{jm} |i\rangle\langle j| \otimes |l\rangle\langle n|$$
$$= \sum_{klmn} \omega_{klmn} |k, l\rangle\langle m, n| = \omega.$$

□

Proposition 3.8 gives us complete relations between the objects (ρ, σ), ω and $(\rho, \mathcal{E})$:

i) Given (ρ, σ) there exists a nonempty set of bipartite states that reduce to (ρ, σ) and a nonempty set of channels that take ρ to σ. Moreover, we can trivially find and element of each of these sets.

ii) Given a bipartite state ω we can find ρ, σ such that $\omega \in \Omega(\rho, \sigma)$ through the partial traces and given a pair $(\rho, \mathcal{E})$, we can find σ such that by applying the channel to ρ.

iii) Finally, Proposition 3.8 tells us that these two sets are isomorphic, that is that given $(\rho, \mathcal{E})$, we can uniquely associate it to a bipartite state ω and vice versa.

iv) If the states are not full-rank the maps still work (with the pseudo-inverse in the place of $\rho^{\frac{1}{2}}$), but Ξ_ρ or its inverse looses injectivity.

3.4.1 Recovery map through the state channel relation

In Proposition 3.8 we find a morphism between state-channel pairs and bipartite states. Here we want to use this isomorphism to define a recovery map $\mathcal{E}^{-1}$. We define $\mathcal{E}^{-1}$ by composing the following operations:

$$(\rho, \mathcal{E}) \mapsto \omega_{\mathcal{E}} \mapsto \mathcal{S}\omega_{\mathcal{E}}\mathcal{S} \mapsto (\mathcal{E}(\rho), \mathcal{E}_{\mathcal{S}\omega_{\mathcal{E}}\mathcal{S}}), \tag{3.45}$$

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

Definition 3.9
In the notation of Proposition 3.8, and writing $\mathcal{S}(\omega) = \mathcal{S}\omega\mathcal{S}$:

$$\mathcal{E}^{-1} = \left(\Xi_\sigma^{-1} \circ S \circ \Xi_\rho\right)(\mathcal{E}). \tag{3.46}$$

Explicitly,

$$\mathcal{E}^{-1}(x) = \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}}\right) (\mathbb{1}_A \otimes \mathcal{E}) (\|\rho\rangle\rangle\langle\langle\rho\|)\right]. \tag{3.47}$$

This recovery channel yields the original input state ρ on input σ, the original output state. We see that this is indeed the case:

Lemma 3.10
Let $\mathcal{E}^{-1}$ be as in the previous definition for given $(\rho, \sigma, \mathcal{E})$ such that $\mathcal{E}(\rho) = \sigma$. Then

$$\mathcal{E}^{-1}(\sigma) = \rho.$$

Proof. The proof is straightforward from the definition:

$$\begin{aligned}\mathcal{E}^{-1}(\sigma) &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma \sigma^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) (\mathbb{1}_A \otimes \mathcal{E}) \left(\|\rho\rangle\!\rangle\langle\!\langle\rho\| \right) \right] \\ &= \mathrm{Tr}_B \left[(\mathbb{1}_A \otimes \mathcal{E}) \left(\|\rho\rangle\!\rangle\langle\!\langle\rho\| \right) \right] = \mathrm{Tr}_B\, \omega = \rho.\end{aligned} \tag{3.48}$$

□

We want to see that this recovery channel yields reasonable results in some special cases:

Example 3.11

0) Let $\omega = \rho \otimes \sigma$. In this case the associated channel $\mathcal{E}$ is

$$\begin{aligned}\mathcal{E}(x) &= \mathrm{Tr}_A \left[\left(\rho^{-\frac{1}{2}} U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_B \right) (\rho \otimes \sigma) \right] \\ &= \mathrm{Tr}_A \left[\left(x^T \otimes \mathbb{1}_B \right) (\mathbb{1}_A \otimes \sigma) \right] = \mathrm{Tr} \left[x^T \right] \sigma.\end{aligned} \tag{3.49}$$

Thus, $\mathcal{E}$ is the constant channel. Now:

$$\begin{aligned}\mathcal{E}^{-1}(x) &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) (\mathbb{1}_A \otimes \mathcal{E}) \left(\|\rho\rangle\!\rangle\langle\!\langle\rho\| \right) \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) \sum_{ij} \sqrt{p_i p_j}\, |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) \sum_{ij} \sqrt{p_i p_j}\, |i\rangle\langle j| \otimes \mathrm{Tr} \left[|j\rangle\langle i| \right] \sigma \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) \left(\sum_{ij} \sqrt{p_i p_j} \delta_{ij}\, |i\rangle\langle j| \otimes \sigma \right) \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes x^T \right) (\rho \otimes \mathbb{1}_B) \right] = \mathrm{Tr} \left[x^T \right] \rho.\end{aligned} \tag{3.50}$$

1) *Let $\sigma = \rho$ and $\mathcal{E} = \mathrm{id}$. Then*

$$\mathcal{E}^{-1}(x) = \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \rho^{-\frac{1}{2}} U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \right) \|\rho\rangle\rangle\langle\langle\rho\| \right]. \tag{3.51}$$

This is not immediately recognisable, but we can use the expression of $\mathcal{E} = \mathrm{id}$ in this form from Proposition 3.8:

$$\mathrm{id}(x) = \mathcal{E}(x) = \mathrm{Tr}_A \left[\left(\rho^{-\frac{1}{2}} U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_B \right) \|\rho\rangle\rangle\langle\langle\rho\| \right]. \tag{3.52}$$

Since $\|\rho\rangle\rangle\langle\langle\rho\|$ is invariant under swap we can identify these two channels, thus $\mathcal{E}^{-1} = \mathrm{id}$.

2) *Classical case: $\rho = \sum_i p_i |i\rangle\langle i|$, $\sigma = \sum_i s_i |i\rangle\langle i|$ and*

$$\mathcal{E}(x) = \sum_i \langle i|x|i\rangle \sum_j p_{j|i} |j\rangle\langle j| \tag{3.53}$$

which, evaluated on ρ yields

$$\mathcal{E}(\rho) = \sum_i \langle i|\rho|i\rangle \sum_j p_{j|i} |j\rangle\langle j| = \sum_j \left(\sum_i p_{j|i} p_i \right) |j\rangle\langle j| = \sum_j s_j |j\rangle\langle j|. \tag{3.54}$$

Note that because ρ and σ commute we can take the partial transpose on the canonical basis, which yields $U_\sigma = \mathbb{1}$. Then the recovery channel is

$$\begin{aligned} \mathcal{E}^{-1}(x) &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \right) (\mathbb{1}_A \otimes \mathcal{E}) (\|\rho\rangle\rangle\langle\langle\rho\|) \right] \qquad (3.55) \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \right) (\mathbb{1}_A \otimes \mathcal{E}) \left(\sum_{ij} \sqrt{p_i p_j} |ii\rangle\langle jj| \right) \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \right) \left(\sum_{ij} \sqrt{p_i p_j} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|) \right) \right] \\ &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \right) \right. \\ &\qquad \left. \left(\sum_{ij} \sqrt{p_i p_j} |i\rangle\langle j| \otimes \sum_k \langle k||i\rangle\langle j||k\rangle \sum_l p_{l|k} |l\rangle\langle l| \right) \right] \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \right) \left(\sum_i p_i \left|i\right\rangle\!\left\langle i\right| \otimes \sum_l p_{l|i} \left|l\right\rangle\!\left\langle l\right| \right) \right] \\
&= \sum_{il} p_{l|i} p_i \left\langle l\right| \sigma^{-\frac{1}{2}} x^T \sigma^{-\frac{1}{2}} \left|l\right\rangle \left|i\right\rangle\!\left\langle i\right| = \sum_{il} \frac{p_{l|i} p_i}{s_l} \left\langle l\right| x^T \left|l\right\rangle \left|i\right\rangle\!\left\langle i\right| \\
&= \sum_{il} p_{i|l} \left\langle l\right| x \left|l\right\rangle \left|i\right\rangle\!\left\langle i\right| ,
\end{aligned}
$$

which is the classical recovery channel on input diagonal in the canonical basis.

3) *Consider a classical-quantum channel with classical input, that is* $\rho = \sum_i p_i \left|i\right\rangle\!\left\langle i\right|$, $\mathcal{E} = \sum_i \left\langle i|x|i\right\rangle \sigma_i$; *therefore* $\sigma = \mathcal{E}(\rho) = \sum_i p_i \sigma_i$. *Then*

$$
\begin{aligned}
\omega &= (\mathbb{1}_A \otimes \mathcal{E}) \left(\|\rho\rangle\!\rangle\langle\!\langle\rho\| \right) = (\mathbb{1}_A \otimes \mathcal{E}) \left(\sum_{ij} \sqrt{p_i p_j} \left|i\right\rangle\!\left\langle j\right| \otimes \left|i\right\rangle\!\left\langle j\right| \right) \\
&= \sum_{ij} \sqrt{p_i p_j} \left|i\right\rangle\!\left\langle j\right| \otimes \delta_{ij} \sigma_i = \sum_i p_i \left|i\right\rangle\!\left\langle i\right| \otimes \sigma_i,
\end{aligned}
\tag{3.56}
$$

and the recovery channel is

$$
\begin{aligned}
\mathcal{E}^{-1} &= \mathrm{Tr}_B \left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \right) \sum_i p_i \left|i\right\rangle\!\left\langle i\right| \otimes \sigma_i \right] \\
&= \sum_i \mathrm{Tr}_B \left[\left|i\right\rangle\!\left\langle i\right| \otimes p_i \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \sigma_i \right] \\
&= \sum_i \mathrm{Tr} \left[p_i U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} \sigma_i \sigma^{-\frac{1}{2}} \right] \left|i\right\rangle\!\left\langle i\right| .
\end{aligned}
\tag{3.57}
$$

Since

$$
\sum_i p_i \sigma^{-\frac{1}{2}} \sigma_i \sigma^{-\frac{1}{2}} = \sigma^{-\frac{1}{2}} \left(\sum_i p_i \sigma_i \right) \sigma^{-\frac{1}{2}} = \mathbb{1},
\tag{3.58}
$$

this forms a POVM (we could use that the trace is invariant under transpose to find it in proper form $\mathrm{Tr}\left[x M_y\right]$*) and encodes the output in an orthonormal basis, thus forming a quantum-classical channel.*

Note that we could now reverse the qc-channel to find a cq-channel.

4) *Let* $\mathcal{E}(x) = UxU^*$. *Then* $\sigma = \mathcal{E}(\rho) = \sum_i p_i U |i\rangle\langle i| U^* = \sum_i p_i |\sigma_i\rangle\langle\sigma_i|$, *and*

$$\begin{aligned}\omega = (\mathbb{1}_A \otimes \mathcal{E})\left(\|\rho\rangle\!\rangle\langle\!\langle\rho\|\right) &= \sum_i \sqrt{p_i p_j}\, |i\rangle\langle j| \otimes U |i\rangle\langle j| U^* \\ &= \sum_i \sqrt{p_i p_j}\, |i\rangle\langle j| \otimes |\sigma_i\rangle\langle\sigma_j|\,.\end{aligned} \tag{3.59}$$

The recovery map then is

$$\begin{aligned}\mathcal{E}^{-1}(x) &= \mathrm{Tr}_B\left[\left(\mathbb{1}_A \otimes \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}}\right) \sum_i \sqrt{p_i p_j}\, |i\rangle\langle j| \otimes |\sigma_i\rangle\langle\sigma_j|\right] \\ &= \sum_i \sqrt{p_i p_j}\, |i\rangle\langle j|\, \mathrm{Tr}\left[\sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} |\sigma_i\rangle\langle\sigma_j|\right] \\ &= \sum_i \sqrt{p_i p_j}\, \langle\sigma_j| \sigma^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* \sigma^{-\frac{1}{2}} |\sigma_i\rangle\, |i\rangle\langle j| \\ &= \sum_i \sqrt{p_i p_j}\, \langle\sigma_j| p_j^{-\frac{1}{2}} U_\sigma x^T U_\sigma^* p_i^{-\frac{1}{2}} |\sigma_i\rangle\, |i\rangle\langle j| \\ &= \sum_i \langle\sigma_j| U_\sigma x^T U_\sigma^* |\sigma_i\rangle\, |i\rangle\langle j| \\ &= \sum_i \left(\langle\sigma_j| U_\sigma x^T U_\sigma^* |\sigma_i\rangle\right)^T U^* |\sigma_i\rangle\langle\sigma_j| U \\ &= U^*\left(\sum_i \left(U_\sigma^* |\sigma_i\rangle\right)^T {x^T}^T \left(\langle\sigma_j| U_\sigma\right)^T |\sigma_i\rangle\langle\sigma_j|\right) U \\ &= U^*\left(\sum_i \left|\sigma_i^T\right\rangle^T x \left\langle\sigma_j^T\right|^T |\sigma_i\rangle\langle\sigma_j|\right) U \\ &= U^*\left(\sum_i \langle\sigma_i| x |\sigma_j\rangle\, |\sigma_i\rangle\langle\sigma_j|\right) U \\ &= U^* x U.\end{aligned} \tag{3.60}$$

5) *Finally, let us now calculate the composition of successive recovery operations. We should obtain the original map. From Lemma 3.10 we have that*

$$\begin{array}{ccccccc} A & \xrightarrow{\mathcal{E}} & B & \xrightarrow{\mathcal{E}^{-1}} & A' & \xrightarrow{(\mathcal{E}^{-1})^{-1}} & B' \\ \rho & \longmapsto & \sigma & \longmapsto & \rho & \longmapsto & \sigma, \end{array} \tag{3.61}$$

where A, B, A', B' are labels that all represent the same Hilbert space. Then

$$\begin{aligned}
\left(\mathcal{E}^{-1}\right)^{-1}(x) &= \operatorname{Tr}_{A'}\left[\left(\rho^{-\frac{1}{2}}U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_{B'}\right)\left(\mathcal{E}^{-1} \otimes \mathbb{1}_{B'}\right)(\|\sigma\rangle\!\rangle\langle\!\langle\sigma\|)\right] \\
&= \operatorname{Tr}_{A'}\left[\left(\rho^{-\frac{1}{2}}U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_{B'}\right)\sum_{ij}\sqrt{s_i s_j}\,\mathcal{E}^{-1}(|\sigma_i\rangle\!\langle\sigma_j|) \otimes |\sigma_i\rangle\!\langle\sigma_j|\right] \\
&= \sum_{ij}\sqrt{s_i s_j}\operatorname{Tr}_{A'}\left[\left(\rho^{-\frac{1}{2}}U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_{B'}\right)\right. \\
&\quad \left.\left(\operatorname{Tr}_B\left[\left(\sigma^{-\frac{1}{2}}U_\sigma\left(|\sigma_i\rangle\!\langle\sigma_j|\right)^T U_\sigma^*\sigma^{-\frac{1}{2}} \otimes \mathbb{1}_{A'}\right)(\mathcal{E}\otimes\mathbb{1}_{A'})(\|\rho\rangle\!\rangle\langle\!\langle\rho\|)\right] \otimes |\sigma_i\rangle\!\langle\sigma_j|\right)\right] \\
&= \sum_{ijkl}\sqrt{s_i s_j p_k p_l}\operatorname{Tr}_{BA'}\left[\left(\mathbb{1}_B \otimes \rho^{-\frac{1}{2}}U_\rho x^T U_\rho^* \rho^{-\frac{1}{2}} \otimes \mathbb{1}_{B'}\right)\right. \\
&\quad \left.\left(\frac{1}{\sqrt{s_i s_j}}|\sigma_j\rangle\!\langle\sigma_i| \otimes \mathbb{1}_{A'} \otimes |\sigma_i\rangle\!\langle\sigma_j|\right)\left(\mathcal{E}\left(|k\rangle\!\langle l|\right) \otimes |k\rangle\!\langle l| \otimes \mathbb{1}_{B'}\right)\right] \\
&= \sum_{ijkl}\sqrt{p_k p_l}\,\langle\sigma_i|\,\mathcal{E}\left(|k\rangle\!\langle l|\right)|\sigma_j\rangle\,\frac{1}{\sqrt{p_k p_l}}\,\langle l|\,U_\rho x^T U_\rho^*\,|k\rangle\,|\sigma_i\rangle\!\langle\sigma_j| \\
&= \sum_{ijkl}\langle\sigma_i|\,\mathcal{E}\left(|k\rangle\!\langle l|\right)|\sigma_j\rangle\,\langle k|\,x\,|l\rangle\,|\sigma_i\rangle\!\langle\sigma_j| \\
&= \sum_{ij}\langle\sigma_i|\,\mathcal{E}\left(\sum_{kl}\langle k|\,x\,|l\rangle\,|k\rangle\!\langle l|\right)|\sigma_j\rangle\,|\sigma_i\rangle\!\langle\sigma_j| \\
&= \sum_{ij}\langle\sigma_i|\,\mathcal{E}\left(x\right)|\sigma_j\rangle\,|\sigma_i\rangle\!\langle\sigma_j| = \mathcal{E}(x).
\end{aligned} \tag{3.62}$$

3.4.2 Composition through the state channel relation

Given ρ, $\mathcal{E}_{B|A}$ and $\mathcal{E}_{C|B}$; there is an extension that allows us to find a tripartite state ω_{ABC} such that its one-system partial traces are ρ, $\mathcal{E}_{B|A}(\rho)$ and$(\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})(\rho)$. The construction is the following:

$$\omega_{ABC} = (\mathrm{id}_A \otimes \mathcal{E}_{B|A} \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}))(\|\rho\rangle\!\rangle\langle\!\langle\rho\|), \tag{3.63}$$

where $\|\rho\rangle\!\rangle$ is the 3-system canonical purification of ρ:

$$\|\rho\rangle\!\rangle = \sum_i \sqrt{p_i}\,|iii\rangle\,. \tag{3.64}$$

This construction yields a positive operator that preserves local states and is extendable to n systems but, due to the lack of convex linearity, does not fulfil

$\mathrm{Tr}_A[\omega_{ABC}] = \omega_{BC}$ or $\mathrm{Tr}_B[\omega_{ABC}] = \omega_{AC}$. We can for example see that the second equality is not fulfilled:

$$\begin{aligned}
\mathrm{Tr}_B[\omega_{ABC}] &= \mathrm{Tr}_B\left[(\mathrm{id}_A \otimes \mathcal{E}_{B|A} \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}))(\|\rho\rangle\!\rangle\langle\!\langle\rho\|)\right] \\
&= \mathrm{Tr}_B\left[\left(\mathrm{id}_A \otimes \mathcal{E}_{B|A} \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})\right)\left(\sum_{ij} \sqrt{p_i p_j}\,|iii\rangle\langle jjj|\right)\right] \\
&= \sum_{ij} \sqrt{p_i p_j}\,|i\rangle\langle j| \otimes \mathrm{Tr}\left[\mathcal{E}_{B|A}(|i\rangle\langle j|)\right] \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})(|i\rangle\langle j|) \\
&= \sum_{ij} \sqrt{p_i p_j}\delta_{ij}\,|i\rangle\langle j| \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})(|i\rangle\langle j|) \\
&= \sum_{i} p_i\,|i\rangle\langle i| \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})(|i\rangle\langle i|).
\end{aligned} \tag{3.65}$$

This is different from

$$\begin{aligned}
\omega_{AC} &= (\mathrm{id}_A \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}))\,(\|\rho\rangle\!\rangle\langle\!\langle\rho\|) \\
&= \sum_{ij} \sqrt{p_i p_j}\,|i\rangle\langle j| \otimes (\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A})(|i\rangle\langle j|).
\end{aligned} \tag{3.66}$$

In general, Proposition 3.8 does not allow for composition due to the non-linear nature of the map[4] [LS13; HHP+17].

3.5 Issues and new direction

The choice of coupling explored in this section has some issues. Conceptually, we have the lack of compositionality of the couplings. Given three states ρ, μ, σ and couplings ω_{AB}, ω_{BC} between ρ, μ and μ, σ, we would like a natural way to define a coupling between ρ, σ. Moreover, the seemingly arbitrary appearance of a projection on the antisymmetric space, which deletes information about the coupling, and the appearance of a square root on the final result are unsatisfactory.

Observe that these problems do not appear in the classical case. Given two joint probability distributions, $p(x,y)$ and $q(y,z)$ such that $p(y) = q(y)$, we can easily use Eq. (1.1) to generate processes to 'transport' the states to one another and a third joint probability distribution as

$$r(x,z) = \sum_y p(x)\frac{p(x,y)}{p(x)}\frac{q(y,z)}{q(y)} = \sum_y p(x)p(y|x)q(z|y). \tag{3.67}$$

[4] With respect to the initial state, it is linear with respect to the map.

Clearly $r(x, z)$ is positive. It also has the correct marginal distributions:

$$\begin{aligned} r(x) &= \sum_z \sum_y p(x)p(y|x)q(z|y) = p(x) \sum_y p(y|x) \sum_z q(z|y) \\ &= p(x) \sum_y p(y|x) = p(x), \\ r(z) &= \sum_x \sum_y p(x)p(y|x)q(z|y) = \sum_y \left(\sum_x p(x)p(y|x) \right) q(z|y) \\ &= \sum_y p(y)q(z|y) = q(z). \end{aligned} \tag{3.68}$$

This issues and the relation between our study of quantum couplings and the work of [Lei06; LS13] on quantum Bayes theorem pushed us to explore this field first to land on a suitable coupling. In the following section we do exactly that.

4

States over time

As shown in Section 3.1 the classical theory of optimal transport, when seen through the primal problem, strongly relies on the definition of the coupling set, $\gamma(\mu, \nu)$. Later in the same Chapter 3, we show the choice of a quantum coupling is non-trivial. With the aim of identifying a suitable coupling, in this chapter we focus on understanding the structure of quantum *states over time.* Later in Chapter 5 and Chapter 7 we use these states over time; first as the couplings for quantum optimal transport and later to generate classical probability distributions and quantum partitions that allow us to provide a single shot chain rule.

In Section 4.1 we introduce the motivation for studying states over time, independently of optimal transport. In Section 4.2 we talk about the literature on states over time, we introduce the candidates and the ties to Section 3.4. In Section 4.3 we further discuss the axioms proposed in [HHP+17]. Finally, in Section 4.4 we talk about the proposed state over time we focus on for quantum optimal transport in future chapters, exploring some of its mathematical properties.

4.1 States in space and states in time

Going back to the introduction, consider

$$p(x, y) = p(y|x)p(x). \tag{4.1}$$

This well known equation relates joint probability distributions $p(x, y)$ with stochastic processes acting on a given initial state $p(x)$.

We gave an example in the introduction of how this equation links two in principle different objects. Figure 4.1 illustrates this idea using quantum systems.

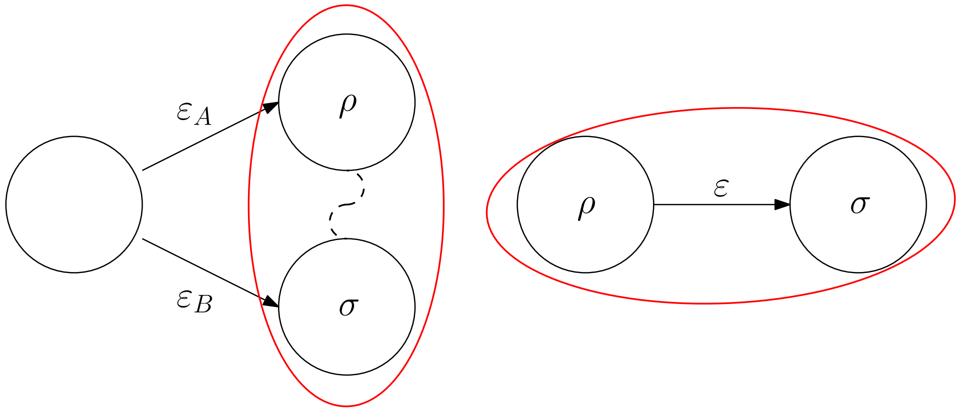

Figure 4.1: Left: depiction of two quantum states that are spatially separated, whose correlations arise from a prior interaction or a shared origin. Right: depiction of two states separated in time, where the correlations instead stem from the dynamical map that transforms the initial state into the final one.

For states separated in space there is a well known quantum object that represents them: the joint state. Given Hilbert spaces, $\mathcal{H}_A$ and $\mathcal{H}_B$, and states $\rho \in \mathcal{S}(\mathcal{H}_a)$, $\sigma \in \mathcal{S}(\mathcal{H}_B)$ the a joint state of ρ and σ is a state $\omega \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\mathrm{Tr}_A[\omega] = \sigma$ and $\mathrm{Tr}_B[\omega] = \rho$.

Unlike in the classical case, from the bipartite quantum state ω— the analogue of the classical joint distribution $p(x, y)$— there is no way to extract a state over time. That is something that encodes an initial state and a channel, the quantum equivalent to $p(y|x)p(x)$. That is, one cannot directly obtain an object that simultaneously encodes an initial state and a dynamical map, the quantum analogue of the factorization $p(y|x)p(x)$.

This questions were asked for the first time around 2005 and have been under study for the last 20 years. The rest of this chapter is dedicated to explaining these results.

4.2 States over time in the literature

The earliest work in this area, to the best of our knowledge, is due to Leifer [Lei06]. In what follows, however, we begin with the contribution of Horsman *et al.* [HHP+17], who collected and systematically compared several proposals for "quantum states over time" that had appeared in the literature. They postulated five desirable properties that such an object should satisfy and analysed which of the existing candidates fulfilled them.

Consider two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, a state $\rho \in \mathcal{S}(\mathcal{H}_A)$ and the Jamiołkowski matrix J associated to a quantum channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$. We use the shorthand notation ρ to stand for $\rho \otimes \mathbb{1}$ when appropriate. This setting will remain standard for the rest of this chapter, unless stated otherwise. The candidates considered in [HHP+17] are as follows:

i) **The Leifer-Spekkens state over time [Lei06; LP08; LS13].** First proposed in [LS13], building on the previous two works. The LS state over time is defined as

$$\rho \star_{LS} J = \rho^{\frac{1}{2}} J \rho^{\frac{1}{2}}. \tag{4.2}$$

ii) **The FJV state over time or Jordan product state over time [FJV15].** Proposed initially by Fitzsimons, Jones and Vedral for qubit spaces. The idea is to define correlations between measurements performed at different times in analogy with spatial correlations. Specifically, one first measures an observable σ_i on system A. The post-measurement state then evolves under the channel $\mathcal{E}$. Finally, a second measurement σ_j is performed on the resulting state. The sequential measurement correlation is then

$$\langle\{\sigma_i, \sigma_j\}\rangle := \sum_{a=\pm 1} a \ \mathrm{Tr}\left[\sigma_j \, \mathcal{E}\left(\Pi_i^a \, \rho \, \Pi_i^a\right)\right], \tag{4.3}$$

where $\Pi_i^{\pm 1} = \frac{1}{2}(\mathbb{1} \pm \sigma_i)$ are projectors onto the eigenstates of σ_i. Here, σ_i denotes the Hermitian Pauli matrices, completed with the identity $\sigma_0 = \mathbb{1}$ to form a complete Hermitian operator basis. This quantity provides the temporal analogue of the spatial correlations $\langle \sigma_i \otimes \sigma_j \rangle$ in a bipartite system.

Equivalently, one can define the FJV state-over-time operator

$$R_{\mathrm{FJV}} = \frac{1}{2} \sum_{i,a=\pm 1} \Pi_i^a \otimes \mathcal{E}(\Pi_i^a \, \rho \, \Pi_i^a), \tag{4.4}$$

such that

$$\langle\{\sigma_i, \sigma_j\}\rangle = \mathrm{Tr}\left[(\sigma_i \otimes \sigma_j)\, R_{\mathrm{FJV}}\right]. \tag{4.5}$$

In turn, one can write the FJV state-over-time in the operator basis as

$$R_{\mathrm{FJV}} = \frac{1}{4}\sum_{i,j=0}^{3} \langle\{\sigma_i, \sigma_j\}\rangle\, \sigma_i \otimes \sigma_j, \tag{4.6}$$

Note that R_{FJV} is Hermitian and reproduces the correct one-time and two-time correlations, but it is not necessarily a positive semi-definite operator, nor are its local reductions.

As shown in [HHP+17], for qubits this definition is equivalent to the Jordan product [Jor32; Jor33; Jor34; JvW35], which can then be applied to arbitrary Hilbert spaces. Thus, the state over time R_{FJV} can also be written as

$$\rho \star_{\mathrm{FJV}} J = \frac{1}{2}\left(\rho J + J\rho\right), \tag{4.7}$$

where J is the Jamiołkowski matrix of the channel $\mathcal{E}$.

iii) **The W state over time.** Here, the idea is to describe a quantum state in a discrete phase space in a manner analogous to a classical joint probability distribution, but where probabilities are replaced by the quasi-probabilities. Specifically, for a d-dimensional system A, a discrete Wigner (W) representation [Woo87] is defined by a set of operators $\Omega^A = \{K_i^A\}$ forming an *operator basis* for $\mathcal{B}(\mathcal{H}_A)$ satisfying

$$\mathrm{Tr}\big[K_i^A K_j^A\big] = d\,\delta_{ij}, \qquad \sum_i K_i^A = d\mathbb{1}, \quad \text{hence } \mathrm{Tr}\big[K_i^A\big] = 1. \tag{4.8}$$

Any density matrix ρ can then be expanded as

$$\rho = \sum_i r^A(i) K_i^A, \tag{4.9}$$

where the real coefficients $r^A(i)$ sum to 1, $\sum_i r^A(i) = 1$, and each lies in $[-1, 1]$. The function $r^A(i)$ is called a *quasi-probability distribution*, providing a discrete phase-space description of the system, analogous to a classical random variable, but allowing for negative values to capture nonclassical features. In the same vein, a quantum channel can be represented in the discrete Wigner representation as

$$E = \frac{1}{d}\sum_{i,j} r_{B|A}(j|i)\, K_i^A \otimes K_j^B, \tag{4.10}$$

where $r_{B|A}(j|i)$ is a real-valued function on $\Omega^A \times \Omega^B$. The factor $1/d$ ensures that $\sum_j r_{B|A}(j|i) = 1$, so the $r_{B|A}(j|i)$ can be interpreted as conditional quasi-probabilities.

Given the quasi-probabilities $r^A(i)$ of the input state ρ, the natural discrete Wigner representation of the composite system AB is

$$r^{AB}(i,j) := r_{B|A}(j|i)\, r^A(i), \tag{4.11}$$

so that the corresponding *W state over time state* reads

$$\rho_{AB}^{(W)} \equiv \sum_{i,j} r_{B|A}(j|i)\, r^A(i)\, K_i^A \otimes K_j^B. \tag{4.12}$$

Expressed directly in terms of ρ and the Jamiołkowski operator, this becomes

$$\rho_{AB}^{(W)} = \rho \star_W J \equiv \frac{1}{d^2} \sum_{i,j} \mathrm{Tr}\left[J\, K_i^A \otimes K_j^B\right]\, \mathrm{Tr}\left[\rho\, K_i^A\right] K_i^A \otimes K_j^B, \tag{4.13}$$

which shows how the state-over-time operator is built from the channel and the initial state in the Wigner representation.

Out of these 3 candidates this thesis mostly focuses on the second one, the Jordan product. In Section 4.4.1 we show the FJV product can not be generalised to arbitrary finite dimension and thus why we use the Jordan product instead of the FJV formulation.

4.2.1 Joint state coupling in the context of states over time

As mentioned in the previous section, the quantum coupling we considered there corresponds to the LS state over time given in i) above. We can develop the expressions in both definitions to make this equivalence evident.

In the product basis that uses the basis of ρ twice, $\rho^{\frac{1}{2}} J \rho^{\frac{1}{2}}$ is

$$\begin{aligned}\rho^{\frac{1}{2}} J \rho^{\frac{1}{2}} &= \sum_{ij} \sqrt{p_i p_j} \, |i\rangle_A\langle i| \, (\mathrm{id}_A \otimes \mathcal{E})(\mathcal{S}) \, |j\rangle_A\langle j| \\ &= \sum_{ijk\ell} \sqrt{p_i p_j} \, |i\rangle_A\langle i| \, (\mathrm{id}_A \otimes \mathcal{E}) \left(|k\ell\rangle\langle \ell k| \right) |j\rangle_A\langle j| \\ &= \sum_{ijk\ell} \sqrt{p_i p_j} \, |i\rangle_A\langle i| \left(|k\rangle\langle \ell| \otimes \mathcal{E}(|\ell\rangle\langle k|) \right) |j\rangle_A\langle j| \\ &= \sum_{ij} \sqrt{p_i p_j} \, |i\rangle\langle j| \otimes \mathcal{E} \left(|j\rangle\langle i| \right) \\ &= \sum_{ij} \sqrt{p_i p_j} (\mathrm{id}_A \otimes \mathcal{E}) \left(|ij\rangle\langle ji| \right) \\ &= (\mathrm{id}_A \otimes \mathcal{E}) \left(\|\rho\rangle\!\rangle\langle\!\langle \rho\| \right) ,\end{aligned} \tag{4.14}$$

which is the definition used in Proposition 3.8 and Section 7.2.

4.3 Axioms for a state over time

Horsman *et al.* [HHP+17] consider five axioms for a quantum state over time. They are as follows.

- **Hermiticity.** Given a state ρ and a Jamiołkowski matrix J, the associated state over time should be Hermitian.
- **Convex Bilinearity.** The state over time should properly preserve statistical mixtures of states and channels. Thus, it is required that

$$\begin{aligned}(p\rho + (1-p)\mu) \star J &= p\,(\rho \star J) + (1-p)\,(\mu \star J)\,, \\ \rho \star (pJ_1 + (1-p)J_2) &= p\,(\rho \star J_1) + (1-p)\,(\rho \star J_2)\,.\end{aligned} \tag{4.15}$$

- **Preservation of the classical limit.** In the classical limit, that is for state $\rho = \sum p_i \, |i\rangle\langle i|$ and a classical channel $\mathcal{E}(\cdot) = \sum_{ij} p(j|i) \, \langle i|\cdot|i\rangle \, |j\rangle\langle j|$ such that the associated Jamiołkowski matrix fulfils $[\rho, J] = 0$, then

$$\rho \star J = \sum_{ij} p(j|i) p(i) \, |ij\rangle\langle ij| \,. \tag{4.16}$$

- **Preservation of marginal states.** The partial traces of the state over time yield the input and output states:

$$\begin{aligned} \mathrm{Tr}_B\,[\rho \star J] &= \rho, \\ \mathrm{Tr}_A\,[\rho \star J] &= \mathcal{E}(\rho). \end{aligned} \tag{4.17}$$

- **Composition.** Finally, given a third Hilbert space $\mathcal{H}_C$ and a Jamiołkowski matrix $J_{B\to C}$, we require the application of composition of maps to be equal to the application of the composition, that is

$$\rho \star (J_{A\to B} \star J_{B\to C}) = (\rho \star J_{A\to B}) \star J_{B\to C}. \tag{4.18}$$

As shown in [HHP+17] the LS state over time violates convex bilinearity and associativity, and the W state over time violates the preservation of the classical limit. As for the Jordan product state over time, [HHP+17] shows that the Jordan product in general violates the compositionality. That said, both [HHP+17] and later [FP22] show that the Jordan product is associative for the relevant objects: $\rho_A \otimes \mathbb{1}_{BC}$, $J_{A\to B} \otimes \mathbb{1}_C$ and $\mathbb{1}_A \otimes J_{B\to C}$. Moreover [LN24] shows that the Jordan product is unique in fulfilling a slightly stronger, and more operationally motivated, set of axioms than the one proposed in [HHP+17] that we have discussed. In particular, [LN24] swaps Hermiticity for the more general time reversal symmetry, which asks that states over time associated to the identity channel are invariant under the swap operator.

4.4 The Jordan product state over time

As we have seen, the Jordan product state over time [FJV15; HHP+17], also called symmetric quantum bloom in literature [FP22; Ful23; FP25], is the best candidate for a state over time due to its unique properties [HHP+17; LN24]. In this section we elaborate on this state over time beyond the axioms. Note that from now on, we refer to the Jordan product state over time as simply *state over time* and we denote the Jordan product by $\star$.

In this section we first define the set of states over time for a pair of quantum spaces; then we show that the original definition in [FJV15] cannot be used in arbitrary dimension spaces; then we show how to recover the input state and the map given an arbitrary state over time and, finally, we show that states over multiple times are well defined.

Definition 4.1
Let $\mathcal{H}_A, \mathcal{H}_B$ be finite dimensional Hilbert spaces and $\mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)$ the set of Jamiołkowski matrices between these two spaces. Let $\rho \in \mathcal{S}(\mathcal{H}_A)$ and $J \in \mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)$. The associated state over time *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:*

$$\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B) = \{\rho \star J | \, J \in \mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B) \text{ and } \rho \in S(\mathcal{H}_A)\} \,. \tag{4.19}$$

4.4.1 General FJV state over time

We want to study whether the FJV state over time, in its original definition, generalises to arbitrary dimension. In fact, we show that it does not generalise to arbitrary dimension, thus giving rise to the necessity of using the Jordan product.

By 'generalising to arbitrary dimension' we mean that for every d, there exists a basis of $\mathrm{Herm}\,\mathbb{C}^d$, $\mathcal{B}_d$, such that

$$\rho \star J = \sum_{\Sigma_A^i, \Sigma_B^j \in \mathcal{B}_d} \left\langle \{\Sigma_A^i, \Sigma_B^j\} \right\rangle \Sigma_A^i \otimes \Sigma_B^j. \tag{4.20}$$

Because $\mathcal{B}_d$ is a basis, this would yield a nice operational interpretation to the state over time. We see in this section this is generally impossible. Because the Jordan product is unique in being fulfilling the axioms [LN24], a state over time defined as in [FJV15] that is not equal to the Jordan product cannot fulfil the axioms, regardless of basis.

First, we accurately define $\langle\{C, D\}\rangle$ for general observables and obtain the following lemma based on the proof of [HHP+17].

Lemma 4.2
Let C, D be observables with spectrums X_C, X_D, ρ a state and J a Jamiołkowski matrix. Then

$$\langle\{C, D\}\rangle = \mathrm{Tr}\left[J \left(\sum_{x \in X_C} x P^x \rho P^x \right) \otimes D \right], \tag{4.21}$$

where P^x is the projector on the subspace of eigenstates of C with eigenvalue x.

Proof. Let $x, y \in X_C, Y_D$ represent eigenvalues of C, D, respectively. Then

$$\begin{aligned}
\langle\{C,D\}\rangle &= \mathbb{E}(xy) = \sum_{x\in X_C, y\in Y_D} xyP(x,y) = \sum_{x\in X_C, y\in Y_D} xyP(y|x)P(x) \\
&= \sum_{x\in X_C, y\in Y_D} xyP(y|x)\operatorname{Tr}\left[P^x\rho P^x\right] \\
&= \sum_{x\in X_C} x\left(\sum_{y\in Y_D} yP(y|x)\right)\operatorname{Tr}\left[P^x\rho P^x\right] \\
&= \sum_{x\in X_C} x\mathbb{E}(y|x)\operatorname{Tr}\left[P^x\rho P^x\right] \\
&= \sum_{x\in X_C} x\operatorname{Tr}\left[\varepsilon\left(\frac{P^x\rho P^x}{\operatorname{Tr}\left[P^x\rho P^x\right]}\right)D\right]\operatorname{Tr}\left[P^x\rho P^x\right] \\
&= \sum_{x\in X_C} x\operatorname{Tr}\left[\varepsilon\left(P^x\rho P^x\right)D\right] \\
&= \sum_{x\in X_C} x\operatorname{Tr}\left[\operatorname{Tr}_A\left[JP^x\rho P^x\right]D\right] \\
&= \operatorname{Tr}\left[J\left(\left(\sum_{x\in X_C} xP^x\rho P^x\right)\otimes D\right)\right].
\end{aligned} \tag{4.22}$$

□

Note that for the particular case $C=\sigma_i$, $D=\sigma_j$ $X_{\sigma_i}=\{\pm 1\}$ we recover Eq. (25) from [HHP+17] and Eq. (4.3)[1].

Lemma 4.2 highlights the importance of $\sum xP^x\rho P^x$ and why this proof works in the case of Pauli matrices: Pauli matrices only have ± 1 eigenvalues, which then allow the projectors to be written as

$$P^{\pm} = \frac{1}{2}\left(\mathbb{1}\pm\sigma_i\right).$$

We expand on this observation in the following lemma. This lemma, together with Lemma 4.2 characterises the key property of the Pauli matrices that allows them to act as observables that characterise the Jordan product.

[1] If we input the Jamiołkowski matrix definition $J = (\mathrm{id}\otimes\mathcal{E})(\mathcal{S})$ and operate a bit.

Lemma 4.3
Let C be an observable not proportional to the identity, X_C its spectrum and $P^x,\ x \in X_C$, the projectors on every eigenspace. Then

$$\sum_{x\in X_C} xP^x\rho P^x = \rho\star C\ \ \forall\rho \quad \Leftrightarrow \quad X_C = \{\pm\lambda\},\ \lambda\in\mathbb{R}\backslash\{0\}. \tag{4.23}$$

Note that the result would hold if $X_C = \{\lambda\}$, since then $C = \lambda\mathbb{1}$ and the left side of the implication is trivial. We explicitly exclude this case in the statement.

Proof. For the $\Leftarrow$ direction, consider an observable C with spectrum $\{\pm\lambda\}$. Then the projectors onto its eigenspaces can be written similarly to the case of the Pauli matrices:

$$P^{\pm} = \frac{1}{2\lambda}\left(\lambda\mathbb{1}\pm C\right). \tag{4.24}$$

Then, a calculation yields

$$\begin{aligned}\sum_{x\in X_C} xP^x\rho P^x &= \lambda P^+\rho P^+ - \lambda P^-\rho P^- \\ &= \frac{1}{4\lambda^2}\left(\lambda\left(\lambda\mathbb{1}+C\right)\rho\left(\lambda\mathbb{1}+C\right) - \lambda\left(\lambda\mathbb{1}-C\right)\rho\left(\lambda\mathbb{1}-C\right)\right) \\ &= \frac{1}{4\lambda^2}\left(\lambda^3\rho+\lambda C\rho C+\lambda^2 C\rho+\lambda^2\rho C-\lambda^3\rho-\lambda C\rho C+\lambda^2\rho C+\lambda^2 C\rho\right) \\ &= \frac{1}{2}\left(\rho C + C\rho\right) = \rho\star C.\end{aligned} \tag{4.25}$$

For $\Rightarrow$, consider the converse statement; if there exists a real number s.t. $\lambda\in X_C$ and a real number $\mu\neq\pm\lambda$ also in the spectrum, then there exists a state ρ such that the equality is false. Consider such an observable C with eigenvalues λ and μ (and possibly others), and corresponding normalized eigenvectors $|\lambda\rangle$ and $|\mu\rangle$. Define the projector onto the subspace orthogonal to these two eigenvectors as

$$P^R = \mathbb{1} - |\mu\rangle\langle\mu| - |\lambda\rangle\langle\lambda|. \tag{4.26}$$

Consider also $\rho = \frac{1}{2}\left(|\mu\rangle+|\lambda\rangle\right)\left(\langle\mu|+\langle\lambda|\right)$. Denote $|w\rangle = \rho\,|\lambda\rangle = \frac{1}{2}\left(|\mu\rangle+|\lambda\rangle\right)$.

Then

$$\begin{aligned}\sum_{x\in X_C} xP^x\rho P^x\left|\lambda\right\rangle - \rho\star C\left|\lambda\right\rangle &= \left(\lambda P^\lambda - \frac{1}{2}\lambda\mathbb{1} - \frac{1}{2}C\right)\left|w\right\rangle \\ &= \left(\lambda P^\lambda - \frac{1}{2}\lambda\left(\left|\mu\right\rangle\!\left\langle\mu\right| + \left|\lambda\right\rangle\!\left\langle\lambda\right| + P^R\right)\right. \\ &\quad \left. - \frac{1}{2}\left(\mu\left|\mu\right\rangle\!\left\langle\mu\right| + \lambda\left|\lambda\right\rangle\!\left\langle\lambda\right| + P^R\sum_{x\in X_C} xP^xP^R\right)\right)\left|w\right\rangle \\ &=\lambda\frac{1}{2}\left|\lambda\right\rangle - \frac{1}{4}\lambda\left(\left|\mu\right\rangle + \left|\lambda\right\rangle\right) - \frac{1}{4}\left(\mu\left|\mu\right\rangle + \lambda\left|\lambda\right\rangle\right) \\ &= -\frac{\lambda+\mu}{4}\left|\mu\right\rangle \neq 0.\end{aligned} \tag{4.27}$$

□

Finally, we can use these results to state to state a general relation between FJV states over time and Jordan product states over time.

Proposition 4.4
Let $n\in\mathbb{N}$. Let σ_i, $i\in\{0,1,2,3\}$ be the Pauli matrices. Then the basis

$$\mathcal{B}_n = \left\{\bigotimes_{i\in[n]}\sigma_{j_i}\right\} \tag{4.28}$$

of $\operatorname{Herm}\mathbb{C}^{2^n}$ *fulfils*

$$J\star\rho = \sum_{\Sigma_A,\Sigma_B\in\mathcal{B}_n} \left\langle\{\Sigma_A,\Sigma_B\}\right\rangle \Sigma_A\otimes\Sigma_B. \tag{4.29}$$

Proof. The spectrum of each element of $\mathcal{B}_n$ is $\{\pm1\}$. Let $\Sigma_A,\Sigma_B\in\mathcal{B}_n$, then by Lemma 4.3

$$\rho\star\Sigma_A = \sum_{x\in\{\pm1\}} xP^x\rho P^x. \tag{4.30}$$

We can now insert this equality into Lemma 4.2 to obtain

$$\left\langle\{\Sigma_A,\Sigma_B\}\right\rangle = \operatorname{Tr}\left[J\left(\rho\star\Sigma_A\right)\otimes\Sigma_B\right] = \operatorname{Tr}\left[\left(\rho\star J\right)\Sigma_A^i\otimes\Sigma_B^j\right]. \tag{4.31}$$

Adding up all possible combinations yields

$$\sum_{\Sigma_A^i,\Sigma_B^j\in\mathcal{B}_n} \left\langle\{\Sigma_A^i,\Sigma_B^j\}\right\rangle \Sigma_A^i\otimes\Sigma_B^j = \sum_{\Sigma_A^i,\Sigma_B^j\in\mathcal{B}_n} \operatorname{Tr}\left[(\rho\star J)\,\Sigma_A^i\otimes\Sigma_B^j\right]\Sigma_A^i\otimes\Sigma_B^j. \tag{4.32}$$

Since $\mathcal{B}_n$ is a basis of $\operatorname{Herm}\mathbb{C}^{2^n}$, the coefficients $\operatorname{Tr}[(\rho\star J)]$ uniquely determine $(\rho\star J)$, yielding the result. □

We have seen that the FJV state over time can be generalised to dimension 2^n with the tensor products of Pauli matrices. More importantly, Lemma 4.3 implies that for Hilbert spaces of dimension $d\neq 2^n$ the FJV state over time can not be generalised with the original form from [FJV15].

Remark 4.5
Known generalisations of Pauli matrices for higher dimensions are the spin matrices and the Gell-Mann matrices [BK08][2]. These matrices all form a basis of $\operatorname{Herm}\mathbb{C}^d$ *but, unlike* $\mathcal{B}_n$*, the elements do not have a spectrum of the form* $\{\pm\lambda\}$*. Therefore, unless there is an unknown generalisation with the correct spectrum for all dimensions, the Jordan product is the only way to extend the FJV state over time to all dimensions.*

4.4.2 Recovering the channel from the state over time

While a state over time defined via the Jordan product has a clear physical interpretation in terms of its factors (the input state and the Jamiołkowski operator of the map) a characterization of which Hermitian matrices ω admit such a decomposition is not known. Even knowingly given a state over time $Q\in\mathcal{Q}(\mathcal{H}_A:\mathcal{H}_B)$, it is not immediately clear how to recover the Jamiołkowski matrix. In this section we show how to identify states over time and recover the associated Jamiołkowski matrix.

We start with the following theorem, which was also shown in [Ful23; LN24].

Theorem 4.6
Let ω *be a Hermitian operator on* $\mathcal{B}(\mathcal{H}_A\otimes\mathcal{H}_B)$ *such that* $\rho=\operatorname{Tr}_B[\omega]\geq 0$*. 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 operator* J *such that, if* p_i *or* p_j *are nonzero then*

$$\langle ik|\,J\,|j\ell\rangle = \frac{2}{p_i+p_j}\langle ik|\,\omega\,|j\ell\rangle\,. \tag{4.33}$$

[2] There is another generalisation: Sylvester's generalised Pauli matrices, but these are not even Hermitian.

Then, $\omega = \rho \star J$. Moreover, if ρ is faithful then J is Hermitian and uniquely specified by the above matrix elements.

Proof. The result can be obtained by writing $\rho \star J$ 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,

$$\rho = \sum_i p_i |i\rangle\langle i| \tag{4.34}$$

$$J = \sum_{ikj\ell} \langle ik| J |j\ell\rangle |ik\rangle\langle j\ell| . \tag{4.35}$$

In this basis we can calculate the Jordan product

$$\begin{aligned} \rho \star J &= \frac{1}{2}\left(\sum_{ikj\ell i'} p_{i'} |i'\rangle\langle i'| \langle ik| J |j\ell\rangle |ik\rangle\langle j\ell| + \langle ik| J |j\ell\rangle |ik\rangle\langle j\ell| p_i |i'\rangle\langle i'|\right) \\ &= \sum_{ikj\ell} \frac{1}{2}(p_i + p_j) \langle ik| J |j\ell\rangle |ik\rangle\langle j\ell| . \end{aligned} \tag{4.36}$$

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

$$\langle ik| J |j\ell\rangle = \frac{2}{p_i + p_j} \langle ik| \rho \star J |j\ell\rangle . \tag{4.37}$$

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

Theorem 4.6 takes an operator $\omega \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with positive partial trace on $\mathcal{H}_A$ and allows us to recover information on operators $J \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that $\omega = \rho \star J$. Note that the Theorem says nothing about whether there is a Jamiołkowski matrix compatible with ω or not.

If ρ is faithful, we obtain a single J compatible with the structure. We can check whether or not this J is a Jamiołkowski matrix to conclude whether or not the original operator ω is a state over time. If ρ is not faithful the coefficients $\langle ik| J |j\ell\rangle$ such that $p_i = p_j = 0$ are not determined by the Theorem. We can fill these coefficients while ensuring that the resulting operator is a Jamiołkowski matrix with the following SDP:

$$\begin{aligned} \min_{J} \quad & f(J) \\ \text{s.t.} \quad & \begin{cases} \langle ik|\, J \,|j\ell\rangle = \dfrac{2}{p_i + p_j} \langle ik|\, \omega \,|j\ell\rangle \,, & \forall i, j \in B \mid p_i + p_j \neq 0,\ \forall k, \ell \in B \\ \operatorname{Tr}_B J = \mathbb{1} \\ J^{T_A} \geq 0 \end{cases} \end{aligned} \tag{4.38}$$

In the SDP, f is an arbitrary linear function, since we are only interested in whether or not the feasible set is empty and, if it is not, obtaining an element of this set, which would be our Jamiołkowski matrix. Appendix A discusses the practical numerical implementation of this SDP, since such feasibility problems can present some computational issues.

This map has also been independently studied in [SAS+25] in a different context, where they provide the following alternative form for the case where ρ is faithful:

$$J = \int_0^\infty e^{-\frac{t}{2}\rho} \omega e^{-\frac{t}{2}\rho} dt. \tag{4.39}$$

This form makes it very easy to prove the following property.

Corollary 4.7
Let $\rho \in \mathcal{S}(\mathcal{H})$ be a state of a Hilbert space $\mathcal{H}$. Then the map

$$x \mapsto \int_0^\infty e^{-\frac{t}{2}\rho} x e^{-\frac{t}{2}\rho} dt \tag{4.40}$$

is completely positive.

Proof. $e^{-\frac{t}{2}\rho}$ will be Hermitian because ρ is. Therefore for a fixed t, the map $x \mapsto e^{-\frac{t}{2}\rho} x e^{-\frac{t}{2}\rho}$ will be CP because it is a Kraus form of a map due to the Hermiticity of $e^{-\frac{t}{2}\rho}$. The integral of CP maps will be CP, thus the original map is CP. □

Additionally, we want to present the following alternative form for the map:

$$J = (U_\rho \otimes \mathbb{1}) \left(U_\rho^* \rho U_\rho \star \left(\left(U_\rho^* \otimes \mathbb{1} \right) \omega \left(U_\rho \otimes \mathbb{1} \right) \right)^{\ominus} \right)^{\ominus} \left(U_\rho^* \otimes \mathbb{1} \right), \tag{4.41}$$

where U_ρ is a unitary that diagonalises ρ from the canonical basis and $\ominus$ 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 $(U_\rho \otimes \mathbb{1}) \cdot \left(U_\rho^* \otimes \mathbb{1}\right)$ and work in the diagonal basis of ρ, as done in Theorem 4.6. Then, note that $\left(U_\rho^* \otimes \mathbb{1}\right) \omega \left(U_\rho \otimes \mathbb{1}\right)$ is just ω written in the diagonal basis of ρ in the first subsystem and the canonical basis in the second, that is

$$\left(\left(U_\rho^* \otimes \mathbb{1}\right) \omega \left(U_\rho \otimes \mathbb{1}\right)\right)_{ikj\ell} = \langle ik| \, \omega \, |j\ell\rangle \, . \tag{4.42}$$

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

$$\left(U_\rho^* \rho U_\rho \star \left(\left(U_\rho^* \otimes \mathbb{1}\right) \omega \left(U_\rho \otimes \mathbb{1}\right)\right)^{\ominus}\right)_{ikj\ell} = \frac{1}{2} \frac{(p_i + p_j)}{\langle ik| \, \omega \, |j\ell\rangle}. \tag{4.43}$$

Now, we only need to invert it element-wise again to obtain the result.

A quantum generalisation of Bayes' Theorem

The previous result serves as a way to recover a map and state from a joint state over time. If the maps and states are classical, we would like that this result recovers the Bayes' Theorem. Here we show that it does:

Remark 4.8
Let $\rho = \sum_i p_i \, |i\rangle\!\langle i|$ and $J_{A\to B} = \sum_{ij} p_{i\to j} \, |ij\rangle\!\langle ij|$, where $p_{i\to j}$ is a classical stochastic map. Then $Q = \sum_{ij} p_{i\to j} p_i \, |ij\rangle\!\langle ij| = \sum_{ij} p_{ij} \, |ij\rangle\!\langle ij|$, where $p_{ij} = p_{i\to j} p_i$ is a joint probability distribution. We can now apply Theorem 4.6 considering B the input space. The partial trace will be $\mathrm{Tr}_A [Q] = \sum_j \left(\sum_i p_{ij}\right) |j\rangle\!\langle j| = \sum_j p_j \, |j\rangle\!\langle j|$. Q is already diagonal in a product basis of the required form so we can directly find $J_{A\leftarrow B} = \sum_{ij} p_{ij}/p_j \, |ij\rangle\!\langle ij| = \sum_{ij} p_{i\leftarrow j} \, |ij\rangle\!\langle ij|$. Combining these expressions, we recover Bayes' Theorem: $p_{ij} = p_{i\leftarrow j} p_j = p_{i\to j} p_i$.

As discussed in Section 5.5 the full inversion of the map's direction is not possible in the quantum case. That is, given a state over time $Q = \rho \star J_{A\to B}$, it is not always possible to find a state σ_B and Jamiołkowski matrix $J_{A\leftarrow B}$ such that $Q = \sigma_B \star J_{A\leftarrow B}$.

4.4.3 States over multiple times

So far we have considered two points in time and a CPTP map connecting them. Part of the advantage of using a coupling that trivially has an interpretation as a channel is that we naturally get states over multiple times. For n times and $n-1$ channels we write the space of states over time associated to this sequence recursively:

Definition 4.9
Consider Hilbert spaces $\mathcal{H}_i$, $i \geq 0$. Then

$$\begin{aligned}\mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_n) &= (\mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_{n-1}) \otimes \mathbb{1}_n) \star \left(\bigotimes_{j=0}^{n-1} \mathbb{1}_j \otimes \mathcal{J}(\mathcal{H}_{n-1} \to \mathcal{H}_n) \right) \\ &\equiv \mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_{n-1}) \star \mathcal{J}(\mathcal{H}_{n-1} \to \mathcal{H}_n).\end{aligned} \tag{4.44}$$

By taking partial traces on specific subsystems we can 'forget' about the state of the system at that slot, and create a state over time over $n-n'$ times, where n' is the number of subsystems that were traced out. Explicitly, $\mathrm{Tr}_i\left[\mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)\right] = \mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_{i-1} : \mathcal{H}_{i+1} : \cdots : \mathcal{H}_n)$. This is true because given two Jamiołkowski matrices $J_{i-1,i},\ J_{i,i+1}$ we can construct a Jamiołkowski matrix

$$\tilde{J}_{i-1,i+1} = \mathrm{Tr}_i\left[(J_{i-1,i} \otimes \mathbb{1}_{i+1}) \star (\mathbb{1}_{i-1} \otimes J_{i,i+1})\right] \tag{4.45}$$

such that the associated channels fulfil $\tilde{\mathcal{E}}_{i-1,i+1} = \mathcal{E}_{i,i+1} \circ \mathcal{E}_{i-1,i}$ [CDP09].

Similarly, it is possible to obtain the state at each time i by tracing out all other times. Indeed, given $Q_n \in \mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)$ the state at time i is $\sigma_i = \mathrm{Tr}_{\hat{i}}\left[Q_n\right]$. In addition, we can recover the channel at each step by tracing out all subsystems but two consecutive ones and then applying Theorem 4.6, using the SDP in Eq. (4.38) if necessary.

5

Optimal transport with Jordan product couplings

This chapter presents the main theoretical framework and general results of our approach to quantum optimal transport. We provide a definition and analyse the properties that a cost matrix must satisfy for the resulting cost to define a distance. Additionally, we examine general properties of interest in this context. In next chapter we will focus particular choices of cost functions.

In Section 5.1 we define the set of coupling based on the Jordan product state over time from Section 4.4. In Section 5.2 we define quantum optimal transport with these couplings. In Section 5.3 we study the conditions for the cost matrix such that the cost is positive; in Section 5.4 the conditions such that the cost is zero for the identity channel; in Section 5.5 the symmetry of the problem; in Section 5.6 the triangle inequality, and in Section 5.7 general properties of the quantum optimal transport.

5.1 The Jordan product state over time as a coupling

We introduced in Section 4.4 the Jordan product as a good candidate for a definition of state over time. We want to use it as a quantum analogue of the joint probability distribution in classical transport theory for the quantum theory. For reference, we restate the definition of states over time via the Jordan product in Definition 4.1, and build our discussion from there.

Definition 5.1
Let $\mathcal{H}_A, \mathcal{H}_B$ be finite dimensional Hilbert spaces and $\mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)$ the set of Jamiołkowski matrices between these two spaces. Let $\rho \in \mathcal{S}(\mathcal{H}_A)$ and $J \in \mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)$. The associated state over time *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:*

$$\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B) = \{\rho \star J | \, J \in \mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B) \text{ and } \rho \in S(\mathcal{H}_A)\} \,. \tag{5.1}$$

From the definition of general states over time between two spaces, we are interested in restricting this set to the states over time between two states. That is, given states $\rho \in \mathcal{S}(\mathcal{H}_A)$ and $\sigma \in \mathcal{S}(\mathcal{H}_B)$ we are interested in the set of states over time generated by channels such that $\mathcal{E}(\rho) = \sigma$. This is written in the following definition.

Definition 5.2
Let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_A), \mathcal{S}(\mathcal{H}_B)$ and $\mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B) = \{J \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B),\ J^{T_A} \geq 0,\ \mathrm{Tr}_B[J] = \mathbb{1}\}$ be the set of Jamiołkowski matrices between these two spaces. The set of states over time between ρ and σ is

$$\begin{aligned} \mathcal{Q}(\rho, \sigma) &= \{\rho \star J | \, J \in \mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B) \text{ and } \mathrm{Tr}_A[\rho J] = \sigma\} \\ &= \{Q \in \mathcal{Q}(\mathcal{H}_A \to \mathcal{H}_B) | \ \mathrm{Tr}_B[Q] = \rho,\ \mathrm{Tr}_A[Q] = \sigma\} \,. \end{aligned} \tag{5.2}$$

The set $\mathcal{Q}(\rho, \sigma)$ contains the quantum couplings between ρ and σ. In the same spirit as classical optimal transport Section 3.1, we want to optimise over this set to find the coupling that optimises a given cost. The main conceptual advantage of this definition is that the resulting optimal coupling has an immediate interpretation as a quantum channel, an interpretation that the definition makes obvious.

With the coupling defined we can finally proceed to building a theory of optimal quantum transport around it. Definition 5.4 below is, in a way, the main result of this chapter, since our main contribution to the field is the implementation of states over time as couplings.

The rest of the chapter is devoted to studying this function's properties. This goes through some technical results on the cone generated by $\mathcal{Q}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)$ in Section 5.3, a simple statement in Section 5.4, some counterexamples in Section 5.5 and more results in Section 5.6 and Section 5.7.

5.2 Definition

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.

Definition 5.3
Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces, let $\rho \in \mathcal{S}(\mathcal{H}_A)$ be a quantum state, and let $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ 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*

$$\kappa(\rho, \mathcal{E}) = \mathrm{Tr}\left[K Q_{\rho,\mathcal{E}}\right], \tag{5.3}$$

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, the one that achieves the smallest possible cost.

Definition 5.4
Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces, let $\rho, \sigma \in \mathcal{S}(\mathcal{H}_{A,B})$ be states on these Hilbert spaces and let $K \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be Hermitian. The quantum optimal transport cost *with cost matrix K between ρ and σ is*

$$\mathcal{K}(\rho, \sigma) = \min_{\mathcal{E}\,|\,Q_{\rho,\mathcal{E}} \in \mathcal{Q}(\rho,\sigma)} \kappa(\rho, \mathcal{E}) = \min_{Q \in \mathcal{Q}(\rho,\sigma)} \mathrm{Tr}\left[KQ\right]. \tag{5.4}$$

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.

In Chapter 6 we study particular cases with specific physical interests in mind. The remainder of this section will be devoted to studying mathematical properties of $\mathcal{K}$, with a particular focus on identifying which set of cost matrices provides a quantum optimal transport cost with certain desired properties.

The optimal cost can be efficiently[1] computed because it can be written as an

[1] See Appendix A for more details.

SDP:

$$\begin{aligned} \min_{J} \quad & \mathrm{Tr}\left[(K \star \rho)J\right] \\ \text{s.t.} \quad & \begin{cases} \mathrm{Tr}_B\, J = \mathbb{1} \\ \mathrm{Tr}_A\left[\rho J\right] = \sigma \\ J^{T_A} \geq 0 \end{cases}, \end{aligned} \tag{5.5}$$

and its dual

$$\begin{aligned} \max_{Y_1, Y_2} \quad & \mathrm{Tr}\left[Y_1\right] + \mathrm{Tr}\left[\sigma Y_2\right] \\ \text{s.t.} \quad & \begin{cases} Y_1 \otimes \mathbb{1} + \rho^T \otimes Y_2 \leq (K \star \rho)^{T_A} \\ Y_1, Y_2 \text{ Hermitian} \end{cases}. \end{aligned} \tag{5.6}$$

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 $\mathrm{Tr}\left[(K \star \rho)J\right] = \mathrm{Tr}\left[K(\rho \star J)\right] = \mathrm{Tr}\left[KQ\right]$. 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.

5.3 Positivity of the cost

5.3.1 The cone of states over time and its dual

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 is Hermitian the cost is a real number, since $\mathrm{Tr}\left[KQ\right] = \langle Q, K\rangle_{HS}$. Due to the fact that the cost is an inner product, it is clear that the set of K for which $\mathcal{K}$ is positive is the dual cone to $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ with respect to the Hilbert-Schmidt inner product. Throughout this section, whenever we refer to duality, we mean it with respect to the Hilbert–Schmidt inner product. Additionally, it is useful to define first the cone generated by $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$.

Definition 5.5
Let $\mathcal{H}_A, \mathcal{H}_B$ be finite dimensional Hilbert spaces and $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ the set of states

over time between these two spaces. The cone of states over time is

$$\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B) = \operatorname{cone}\left(\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)\right) = \operatorname{cone}\left(\bigcup_{\substack{\rho \in S(\mathcal{H}_A) \\ \sigma \in S(\mathcal{H}_B)}} \mathcal{Q}(\rho, \sigma)\right). \tag{5.7}$$

Note that this cone is not a completely unphysical construction, as it can be obtained by considering an ancillary system.

Remark 5.6
The cone $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$ can be understood as the set of states over time arising from an extended scenario, where an ancillary system R is added and one considers suitable CPTP maps acting on the composite system AR. More precisely, consider finite dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_R$ and the following: let

$$\rho = \sum_t p_t \rho_t \in \mathcal{S}(\mathcal{H}_A), \quad \left\{\mathcal{E}_t^{A\to B}\right\}, \tag{5.8}$$

where $\mathcal{E}_t^{A\to B}$ 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

$$\rho^{AR} = \sum_t p_t \rho_t^A \otimes |t\rangle\!\langle t|^R, \quad \mathcal{E}^{AR\to B} = \sum_t \mathcal{E}_t^{A\to B} \langle t|\cdot|t\rangle, \tag{5.9}$$

with Jamiołkowski operator $J^{AR\to B} = \sum_t J_t^{A\to B} \otimes |t\rangle\!\langle t|^R$. Now

$$J^{AR\to B} \star \rho^{AR} = \sum_t p_t J_t^{A\to B} \star \rho_t^A \otimes |t\rangle\!\langle t|^R, \tag{5.10}$$

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

$$\operatorname{Tr}_R\left[J^{AR\to B} \star \rho^{AR}\right] = \sum_t p_t J_t^{A\to B} \star \rho_t^A, \tag{5.11}$$

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

Finally, we show the cone of states over time and its dual are pointed and spanning, both desirable properties for any cone. In particular, we are interested in the spanning property of the dual cone, which ensures that $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ is not a measure zero set. Our search for this cone is justified, since we know that it is a spanning set. Recall that the dual cone is of particular interest here because it characterizes the set of cost matrices that yield a nonnegative quantum transport cost of any transport plan —not jut the optimal one.

Proposition 5.7
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 $Q \in \hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$, $\mathrm{Tr}\,[\mathbb{1}Q] = \mathrm{Tr}\,[\rho \star J] = \mathrm{Tr}\,[\rho] = 1 > 0$. By Lemma 2.13, $\mathcal{Q}(\mathcal{H}_A : \mathcal{H}_B)$ 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 [BV04]. 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. □

5.3.2 Partial characterisation of the dual cone

Our objective now is to establish Theorem 5.11, which provides a partial characterization of the dual cone $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)$, the set of cost matrices that ensure a positive quantum transport cost. Despite significant research effort, we have not been able to obtain a complete characterization of this cone, and doing so remains one of the main open questions left by this thesis.

We present four technical results that progressively explore the structure of the dual cone: we first find the dual cone to the set of matrices associated to trace scaling maps; then we show the relation between the dual cones to the Choi and Jamiołkowski matrices; then we find the dual cone to the set of Choi matrices; and finally, with these results and some technical lemmas from Section 2.1.3 we obtain Theorem 5.11.

The following technical lemma characterises the cone dual to the Choi matrices[2] associated to trace scaling maps.

Lemma 5.8
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*

$$\mathcal{C}_2^* = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A), \mathbb{1} \in \mathcal{B}(\mathcal{H}_B),\ \mathrm{Tr}\,[A] = 0\}\,. \tag{5.12}$$

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:

$$\mathrm{Tr}\,[(A \otimes \mathbb{1})\,C] = \mathrm{Tr}_A\,[\mathrm{Tr}_B\,[(A \otimes \mathbb{1})\,C]] = \mathrm{Tr}\,[A\,\mathrm{Tr}_B\,[\mathbb{1}C]] = z\,\mathrm{Tr}\,[A\mathbb{1}] = 0. \tag{5.13}$$

[2] Generally we talk about Choi matrices for CPTP maps, but any linear map has an associated Choi matrix, as it is clear from Theorem 2.28.

To see that they are equal, note that $\mathcal{C}_2$ (and thus the orthogonal $\mathcal{C}_2^*$ [BW11]) and $\mathcal{A}$ are real subspaces of $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, considered as a $2d_A^2 d_B^2$ dimensional real vector space- 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 $\mathrm{Tr}\left[A\right] = 0$. That is

$$\dim \mathcal{A} = \dim \mathcal{B}(\mathcal{H}) - 2 = 2d_A^2 - 2. \tag{5.14}$$

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 $\mathrm{Tr}_B\left[C\right] \propto_{\mathbb{C}} \mathbb{1}$. 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

$$\begin{aligned} \dim \mathcal{C}_2 &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A(d_A - 1) - 2d_A + 2 \\ &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A^2 + 2d_A - 2d_A + 2 \\ &= \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - 2d_A^2 + 2. \end{aligned} \tag{5.15}$$

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

$$\dim \mathcal{C}_2^* = \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) - \dim \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) + 2d_A^2 - 2 = 2d_A^2 - 2. \tag{5.16}$$

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:

$$\mathcal{C}_2^* = \mathcal{A} = \left\{ A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A),\ \mathrm{Tr}\left[A\right] = 0 \right\}. \tag{5.17}$$

□

The following corollary of Lemma 2.22 and Lemma 2.14 notes that the partial transpose constitutes an isomorphism between the dual cones to the Choi and Jamiołkowski operators. This makes it so we can work with either the Choi or Jamiołkowski operators interchangeably, depending on which have better properties for a particular task.

Corollary 5.9

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,

$$\mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)^* = T_A(\mathcal{C}(\mathcal{H}_\mathcal{A} \to \mathcal{H}_\mathcal{B}))^* = T_A(\mathcal{C}(\mathcal{H}_A \to \mathcal{H}_B)^*). \tag{5.18}$$

Proof. Note that the partial transpose is self adjoint from Lemma 2.22 and self inverse and apply Lemma 2.14. □

As a partial result to the characterisation of $\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^*$ we characterise the dual to the set of Choi matrices in the following proposition.

Proposition 5.10
Let $\mathcal{C} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be the minimal cone that contains the Choi matrices[3]. Then,

$$\begin{aligned}\mathcal{C}^* &= \overline{\mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B) + \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A),\ \mathrm{Tr}\,[A] = 0\}} \\ &= \overline{\{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B),\ \mathrm{Tr}\,[A] = 0\}} \\ &= \{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B),\ \mathrm{Tr}\,[A] = 0\}\,.\end{aligned} \tag{5.19}$$

Proof. Consider the following:

$$\mathcal{C}_1 = \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \tag{5.20}$$

$$\mathcal{C}_2 = \{C \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \mathrm{Tr}_B\,[C] \propto_{\mathbb{C}} \mathbb{1}\}. \tag{5.21}$$

These two are closed cones and

$$\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2. \tag{5.22}$$

Moreover (the cone of psd matrices is self dual [BW11] and Lemma 5.8):

$$\mathcal{C}_1^* = \mathcal{C}_1 = \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B), \tag{5.23}$$

$$\mathcal{C}_2^* = \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A),\ \mathrm{Tr}\,[A] = 0\}\,. \tag{5.24}$$

Now, we can use Lemma 2.12, setting $I = \{1, 2\}$ and the duals in the theorem, to find the dual of $\mathcal{C}$:

$$\begin{aligned}\mathcal{C}^* &= (\mathcal{C}_1 \cap \mathcal{C}_2)^* = (\overline{\mathcal{C}_1} \cap \overline{\mathcal{C}_2})^* = \overline{\mathcal{C}_1^* + \mathcal{C}_2^*} \\ &= \overline{\mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B) + \{A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid A \in \mathcal{B}(\mathcal{H}_A),\ \mathrm{Tr}\,[A] = 0\}},\end{aligned} \tag{5.25}$$

where we used $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$ first; the closedness of $\mathcal{C}_1$ and $\mathcal{C}_2$ second, then Lemma 2.12; and finally the duals of $\mathcal{C}_1$ and $\mathcal{C}_2$. Note that the set

$$\{\omega + A \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B) \mid \omega \in \mathcal{B}_+(\mathcal{H}_A \otimes \mathcal{H}_B),\ \mathrm{Tr}\,[A] = 0\} \tag{5.26}$$

is closed. □

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

With the previous results and the lemmas from Section 2.1.3 we can proceed to the main result of this section, as well as its proof.

Theorem 5.11
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

$$\hat{\mathcal{Q}}(\mathcal{H}_A : \mathcal{H}_B)^* = \bigcap_{U \in U(\mathcal{H}_A)} (U \otimes \mathbb{1}) \left(\bigcap_{s \in \mathbb{R}_+^{d_A}} \varphi_{D_s}^{-1}(\mathcal{J}(\mathcal{H}_A \to \mathcal{H}_B)^*) \right) (U^* \otimes \mathbb{1}) , \tag{5.27}$$

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

Proof. For simplicity, we ignore the specific Hilbert space dependencies. Start with the definition of $\hat{\mathcal{Q}}$, then apply Lemma 2.12 and Lemma 2.14:

$$\hat{\mathcal{Q}}^* = \left(\sum_\rho (\rho \star \mathcal{J}) \right)^* = \left(\sum_\rho \varphi_\rho(\mathcal{J}) \right)^* = \bigcap_\rho \varphi_\rho(\mathcal{J})^* = \bigcap_\rho \varphi_\rho^{-1}(\mathcal{J}^*). \tag{5.28}$$

Note that we can use Lemma 2.14 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 4.6. 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 $U(n)$; such that $\rho = U_\rho D_{s_\rho} U_\rho^*$. Moreover, $\mathcal{J}^*$ is invariant under local unitaries, thus

$$\begin{aligned} \varphi_\rho^{-1}(\mathcal{J}^*) &= (U_\rho \otimes \mathbb{1}) \left(U_\rho^* \rho U_\rho \star \left(\left(U_\rho^* \otimes \mathbb{1} \right) \mathcal{J}^* \left(U_\rho \otimes \mathbb{1} \right) \right)^\ominus \right)^\ominus \left(U_\rho^* \otimes \mathbb{1} \right) \\ &= (U_\rho \otimes \mathbb{1}) \left(D_{s_\rho} \star (\mathcal{J}^*)^\ominus \right)^\ominus \left(U_\rho^* \otimes \mathbb{1} \right) \\ &= (U_\rho \otimes \mathbb{1}) \, \varphi_{D_{s_\rho}}^{-1}(\mathcal{J}^*) \left(U_\rho^* \otimes \mathbb{1} \right) . \end{aligned} \tag{5.29}$$

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

$$\begin{aligned} \hat{\mathcal{Q}}^* = \bigcap_\rho \varphi_\rho^{-1}(\mathcal{J}^*) &= \bigcap_{U \in U(\mathcal{H}_A)} \bigcap_{s \in \mathbb{R}_+^{d_A}} (U \otimes \mathbb{1}) \, \varphi_{D_s}^{-1}(\mathcal{J}^*) \, (U^* \otimes \mathbb{1}) \\ &= \bigcap_{U \in U(\mathcal{H}_A)} (U \otimes \mathbb{1}) \left(\bigcap_{s \in \mathbb{R}_+^{d_A}} \varphi_{D_s}^{-1}(\mathcal{J}^*) \right) (U^* \otimes \mathbb{1}) , \end{aligned} \tag{5.30}$$

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

Theorem 5.11 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 Section 4.4.2, we have not been able to use it to obtain a complete and succinct characterization of the dual cone.

Finally, to end this section, we show how the hierarchy defined in Definition 4.9 behaves under duality.

Theorem 5.12
Let $\hat{\mathcal{Q}}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)$ be defined as in Definition 4.9. The dual of this hierarchy fulfils

$$\mathrm{Tr}_i\left[\hat{\mathcal{Q}}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)^*\right] \supseteq \mathrm{Tr}_i\left[\hat{\mathcal{Q}}(\mathcal{H}_0 : \cdots : \mathcal{H}_n)\right]^* . \tag{5.31}$$

Proof. We can show this for general cones using the proof of Lemma 2.14. Let $K \subseteq \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a cone and, then

$$\begin{aligned} x \in \mathrm{Tr}_B(K)^* \quad &\Leftrightarrow \quad \langle x, \mathrm{Tr}_B(y)\rangle \geq 0 \;\; \forall y \in K \quad \Leftrightarrow \quad \langle x \otimes \mathbb{1}, y\rangle \geq 0 \;\; \forall y \in K \\ &\Leftrightarrow \quad x \otimes \mathbb{1} \in K^* \quad \Rightarrow \quad x \in \mathrm{Tr}_B(K^*). \end{aligned} \tag{5.32}$$

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

5.3.3 Further work

A proper characterisation of the cone of states over time and its dual remains an open question of great importance to this work. After a huge effort trying to answer this question, a different direction one can take is to instead study the complexity of this cone. In quantum mechanics, there is the important example of the cone of separable states, which is of key importance to entanglement theory, that is known to be NP hard to characterise. The cone of quasistates shares some qualitative similarities with the set of separable states, which might indicate a similar complexity in the characterisation.

Let $\mathcal{H}_A, \mathcal{H}_B$ be Hilbert spaces. The set of product states on the tensor product is defined as

$$\mathrm{PROD}(A, B) = \{\rho_A \otimes \rho_B \mid \rho_A \in \mathcal{S}(\mathcal{H}_A),\ \rho_B \in \mathcal{S}(\mathcal{H}_B)\} . \tag{5.33}$$

The closure of this set is called the separable set (see the end of Section 2.2.2)

$$\mathrm{SEP}(A, B) = \mathrm{cone}\left(\mathrm{PROD}(A, B)\right) = \tag{5.34}$$

$$= \left\{ \sum_k p_k \rho_A^k \otimes \rho_B^k \mid \rho_A^k \in \mathcal{S}(\mathcal{H}_A),\ \rho_B^k \in \mathcal{S}(\mathcal{H}_B),\ p_k \geq 0\ \forall k,\ \sum_k p_k = 1 \right\}.$$

A state in the tensor product space can be trivially checked to be in $\mathrm{PROD}(A, B)$ by taking the partial traces and constructing the tensor product, then checking that the result is the same operator we started with. That is

$$\rho_{AB} \in \mathrm{PROD}(A, B) \quad \Leftrightarrow \quad \mathrm{Tr}_B\left[\rho_{AB}\right] \otimes \mathrm{Tr}_A\left[\rho_{AB}\right] = \rho_{AB}. \tag{5.35}$$

Despite this trivial characterisation of the boundary, membership to $\mathrm{SEP}(A, B)$ is a well known NP-hard problem [Gur03]. Despite this, partial solutions to the problem exist [Per96; HHH96], with characterisation provided in low dimensions in [HHH96].

This analogy motivates the study of the cone of states over time: just as the separable set is computationally hard to characterize despite its simple building blocks, the cone of states over time may similarly exhibit a rich and complex structure. Understanding its complexity, even partially, can therefore provide valuable insights and guide the development of approximate or tractable methods for working with states over time in quantum information tasks.

5.4 Zero cost for the identity channel

Another property of interest is that the cost should vanish if and only if $\rho = \sigma$. Furthermore, we require that the channel achieving this zero cost is the identity channel. The motivation is straightforward: an operation that leaves the state unchanged should incur no cost. This property is neatly characterised in the following theorem.

Theorem 5.13
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

$$\mathrm{Tr}_B\left[\mathcal{S} \star K\right] = 0, \tag{5.36}$$

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

Proof. The Jamiołkowski operator associated to the identity channel is the swap operator $\mathcal{S}$, clearly from Theorem 2.29: $J_{\mathrm{id}} = (\mathrm{id} \otimes \mathrm{id})(\mathcal{S}) = \mathcal{S}$. Now, let K be a matrix such that the cost $\kappa(\rho, \mathrm{id}) = \mathrm{Tr}\left[(\rho \star \mathcal{S})K\right] = 0\ \forall \rho \geq 0$. 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:

$$\begin{aligned}\operatorname{Tr}\left[((\rho\otimes\mathbb{1})\star\mathcal{S})K\right] &= \frac{1}{2}\operatorname{Tr}\left[((\rho\otimes\mathbb{1})\mathcal{S}+\mathcal{S}(\rho\otimes\mathbb{1}))K\right] \\ &= \frac{1}{2}\operatorname{Tr}\left[(\rho\otimes\mathbb{1})(\mathcal{S}K+K\mathcal{S})\right] \\ &= \operatorname{Tr}\left[\rho\operatorname{Tr}_B\left[\mathcal{S}\star K\right]\right]=0.\end{aligned} \tag{5.37}$$

The set of positive matrices generates the whole space [SSŻ09; BW11], therefore this is equivalent to saying that the Hilbert-Schmidt inner product of $\operatorname{Tr}_B\left[\mathcal{S}\star K\right]$ with all other elements is 0. Therefore $\operatorname{Tr}_B\left[\mathcal{S}\star K\right]=0$. The converse is immediate, completing the proof. □

5.4.1 Combination of positive and cost zero cost matrices

We can assume both positivity of the cost and 0 cost for the identity channel to obtain the following results in the case where $\mathcal{H}_A=\mathcal{H}_B$:

Theorem 5.14
Let $\mathcal{H}$ be a finite dimensional Hilbert space. Then

$$K\in\mathcal{J}(\mathcal{H}\to\mathcal{H})^*\cap\left\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[\mathcal{S}\star C\right]=0\right\} \tag{5.38}$$

if and only if

$$K=T_A\left(\omega\right)-\left(\operatorname{Tr}_B\left[\mathcal{S}\star T_A\left(\omega\right)\right]\otimes\mathbb{1}\right),\quad\omega\geq0,\quad\omega\perp|\Phi_+\rangle\langle\Phi_+|\,. \tag{5.39}$$

Proof. Similarly to before, we ignore the Hilbert space dependencies for the proof. Note that, in general, 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 not true. However, it holds for the partial transpose since this map is invertible. Thus, we can transform the target set as follows:

$$\begin{aligned}&\mathcal{J}^*\cap\left\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[\mathcal{S}\star C\right]=0\right\}\\ &\quad=T_A(T_A(\mathcal{J}^*\cap\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[\mathcal{S}\star C\right]=0\}))\\ &\quad=T_A(T_A(\mathcal{J}^*)\cap T_A(\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[\mathcal{S}\star C\right]=0\}))\\ &\quad=T_A(\mathcal{C}^*\cap\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[\mathcal{S}\star T_A(C)\right]=0\})\\ &\quad=T_A(\mathcal{C}^*\cap\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid T_A(\operatorname{Tr}_B\left[\mathcal{S}\star T_A(C)\right])=T_A(0)\})\\ &\quad=T_A(\mathcal{C}^*\cap\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[T_A(\mathcal{S})\star C\right]=0\})\\ &\quad=T_A(\mathcal{C}^*\cap\{C\in\mathcal{B}(\mathcal{H}\otimes\mathcal{H})\mid\operatorname{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+|\star C\right]=0\}),\end{aligned} \tag{5.40}$$

where we used Lemma 2.22.

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

$$\begin{aligned}0 &= \mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star K\right] = \mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star (\omega + A \otimes \mathbb{1})\right] \\ &= \mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star \omega\right] + A,\end{aligned} \tag{5.41}$$

thus $A = -\mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star \omega\right]$. Moreover if we take the trace of this expression, since $\mathrm{Tr}\left[A\right] = 0$, 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

$$\begin{aligned}&\mathcal{J}^* \cap \{C \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \mid \mathrm{Tr}_B\left[\mathcal{S} \star C\right] = 0\} \\ &\quad = T_A(\{\omega - \mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star \omega\right] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0,\ \omega \perp |\Phi_+\rangle\langle\Phi_+|\}) \\ &\quad = \{T_A(\omega) - T_A(\mathrm{Tr}_B\left[|\Phi_+\rangle\langle\Phi_+| \star \omega\right] \otimes \mathbb{1}) \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0,\ \omega \perp |\Phi_+\rangle\langle\Phi_+|\} \\ &\quad = \{T_A(\omega) - \mathrm{Tr}_B\left[T_A(|\Phi_+\rangle\langle\Phi_+|) \star T_A(\omega)\right] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0,\ \omega \perp |\Phi_+\rangle\langle\Phi_+|\} \\ &\quad = \{T_A(\omega) - \mathrm{Tr}_B\left[\mathcal{S} \star T_A(\omega)\right] \otimes \mathbb{1} \in \mathcal{B}(\mathcal{H}^2) \mid \omega \geq 0,\ \omega \perp |\Phi_+\rangle\langle\Phi_+|\}.\end{aligned} \tag{5.42}$$

□

In the Theorem, K is a matrix that is dual to the Jamiołkowski matrices and generates cost 0 for the identity (see Theorem 5.13, Proposition 5.10 and Lemma 2.22). This yields a set that contains the intersection between the dual $\mathcal{Q}^*$ and the cost matrices that yield zero for the identity channel, thus constituting a good set to look for operators that fulfil both properties.

5.5 Symmetry

Finally we address the symmetry of the optimal cost, that is $\mathcal{K}(\rho, \sigma) = \mathcal{K}(\sigma, \rho)$ for all states ρ and σ.

A naive guess or first thought might suggest that symmetry of the cost matrix under input–output exchange, i.e., $\mathcal{S}K\mathcal{S} = K$, 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 $\mathcal{S}\mathcal{Q}(\rho, \sigma)\mathcal{S} = \mathcal{Q}(\sigma, \rho)$.

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 4.6, 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 $\mathcal{S}K\mathcal{S} = K$ is not a sufficient condition on the cost matrix to give rise to symmetry of the cost.

Example 5.15
Within this example we indicate the direction of time we are observing with the subindices $A \to B$ and $A \leftarrow B$. The order of the subsystems in the matrix notations will always be $\mathcal{H}_A \otimes \mathcal{H}_B$. ρ will be the state associated to subsystem A and σ the state associated to subsystem B.

i) ***Replacement channel:*** *Let ρ, σ be any states and let $J_{A\to B}$ be the Jamiołkowski matrix associated to the constant channel $\mathcal{E}(\rho) = \mathrm{Tr}(\rho)\sigma$, that is $J_{A\to B} = \mathbb{1} \otimes \sigma$. The associated state over time is $Q = \rho \star J_{A\to B} = \rho \otimes \sigma$. 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 $J_{A\to B}$ 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,*

$$Q = \rho \star \mathcal{S} = \begin{bmatrix} p & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1-p \end{bmatrix}. \tag{5.43}$$

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, WLOG, we will set $\rho = |0\rangle\langle 0|$. Let $J_{A\to B}$ be the Jamiołkowski matrix associated to the depolarising channel $\mathcal{E}(\rho) = (1-p)\rho + p\,\mathrm{Tr}(\rho)\frac{\mathbb{1}}{2}$, that is $J_{A\to B} = (1-p)\mathcal{S} + \frac{p}{2}\mathbb{1}$. The resulting state over time is*

$$Q = \rho \star J_{A\to B} = \frac{1}{2}\begin{bmatrix} 2-p & 0 & 0 & 0 \\ 0 & p & 1-p & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.44}$$

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 4.6 to Q yields

$$J_{A\leftarrow B}^{T_B} = \begin{bmatrix} 1 & 0 & 0 & 1-p \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1-p & 0 & 0 & 0 \end{bmatrix}, \tag{5.45}$$

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 $J_{A\to B}$ 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*

$$J_{A\to B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{2p-1}{2} & 0 \\ 0 & \frac{2p-1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.46}$$

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

$$Q = \rho \star J_{A\to B} = \frac{1}{4}\begin{bmatrix} 2 & 2p-1 & 1 & 0 \\ 2p-1 & 0 & 2p-1 & 1 \\ 1 & 2p-1 & 0 & 2p-1 \\ 0 & 1 & 2p-1 & 2 \end{bmatrix}. \tag{5.47}$$

We can now calculate $J_{A\leftarrow B}$ from Theorem 4.6[4]*, which yields*

$$J_{A\leftarrow B} = \frac{1}{2}\begin{bmatrix} 2 & 0 & 1 & 1-2p \\ 0 & 0 & 2p-1 & 1 \\ 1 & 2p-1 & 0 & 0 \\ 1-2p & 1 & 0 & 2 \end{bmatrix}. \tag{5.48}$$

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 $\{\frac{1}{2}(1 \pm \sqrt{2})\}$.

[4]Note that even though ρ has rank 1, σ has rank 2 and therefore allows us to uniquely apply Theorem 4.6.

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

$$\varepsilon(x) = \sum_i \operatorname{Tr}[M_i x]\,\sigma_i, \tag{5.49}$$

where $\{M_i\}$ *is a POVM and* $\sigma_i \in \mathcal{S}(\mathcal{H}_B)$ *are states. Then,*

$$J_{A\to B} = \sum M_i \otimes \rho_i \quad \text{and} \quad Q = \sum (\rho \star M_i) \otimes \sigma_i. \tag{5.50}$$

Because the map in Theorem 4.6 is CP, as seen in Corollary 4.7, $J^{T_A}_{A\leftarrow B}$ *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 in 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* $J_{A\to B} = |0,0\rangle\langle 0,0| + |1,+\rangle\langle 1,+|$*, 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*

$$Q = \rho \star J_{A\to B} = \frac{1}{2}\begin{bmatrix} 2p & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & 1-p \\ 0 & 0 & 1-p & 1-p \end{bmatrix}. \tag{5.51}$$

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

$$J_{A\leftarrow B} = \frac{1}{2}\begin{bmatrix} 1+2p & 0 & 2p-1 & 0 \\ 0 & 1-2p & 0 & 1-2p \\ 2p-1 & 0 & 1-2p & 0 \\ 0 & 1-2p & 0 & 1+2p \end{bmatrix}, \tag{5.52}$$

which is positive under partial transposition.

This example does not show that the cost is not symmetric under symmetric cost matrices, since it could be that the two different optimal channels happen to yield the same cost. We show in Section 6.5 that this is not the case in the particular of the unitary invariant cost matrix, which provides a concrete counterexample for the general case. This demonstrates that a symmetric cost matrix does not in general imply a symmetric quantum transport cost.

5.6 Triangle inequality

In this section we consider the triangle inequality for the quantum optimal transport cost $\mathcal{K}(\rho,\sigma)$. In Theorem 5.16 we state a sufficient condition for its fulfilment. Since this condition relies on $\mathcal{Q}(\mathcal{H}_A:\mathcal{H}_B:\mathcal{H}_C)^*$, the characterisation of the dual space explored in Section 5.3 is also relevant in this section.

Theorem 5.16
Consider Hilbert spaces $\mathcal{H}_A$, $\mathcal{H}_B$ and $\mathcal{H}_C$, and states ρ,σ,μ in $\mathcal{S}(\mathcal{H}_A),\mathcal{S}(\mathcal{H}_B),\mathcal{S}(\mathcal{H}_C)$, respectively. The inequality

$$\mathcal{K}_{AB}(\rho,\sigma)+\mathcal{K}_{BC}(\sigma,\mu)\geq\mathcal{K}_{AC}(\rho,\mu) \tag{5.53}$$

will be fulfilled for all input states if the cost matrices fulfil the following identity:

$$K_{AB}\otimes\mathbb{1}_C+\mathbb{1}_A\otimes K_{BC}-K_{AC}\otimes\mathbb{1}_B\in\mathcal{Q}(\mathcal{H}_A:\mathcal{H}_B:\mathcal{H}_C)^*. \tag{5.54}$$

Proof. Consider first an admissible Jamiołkowski matrix in systems AB and a cost matrix K_{AB}. Because $\mathbb{1}_B=\mathrm{Tr}_C\left[J_{BC}\right]$ for any admissible Jamiołkowski matrix in systems BC, and the partial associativity of the Jordan product [HHP+17; FJV15] we can rewrite the cost as

$$\begin{aligned}\mathrm{Tr}\left[K_{AB}(\rho\star J_{AB})\right]&=\mathrm{Tr}\left[K_{AB}((\rho\star J_{AB})\star(\mathbb{1}_A\otimes\mathrm{Tr}_C\left[J_{BC}\right]))\right]\\&=\mathrm{Tr}\left[(K_{AB}\star(\rho\star J_{AB}))(\mathbb{1}_A\otimes\mathrm{Tr}_C\left[J_{BC}\right])\right]\\&=\mathrm{Tr}\left[((K_{AB}\star(\rho\star J_{AB}))\otimes\mathbb{1}_C)(\mathbb{1}_A\otimes J_{BC})\right]\\&=\mathrm{Tr}\left[(K_{AB}\otimes\mathbb{1}_C)((\rho\star(J_{AB}\otimes\mathbb{1}_C))\star(\mathbb{1}_A\otimes J_{BC}))\right]\\&=\mathrm{Tr}\left[(K_{AB}\otimes\mathbb{1}_C)\left(\rho\star((J_{AB}\otimes\mathbb{1}_C)\star(\mathbb{1}_A\otimes J_{BC}))\right)\right].\end{aligned} \tag{5.55}$$

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

$$\begin{aligned}\mathrm{Tr}\left[K_{BC}(\sigma\star J_{BC})\right]&=\mathrm{Tr}\left[K_{BC}((\mathrm{Tr}_A\left[\rho\star J_{AB}\right]\otimes\mathbb{1}_C)\star J_{BC})\right]\\&=\mathrm{Tr}\left[(J_{BC}\star K_{BC})(\mathrm{Tr}_A\left[\rho\star J_{AB}\right]\otimes\mathbb{1}_C)\right]\\&=\mathrm{Tr}\left[((\mathbb{1}_A\otimes J_{BC})\star(\mathbb{1}_A\otimes K_{BC}))(\rho\star J_{AB}\otimes\mathbb{1}_C)\right]\\&=\mathrm{Tr}\left[(\mathbb{1}_A\otimes K_{BC})((\rho\star(J_{AB}\otimes\mathbb{1}_C))\star(\mathbb{1}_A\otimes J_{BC}))\right]\\&=\mathrm{Tr}\left[(\mathbb{1}_A\otimes K_{BC})(\rho\star((J_{AB}\otimes\mathbb{1}_C)\star(\mathbb{1}_A\otimes J_{BC})))\right].\end{aligned} \tag{5.56}$$

Finally, for systems AC, consider the link product [CDP09] of any two Jamiołkowski matrices as the Jamiołkowski matrix of AC:

$$\begin{aligned} \mathrm{Tr}\left[K_{AC}(\rho \star J_{AC})\right] &= \mathrm{Tr}\left[K_{AC}(\rho \star \mathrm{Tr}_B\left[(J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})\right])\right] \\ &= \mathrm{Tr}\left[(K_{AC} \star \rho)\, \mathrm{Tr}_B\left[(J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})\right]\right] \\ &= \mathrm{Tr}\left[((K_{AC} \otimes \mathbb{1}_B) \star \rho)((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC}))\right] \\ &= \mathrm{Tr}\left[(K_{AC} \otimes \mathbb{1}_B)(\rho \star ((J_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J_{BC})))\right]. \end{aligned} \tag{5.57}$$

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:

$$\begin{aligned} \mathcal{K}_{AC} &= \mathrm{Tr}\left[K_{AC}(\rho \star J^o_{AC})\right] \leq \mathrm{Tr}\left[K_{AC}(\rho \star J_{AC})\right] \\ &= \mathrm{Tr}\left[K_{AC} \otimes \mathbb{1}_B(\rho \star ((J^o_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J^o_{BC})))\right] \\ &= \mathrm{Tr}\left[(K_{AB} \otimes \mathbb{1}_C + \mathbb{1}_A \otimes K_{BC} - K')(\rho \star ((J^o_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J^o_{BC})))\right] \\ &= \mathrm{Tr}\left[(\mathbb{1}_A \otimes K_{BC})(\rho \star ((J^o_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J^o_{BC})))\right] \\ &+ \mathrm{Tr}\left[(K_{AB} \otimes \mathbb{1}_C)\,(\rho \star ((J^o_{AB} \otimes \mathbb{1}_C) \star (\mathbb{1}_A \otimes J^o_{BC})))\right] \\ &- \mathrm{Tr}\left[K'\,(\rho \star (J^o_{AB} \star J^o_{BC}))\right] \\ &= \mathrm{Tr}\left[K_{AB}(\rho \star J^o_{AB})\right] + \mathrm{Tr}\left[K_{BC}(\sigma \star J^o_{BC})\right] - \mathrm{Tr}\left[K'\,(\rho \star (J^o_{AB} \star J^o_{BC}))\right] \\ &= \mathcal{K}_{AB} + \mathcal{K}_{BC} - \mathrm{Tr}\left[K'\,(\rho \star (J^o_{AB} \star J^o_{BC}))\right]. \end{aligned} \tag{5.58}$$

Finally, because K' is in the dual of $\mathcal{Q}_3$, $\mathrm{Tr}\left[K'\,(\rho \star (J^o_{AB} \star J^o_{BC}))\right] \geq 0$ and therefore $\mathcal{K}_{AB} + \mathcal{K}_{BC} \geq \mathcal{K}_{AC}$. □

The proof also provides some insight in what a necessary and sufficient condition would require. If we fix the cost matrices K_{AB}, K_{BC} and K_{AC}, and K' as defined in the proof, we require that the value $\mathrm{Tr}\left[K'\,(\rho \star (J^o_{AB} \star J^o_{BC}))\right]$ is positive for all triples $(\rho, J^o_{AB}, J^o_{BC})$, with J^o_{AB}, J^o_{BC} fixed by the σ, μ and the cost matrices. Finding this space analytically seems reliant on knowing the space of optimal maps, which is in general very hard due to the need for convex optimisation to approximate each one for general cost matrices.

Finally, we state Theorem 5.16 for the case where $\mathcal{H}_A = \mathcal{H}_B = \mathcal{H}_C = \mathcal{H}$ and all cost matrices are the same, since this is when the triangle inequality is generally relevant, namely for distinguishability measures on elements of the same space.

Corollary 5.17
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

$$K \otimes \mathbb{1}_3 + \mathbb{1}_1 \otimes K - K \otimes \mathbb{1}_2 \in \mathcal{Q}(\mathcal{H} : \mathcal{H} : \mathcal{H})^*, \tag{5.59}$$

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

5.7 General properties

Finally, we show some general mathematical properties of the optimal quantum transport. Proposition 5.18 shows subadditivity related properties and Proposition 5.20 shows how we can relate the transport cost on the tensor product to the transport cost on each subsystem.

Proposition 5.18
Let p_x be a probability distribution. The optimal quantum cost fulfils the following:

i) Subadditivity: $\mathcal{K}\left(\rho, \sum_x p_x\sigma_x\right) \leq \sum_x p_x \mathcal{K}\left(\rho, \sigma_x\right)$.

Moreover, if the triangle inequality is fulfilled:

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

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

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 fulfills $\mathcal{E}_{J_\Sigma}(\rho) = \sum_x p_x\sigma_x$. Thus,

$$\begin{aligned} \mathcal{K}\left(\rho, \sum_x p_x\sigma_x\right) &= \mathrm{Tr}\left[K\left(\rho \star J_o\right)\right] \leq \mathrm{Tr}\left[K\left(\rho \star J_\Sigma\right)\right] \\ &= \mathrm{Tr}\left[K\left(\rho \star \left(\sum_x p_x J_x\right)\right)\right] = \sum_x p_x \,\mathrm{Tr}\left[K\left(\rho \star J_x\right)\right] \\ &= \sum_x p_x \mathcal{K}(\rho, \sigma_x), \end{aligned} \tag{5.60}$$

where we used the bilinearity of the Jordan product [HHP+17] and the linearity of the trace.

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

$$\begin{aligned}\sum_x p_x \mathcal{K}(\rho_x, \sigma) &\leq \sum_x p_x \left(\mathcal{K}\left(\rho_x, \sum_{x'} p_{x'}\rho_{x'}\right) + \mathcal{K}\left(\sum_{x'} p_{x'}\rho_{x'}, \sigma\right)\right) \\ &= \mathcal{K}\left(\sum_x p_x\rho_x, \sigma\right) + \sum_x p_x \mathcal{K}\left(\rho_x, \sum_{x'} p_{x'}\rho_{x'}\right).\end{aligned} \tag{5.61}$$

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

$$\begin{aligned} p_x\mathcal{K}(\rho,\sigma) &\leq p_x\mathcal{K}(\rho,\rho_x) + p_x\mathcal{K}(\rho_x,\sigma) \\ \Rightarrow \quad \sum_x p_x\mathcal{K}(\rho,\sigma) &\leq \sum_x p_x\mathcal{K}(\rho,\rho_x) + \sum_x p_x\mathcal{K}(\rho_x,\sigma) \\ \Rightarrow \quad \mathcal{K}(\rho,\sigma) &\leq \sum_x p_x\mathcal{K}(\rho,\rho_x) + \sum_x p_x\mathcal{K}(\rho_x,\sigma),\end{aligned} \tag{5.62}$$

where we first used the triangle inequality and then we added all the inequalities together. □

Remark 5.19
A similar proof does not work for subadditivity 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

$$\mathcal{K}\left(\sum_x p_x\rho_x, \sigma\right) = \operatorname{Tr}\left[K\left(\sum_x p_x\rho_x \star J_o\right)\right] = \sum_x p_x \operatorname{Tr}\left[K\left(\rho_x \star J_o\right)\right]. \tag{5.63}$$

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 5.18 it was possible to define the joint channel J_Σ because we could send ρ to each element of the ensemble $\{(p_x, \sigma_x)\}$ and that would in total define a channel that sends ρ to σ.

We can define $\sigma_x = \mathcal{E}_{J_o}(\rho_x)$ and observe that $\sigma = \sum_x p_x\sigma_x$ to lower bound this quantity obtaining

$$\mathcal{K}\left(\sum_x p_x\rho_x, \sum_x p_x\sigma_x\right) \geq \sum_x p_x\mathcal{K}(\rho_x, \sigma_x). \tag{5.64}$$

This joint superadditivity 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.

This last expression allows us to prove that subadditivity on the first input is false in general. Let $\mathcal{H} = \mathbb{C}^2$ and let K be an associated cost matrix that yields a positive optimal transport cost that is 0 for the identity channel. Now consider $\rho = \sigma = \mathbb{1}_2$ and the ensambles $\{\left(\frac{1}{2}, |0\rangle\langle 0|\right), \left(\frac{1}{2}, |1\rangle\langle 1|\right)\}$ and $\{\left(\frac{1}{2}, |+\rangle\langle +|\right), \left(\frac{1}{2}, |-\rangle\langle -|\right)\}$.

Proposition 5.20
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 associated to $\mathcal{H}_1 \otimes \mathcal{H}_2$, and K_i be cost matrices associated to $\mathcal{H}_i$. Then the optimal transport cost of $\mathcal{K}\left(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2\right)$ 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,

$$\begin{aligned}
\mathcal{K}(\rho_1, \sigma_1) \cdot \mathcal{K}(\rho_2, \sigma_2) &= \mathrm{Tr}\left[K_1\left(\rho_1 \star J_1^o\right)\right] \cdot \mathrm{Tr}\left[K_2\left(\rho_2 \star J_2^o\right)\right] \\
&= \mathrm{Tr}\left[\left(K_1\left(\rho_1 \star J_1^o\right)\right) \otimes \left(K_2\left(\rho_2 \star J_2^o\right)\right)\right] \\
&= \mathrm{Tr}\left[\left(K_1 \otimes K_2\right)\left(\left(\rho_1 \otimes \rho_2\right) \star \left(J_1^o \otimes J_2^o\right)\right)\right] \\
&\geq \mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2).
\end{aligned} \tag{5.65}$$

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

$$\begin{aligned}
\mathcal{K}(\rho_1 \otimes \rho_2, \sigma_1 \otimes \sigma_2) &= \mathrm{Tr}\left[K_{12}\left(\rho_1 \otimes \rho_2\right) \star J_{12}^o\right] \\
&\leq \mathrm{Tr}\left[\left(K_1 \otimes \mathbb{1}_2 + \mathbb{1}_1 \otimes K_2\right)\left(\rho_1 \otimes \rho_2\right) \star \left(J_1^o \otimes J_2^o\right)\right] \\
&= \mathrm{Tr}\left[K_1 \rho_1 \star J_1^o\right] \mathrm{Tr}\left[\rho_2 \star J_2^o\right] \\
&\quad + \mathrm{Tr}\left[\rho_1 \star J_1^o\right] \mathrm{Tr}\left[K_2 \rho_2 \star J_2^o\right] \\
&= \mathcal{K}(\rho_1, \sigma_1) + \mathcal{K}(\rho_2, \sigma_2).
\end{aligned} \tag{5.66}$$

□

6

The unitary invariant quantum transport cost

In this chapter, 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:

$$\kappa(\rho, \mathcal{E}) = \kappa(U\rho U^*, \mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^{-1}) \quad \forall U \in U(d),\ \rho, \sigma \in \mathcal{S}(\mathcal{H}). \tag{6.1}$$

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 diagram:

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

In Section 6.1, we show that unitary invariance fully determines the cost matrix up to a single parameter. In Section 6.2, 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 Section 6.3, we obtain analytic expressions for the optimal cost in the special case of commuting states. Section 6.4 explores the striking modification of the cost induced by embedding the problem into a higher-dimensional Hilbert space, and study its asymptotic behaviour. In Section 6.5, we present an example that illustrates several noteworthy features. Finally, in Section 6.6, we introduce the notion of a Hamiltonian-based quantum cost, which also exhibits unitary invariance, though restricted to the subgroup of unitaries generated by a fixed Hamiltonian.

6.1 Unitary invariant cost matrix

We can immediately see that there is single cost matrix that generates a unitary invariant cost with good properties:

Theorem 6.1
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 Theorem 5.13 and are unitarily invariant as defined in Eq. (6.1), are positive multiples of*

$$K_0 = d\mathbb{1} - \mathcal{S}. \tag{6.2}$$

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. (2.54)

$$\begin{aligned} J'_U &= (\mathrm{id}\otimes\mathcal{E}'_U)\,(\mathcal{S}) = (\mathrm{id}\otimes\mathcal{U})\circ(\mathrm{id}\otimes\mathcal{E})\circ(\mathrm{id}\otimes\mathcal{U}^{-1})(\mathcal{S}) \\ &= (\mathbb{1}\otimes U)\,(\mathrm{id}\otimes\mathcal{E})((\mathbb{1}\otimes U^*)\mathcal{S}(\mathbb{1}\otimes U))\,(\mathbb{1}\otimes U)^* \end{aligned} \tag{6.3}$$

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.

$$\begin{aligned} J'_U &= (\mathbb{1}\otimes U)(U\otimes\mathbb{1})\,(\mathrm{id}\otimes\mathcal{E})(\mathcal{S})\,(U^*\otimes\mathbb{1})(\mathbb{1}\otimes U^*) \\ &= (U\otimes U)\,J\,(U^*\otimes U^*). \end{aligned} \tag{6.4}$$

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

$$\begin{aligned} \mathrm{Tr}\left[K_0\,(\rho\star J)\right] &= \mathrm{Tr}\left[K_0\,((U\rho U^*)\star((U\otimes U)J(U^*\otimes U^*)))\right] \\ &= \mathrm{Tr}\left[(U\otimes U)K_0(U^*\otimes U^*)(\rho\star J)\right]. \end{aligned} \tag{6.5}$$

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 5.7), the scalar equality Eq. (6.5) 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 whole 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 [Wer89]. Therefore $K_0 = a(b\mathbb{1} - \mathcal{S})$ with real a and b, to preserve Hermiticity.

We can now impose the second condition, Theorem 5.13:

$$0 = \mathrm{Tr}_B [\mathcal{S} \star K_0] = a\,\mathrm{Tr}_B [\mathcal{S} \star (b\mathbb{1} - \mathcal{S})] = a\,\mathrm{Tr}_B [b\mathcal{S} - \mathbb{1}] = a(b-d)\mathbb{1}. \tag{6.6}$$

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

$$\begin{aligned} \mathrm{Tr}[K_0(\rho \star J)] &= a\,\mathrm{Tr}[(d\mathbb{1} - \mathcal{S})(\rho \star J)] \\ &= ad\,\mathrm{Tr}[\rho \star J] - a\,\mathrm{Tr}[\mathcal{S}(\rho \star J)] \\ &= a\left[d - \mathrm{Tr}[\rho(\mathcal{S} \star J)]\right]. \end{aligned} \tag{6.7}$$

We can now bound the remaining term using the operator norm:

$$\begin{aligned} \mathrm{Tr}[\rho(\mathcal{S} \star J)] &= \sum_i p_i \langle i | J \star \mathcal{S} | i \rangle \leq \sum_i p_i \| J \star \mathcal{S} \, | i \rangle \| = \sum_i p_i \| J \star S \| \| | i \rangle \| \\ &= \left(\sum_i p_i \right) \| J \star \mathcal{S} \| \leq \| J \| \| \mathcal{S} \| \leq d. \end{aligned} \tag{6.8}$$

Thus we have that a times a positive constant has to be positive, therefore a is positive. □

Throughout this section we refer to K_0 as the UI cost matrix and $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$ the normalised UI cost matrix. Since they mostly share the same properties, we generally use them interchangeably depending on the situation. With $\tilde{K}_0$ the maximum achievable cost is 1, which allows us to compare costs across dimensions.

With these cost matrices, we can write the cost in different ways depending on the choice of channel representation (see Section 2.3). These forms will be useful later.

Remark 6.2
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

$$\kappa(\rho, \mathcal{E}) = \mathrm{Tr}\left[\tilde{K}_0(\rho \star J)\right] = 1 - \frac{1}{d}\langle \Phi_+|\rho^T \star C|\Phi_+\rangle \tag{6.9}$$

$$= 1 - \frac{1}{d}\mathfrak{Re}(\langle\rho| C |\Phi_+\rangle) \tag{6.10}$$

$$= 1 - \frac{1}{d}\sum_{ij} \frac{p_i + p_j}{2} \langle i| \mathcal{E}(|i\rangle\langle j|) |j\rangle \tag{6.11}$$

$$= 1 - \frac{1}{d}\sum_{i} p_i \sum_{j} \mathfrak{Re}(\langle i| \mathcal{E}(|i\rangle\langle j|) |j\rangle) \tag{6.12}$$

$$= 1 - \frac{1}{d}\sum_{k} \mathfrak{Re}\left(\mathrm{Tr}\left[E_k^*\right] \mathrm{Tr}\left[E_k\rho\right]\right). \tag{6.13}$$

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 $\mathrm{Tr}\left[\mathcal{S}(\rho\otimes J)\right]$, and we will focus on that.

Eq. (6.9) and Eq. (6.10) are a direct consequence of Lemma 2.22, recalling that $\mathcal{S}^{T_A} = |\Phi_+\rangle\langle\Phi_+|$.

Eq. (6.11) comes from the definition of the Jamiołkowski matrix, $J = \sum_{ij} |i\rangle\langle j| \otimes \mathcal{E}(|j\rangle\langle i|)$ and Theorem 4.6, 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 $\mathrm{Tr}\left[\mathcal{S}(\rho\star J)\right]$:

$$\mathrm{Tr}\left[\mathcal{S}(\rho \star J)\right] = \mathrm{Tr}\left[\sum_{ij} \frac{p_i + p_j}{2} \sum_{i'j'} \left(|i'\rangle\langle j'| \otimes |j'\rangle\langle i'|\right)\left(|i\rangle\langle j| \otimes \mathcal{E}(|j\rangle\langle i|)\right)\right]$$
$$= \sum_{ij} \frac{p_i + p_j}{2} \sum_{i'j'} \delta_{ij'}\delta_{ji'} \langle i'| \mathcal{E}(|j\rangle\langle i|) |j'\rangle \tag{6.14}$$
$$= \sum_{ij} \frac{p_i + p_j}{2} \langle i| \mathcal{E}(|i\rangle\langle j|) |j\rangle .$$

Finally, for Eq. (6.13) consider J written as a function of the Kraus operators:

$$J = (\mathrm{id} \otimes \mathcal{E})(\mathcal{S}) = \sum_k (\mathbb{1} \otimes E_k)\mathcal{S}(\mathbb{1} \otimes E_k^*) = \sum_k (E_k^* \otimes E_k)\mathcal{S}. \tag{6.15}$$

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

$$\begin{aligned}
\operatorname{Tr}[\mathcal{S}(\rho \star J)] &= \sum_k \operatorname{Tr}[\mathcal{S}(\rho \star (E_k^* \otimes E_k)\mathcal{S})] \\
&= \frac{1}{2}\sum_k \left(\operatorname{Tr}[\mathcal{S}(\rho(E_k^* \otimes E_k)\mathcal{S})] + \operatorname{Tr}[\mathcal{S}((E_k^* \otimes E_k)\mathcal{S}\rho)]\right) \\
&= \frac{1}{2}\sum_k \left(\operatorname{Tr}[\rho E_k^* \otimes E_k] + \operatorname{Tr}[E_k\rho \otimes E_k^*]\right) \qquad (6.16)\\
&= \frac{1}{2}\sum_k \left(\operatorname{Tr}[\rho E_k^*]\operatorname{Tr}[E_k] + h.c.\right) \\
&= \sum_k \mathfrak{Re}\left(\operatorname{Tr}[E_k^*]\operatorname{Tr}[\rho E_k]\right).
\end{aligned}$$

This concludes the proof. □

A channel can have more than one Kraus representation [NC10, Theorem 8.2], also shown in Section 2.3. The following remark shows that Eq. (6.13) holds for all of them, as it should.

Remark 6.3
Note that by Eq. (6.13), for every Kraus representation of a channel

$$1 - \frac{1}{d}\sum_k \mathfrak{Re}\left(\operatorname{Tr}[E_k^*]\operatorname{Tr}[\rho E_k]\right) = \operatorname{Tr}[S(\rho \star J)], \qquad (6.17)$$

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

$E_i = \sum_j u_{ij} F_j$ *[NC10]. Then*

$$\begin{aligned}\sum_i \mathfrak{Re}\left(\operatorname{Tr}\left[E_i^*\right]\operatorname{Tr}\left[\rho E_i\right]\right) &= \sum_i \mathfrak{Re}\left(\operatorname{Tr}\left[\sum_j \bar{U}_{ij} F_j^*\right]\operatorname{Tr}\left[\rho \sum_{j'} U_{ij'} F_{j'}\right]\right)\\ &= \sum_{jj'}\left(\sum_i U_{ij'}\bar{U}_{ij}\right)\mathfrak{Re}\left(\operatorname{Tr}\left[F_j^*\right]\operatorname{Tr}\left[\rho F_{j'}\right]\right)\\ &= \sum_{jj'}\left(U^T\left(U^T\right)^*\right)_{j'j}\mathfrak{Re}\left(\operatorname{Tr}\left[F_j^*\right]\operatorname{Tr}\left[\rho F_{j'}\right]\right)\\ &= \sum_{jj'}\delta_{jj'}\mathfrak{Re}\left(\operatorname{Tr}\left[F_j^*\right]\operatorname{Tr}\left[\rho F_{j'}\right]\right)\\ &= \sum_j \mathfrak{Re}\left(\operatorname{Tr}\left[F_j^*\right]\operatorname{Tr}\left[\rho F_j\right]\right).\end{aligned} \tag{6.18}$$

6.2 Analytical analysis of the UI cost

In this section we study the analytical solution of the unitary invariant cost. While no analytical solution exists in general, we show that the solution depends only on the joint support of ρ and σ and we analytically solve the particular case when ρ and σ are related by a unitary channel.

First, we calculate the cost associated to two important quantum channels.

Example 6.4

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

$$\kappa(\rho, \mathcal{E}_R) = \operatorname{Tr}\left[(d\mathbb{1} - \mathcal{S})\rho_A \star \sigma_B\right] = d - \operatorname{Tr}\left[\rho\sigma\right] \geq d - 1, \tag{6.19}$$

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

2) ***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. (6.13) in Remark 6.2 to calculate the cost, since the Kraus operators associated to a Unitary channel are just* $\{U\}$*. This cost is then*

$$\kappa(\rho, \mathcal{E}_U) = \operatorname{Tr}\left[(\rho_A \star (U_B\mathcal{S}U_B^*))(d\mathbb{1} - \mathcal{S})\right] = d - \mathfrak{Re}\left(\operatorname{Tr}\left[U^*\right]\operatorname{Tr}\left[\rho U\right]\right) \tag{6.20}$$

We can find an expression for the optimal U when ρ and σ are pure states. Without loss of generality due to the unitary invariance, consider $\rho = |0\rangle\langle 0|$ 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 $\langle\{|0\rangle, |1\rangle\}\rangle$ invariant and have 1 in the diagonal elements outside this subspace. Therefore the optimal (i.e. largest) value is

$$\mathfrak{Re}\left(\mathrm{Tr}\left[U^*\right]\mathrm{Tr}\left[\rho U\right]\right) = \alpha(d-2+2\alpha) \tag{6.21}$$

with associated cost

$$\kappa(\rho, \mathcal{E}_U) = d-\alpha(d-2+2\alpha) = d(1-\alpha)+2\alpha(1-\alpha) = (1-\alpha)(d+2\alpha). \tag{6.22}$$

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 $d \to \infty$.

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.

In the limit of high $d \to \infty$ 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.

The second example computes the optimal cost associated to the restricted unitary channels. We can then numerical compare it to the general UI cost, which we do in Fig. 6.1. We observe that for dimension $d = 4$ the general cost is smaller for mixed states, but it appears equal for pure states. We show in Proposition 6.7 that the optimal channel when both states are pure is indeed unitary.

For the proof that the optimal map on pure inputs is unitary, we first need to show the following technical lemma, which will again be important in Section 6.4. In the following we refer several times to the 'joint support of ρ and σ'. By that we mean $\mathcal{H}_S = \mathrm{supp}\,\rho + \mathrm{supp}\,\sigma$, which allows us to write the Hilbert space as $\mathcal{H} = \mathcal{H}_S \oplus \mathcal{H}_\perp$.

Lemma 6.5
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

$$E = E_S \oplus c\Pi_\perp. \tag{6.23}$$

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.

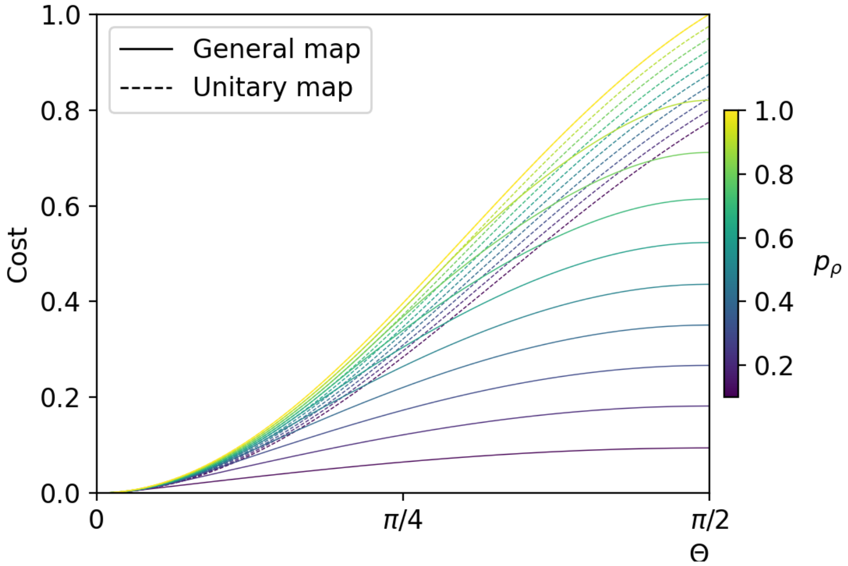

Figure 6.1: Plot comparing the global optimal cost with the result of optimising only over unitaries for various values of p_ρ, where the states are $\rho = p_\rho |0\rangle\langle 0| + (1 - p_\rho)\frac{1}{d}\mathbb{1}$ and $\sigma = p_\rho |\varphi\rangle\langle\varphi| + (1 - p_\rho)\frac{1}{d}\mathbb{1}$.

The proof of this lemma is conceptually fairly simple, but in practise becomes a very long calculation. Before presenting the proof we have a sketch with the main ideas of the proof. In Section 6.6 we see a generalisation of this result with a different approach to the proof, using results from operator theory.

Sketch of the proof. A seen in Section 2.3, we can find the Kraus decomposition of a channel by diagonalising the Choi matrix. From [Val09], we know that there exists an optimal Choi matrix that is unitary invariant. We diagonalise a generic unitary invariant Choi matrix. Then we show which structures its eigenvectors, and therefore the Kraus operators of the map, can have. Then, we show that one of the structures corresponds to Kraus operators of the form $E = E_S \oplus c\Pi_\perp$. Finally, we see that every other kind of admissible Kraus operator can be exchanged for a direct sum Kraus operator that does not change the result of the SDP or the semidefinite conditions. □

Proof. Let ρ, σ have joint support $\mathcal{H}_S$ of dimension n such that $\mathcal{H} = \mathcal{H}_S \oplus \mathcal{H}_\perp$. The dimension of $\mathcal{H}$ will be d and therefore the dimension of $\mathcal{H}_\perp$ is $d - n$. The idea of the proof is that the unitaries $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$, where Π_S is the projector onto $\mathcal{H}_S$ and $U_\perp$ are unitaries on $\mathcal{H}_\perp$, form a symmetry of the SDP in Eq. (5.5) with $K = \tilde{K}_0$. Due to [Val09], there will be an optimal Jamiołkowski matrix of the problem that is invariant under these unitaries. After showing that $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$ is indeed a symmetry of the SDP we will find the subspace of invariant matrices under $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$ as a direct sum of blocks, these blocks are in Eqs. (6.28), (6.29), (6.30), (6.31), (6.32), (6.33). With the whole subspace, we will calculate the partial transpose to find the subspace of Choi matrices that the symmetry allows. We can then find the canonical Kraus operators by finding the general eigenstates of the elements in the subspace of allowed Choi matrices. With these Kraus operators, we can then show that any any operator that has a form different to Eq. (6.23) can be simulated[1] with an operator with form as in Eq. (6.23). As we have seen in Remark 6.3 specifically working with the canonical Kraus representation should not be an issue.

First, let us show that $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$ is indeed a symmetry of the SDP. Clearly $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$ commutes with $\mathbb{1}$ and $\mathcal{S}$, as well as ρ and σ. Therefore,

$$\begin{aligned}
&\operatorname{Tr}\left[(\tilde{K}_0 \star \rho)((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))\right] \\
&\quad = \operatorname{Tr}\left[((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))(\tilde{K}_0 \star \rho)((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J\right] \\
&\quad = \operatorname{Tr}\left[(\tilde{K}_0 \star \rho)J\right],
\end{aligned} \tag{6.24}$$

and similarly the constraints of the problem are also invariant:

$$\begin{aligned}
&\operatorname{Tr}_A\left[\rho((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))\right] \\
&\quad = \operatorname{Tr}_A\left[((\Pi_S \oplus U_\perp^*) \otimes \mathbb{1})\rho((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J(\mathbb{1} \otimes (\Pi_S \oplus U_\perp^*))\right] \\
&\quad = (\Pi_S \oplus U_\perp)\operatorname{Tr}_A\left[\rho J\right](\Pi_S \oplus U_\perp^*) = (\Pi_S \oplus U_\perp)\sigma(\Pi_S \oplus U_\perp^*) = \sigma;
\end{aligned} \tag{6.25}$$

$$\begin{aligned}
&\operatorname{Tr}_B\left[((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))\right] \\
&\quad = \operatorname{Tr}_B\left[(\mathbb{1} \otimes (\Pi_S \oplus U_\perp^*))((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J((\Pi_S \oplus U_\perp^*) \otimes \mathbb{1})\right] \\
&\quad = (\Pi_S \oplus U_\perp)\operatorname{Tr}_B\left[J\right](\Pi_S \oplus U_\perp^*) = (\Pi_S \oplus U_\perp)\mathbb{1}(\Pi_S \oplus U_\perp^*) = \mathbb{1}.
\end{aligned} \tag{6.26}$$

Finally, $((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))J((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*))$ is clearly positive under partial transpose because the unitaries $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$ are local.

[1] By simulate we mean that we can exchange it for another Kraus operator that has the same contributions to the cost and the constraints of the SDP.

Because we have seen that the SDP is invariant under the unitaries $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$, we know thanks to [Val09] that there exists an optimal Jamiołkowski matrix that is invariant under these unitaries. We can twirl an arbitrary matrix to find the structure of these matrices [Wer89]. For this calculation, we take a basis of $\mathcal{H}$ that we can divide into a basis of $\mathcal{H}_S$ and a basis of $\mathcal{H}_\perp$. We will use latin letters to refer to elements of the basis of $\mathcal{H}_S$ and greek letters to refer to elements of the basis of $\mathcal{H}_\perp$. We will get that a general matrix can be written in terms of 16 blocks[2], ten of which will turn out to vanish. We consider an arbitrary matrix M and the twirling will bring aout a structure for the twirling of M.

$$\begin{aligned}
&\int dU_\perp((\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp))M((\Pi_S \oplus U_\perp^*) \otimes (\Pi_S \oplus U_\perp^*)) \\
&\quad = \int dU_\perp(\Pi_S \otimes U_\perp)M(\Pi_S \otimes U_\perp^*) + \int dU_\perp(U_\perp \otimes \Pi_S)M(U_\perp^* \otimes \Pi_S) \\
&\quad + \int dU_\perp(U_\perp \otimes \Pi_S)M(\Pi_S \otimes U_\perp^*) + \int dU_\perp(\Pi_S \otimes U_\perp)M(U_\perp^* \otimes \Pi_S) \\
&\quad + \int dU_\perp(U_\perp \otimes U_\perp)M(U_\perp^* \otimes U_\perp^*) + \int dU_\perp(\Pi_S \otimes \Pi_S)M(\Pi_S \otimes \Pi_S) \\
&\quad + \int dU_\perp(\Pi_S \otimes \Pi_S)M(U_\perp^* \otimes U_\perp^*) + \int dU_\perp(U_\perp \otimes U_\perp)M(\Pi_S \otimes \Pi_S) \\
&\quad + \int dU_\perp(U_\perp \otimes U_\perp)M(U_\perp^* \otimes \Pi_S) + \int dU_\perp(\Pi_S \otimes U_\perp)M(U_\perp^* \otimes U_\perp^*) \\
&\quad + \int dU_\perp(U_\perp \otimes \Pi_S)M(U_\perp^* \otimes U_\perp^*) + \int dU_\perp(U_\perp \otimes U_\perp)M(\Pi_S \otimes U_\perp^*) \\
&\quad + \int dU_\perp(U_\perp \otimes \Pi_S)M(\Pi_S \otimes \Pi_S) + \int dU_\perp(\Pi_S \otimes U_\perp)M(\Pi_S \otimes \Pi_S) \\
&\quad + \int dU_\perp(\Pi_S \otimes \Pi_S)M(U_\perp^* \otimes \Pi_S) + \int dU_\perp(\Pi_S \otimes \Pi_S)M(\Pi_S \otimes U_\perp^*).
\end{aligned} \tag{6.27}$$

The terms with an odd number of $U_\perp$, as well as the four terms with double $U_\perp$ on one side and double Π_S on the other are zero due to the phase symmetry of unitary matrices. That leaves the first six terms as nonzero. Using [PM17] we calculate the

[2] Each block corresponds to the choice of one of Π_S or $U_\perp$ in each element of the tensor product in each side of M, for a total of 4 binary elections. We write this explicitly.

associated subspace for each nonzero term[3]. The first term is

$$
\begin{aligned}
\int & dU_\perp (\Pi_S \otimes U_\perp) M (\Pi_S \otimes U_\perp^*) \\
&= \int dU_\perp \left(\sum_i |i\rangle\langle i| \otimes \sum_{\alpha\beta} u_{\alpha\beta} |\alpha\rangle\langle\beta| \right) M \left(\sum_j |j\rangle\langle j| \otimes \sum_{\mu\nu} \bar{u}_{\nu\mu} |\mu\rangle\langle\nu| \right) \\
&= \sum_{ij\alpha\beta\mu\nu} M_{i\beta j\mu} |i\alpha\rangle\langle j\nu| \int dU_\perp u_{\alpha\beta} \bar{u}_{\nu\mu} \\
&= \sum_{ij\alpha\beta\mu\nu} M_{i\beta j\mu} |i\alpha\rangle\langle j\nu| \frac{1}{d-n} \delta_{\alpha\nu} \delta_{\beta\mu} \\
&= \sum_{ij\alpha} \left(\frac{1}{d-n} \sum_\beta M_{i\beta j\beta} \right) |i\alpha\rangle\langle j\alpha| \\
&= \sum_{ij} B_{ij} |i\rangle\langle j| \otimes \sum_\alpha |\alpha\rangle\langle\alpha|.
\end{aligned}
\tag{6.28}
$$

As we mentioned, due to the symmetry the second term is

$$
\begin{aligned}
\int dU_\perp (U_\perp \otimes \Pi_S) M (U_\perp^* \otimes \Pi_S) &= \sum_{ij\alpha} \left(\frac{1}{d-n} \sum_\beta M_{\beta i \beta j} \right) |\alpha i\rangle\langle \alpha j| \\
&= \sum_\alpha |\alpha\rangle\langle\alpha| \otimes \sum_{ij} B_{ij}^* |i\rangle\langle j| .
\end{aligned}
\tag{6.29}
$$

[3] The first and second terms are equivalent under swap, as well as the third and fourth, so we have to calculate four distinct terms.

The third term is

$$\begin{aligned}
&\int dU_\perp (U_\perp \otimes \Pi_S) M (\Pi_S \otimes U_\perp^*) \\
&\qquad = \int dU_\perp \left(\sum_{\alpha\beta} u_{\alpha\beta} |\alpha\rangle\langle\beta| \otimes \sum_i |i\rangle\langle i| \right) M \left(\sum_j |j\rangle\langle j| \otimes \sum_{\mu\nu} \bar{u}_{\nu\mu} |\mu\rangle\langle\nu| \right) \\
&\qquad = \sum_{ij\alpha\beta\mu\nu} M_{\beta ij\mu} |\alpha i\rangle\langle j\nu| \int dU_\perp u_{\alpha\beta} \bar{u}_{\nu\mu} \\
&\qquad = \sum_{ij\alpha\beta\mu\nu} M_{\beta ij\mu} |\alpha i\rangle\langle j\nu| \frac{1}{d-n} \delta_{\alpha\nu}\delta_{\beta\mu} \\
&\qquad = \sum_{ij\alpha} \left(\frac{1}{d-n} \sum_\beta M_{\beta ij\beta} \right) |\alpha i\rangle\langle j\alpha| = \sum_{ij\alpha} A_{ij} |\alpha i\rangle\langle j\alpha| .
\end{aligned} \tag{6.30}$$

Again due to the symmetry, the fourth term is

$$\begin{aligned}
\int dU_\perp (\Pi_S \otimes U_\perp) M (U_\perp^* \otimes \Pi_S) &= \sum_{ij\alpha} \left(\frac{1}{d-n} \sum_\beta M_{i\beta\beta j} \right) |i\alpha\rangle\langle\alpha j| \\
&= \sum_{ij\alpha} A_{ij}^* |i\alpha\rangle\langle\alpha j| .
\end{aligned} \tag{6.31}$$

The fifth term is

$$\begin{aligned}
&\int dU_\perp (U_\perp \otimes U_\perp) M (U_\perp^* \otimes U_\perp^*) \\
&\quad = \int dU_\perp \left(\sum_{\alpha\beta} u_{\alpha\beta} |\alpha\rangle\langle\beta| \otimes \sum_{\gamma\epsilon} u_{\gamma\epsilon} |\gamma\rangle\langle\epsilon| \right) M \left(\sum_{\xi\tau} \bar{u}_{\tau\xi} |\xi\rangle\langle\tau| \otimes \sum_{\mu\nu} \bar{u}_{\nu\mu} |\mu\rangle\langle\nu| \right) \\
&\quad = \sum_{\alpha\beta\gamma\epsilon\xi\tau\mu\nu} M_{\beta\epsilon\xi\mu} |\alpha\gamma\rangle\langle\tau\nu| \int dU_\perp u_{\alpha\beta} u_{\gamma\epsilon} \bar{u}_{\tau\xi} \bar{u}_{\nu\mu} \\
&\quad = \sum_{\alpha\beta\gamma\epsilon\xi\tau\mu\nu} M_{\beta\epsilon\xi\mu} |\alpha\gamma\rangle\langle\tau\nu| \left(\frac{\delta_{\alpha\tau}\delta_{\gamma\nu}\delta_{\beta\xi}\delta_{\epsilon\mu} + \delta_{\alpha\nu}\delta_{\gamma\tau}\delta_{\beta\mu}\delta_{\epsilon\xi}}{(d-n)^2-1} - \frac{\delta_{\alpha\tau}\delta_{\gamma\nu}\delta_{\beta\mu}\delta_{\epsilon\xi} + \delta_{\alpha\nu}\delta_{\gamma\tau}\delta_{\beta\xi}\delta_{\epsilon\mu}}{(d-n)((d-n)^2-1)} \right) \\
&\quad = C \sum_{\alpha\gamma} |\alpha\gamma\rangle\langle\alpha\gamma| + D \sum_{\alpha\gamma} |\alpha\gamma\rangle\langle\gamma\alpha| ,
\end{aligned} \tag{6.32}$$

where in the last step, we simplified the coefficients because their exact values are not relevant —only that these terms are nonzero and independent of α, γ.

Finally, we can calculate the sixth, and last nonzero, term:

$$\begin{aligned}
&\int dU_\perp (\Pi_S \otimes \Pi_S) M (\Pi_S \otimes \Pi_S) \\
&\quad = \int dU_\perp \left(\sum_i |i\rangle\langle i| \otimes \sum_k |k\rangle\langle k| \right) M \left(\sum_j |j\rangle\langle j| \otimes \sum_\ell |\ell\rangle\langle \ell| \right) \\
&\quad = \sum_{ikj\ell} M_{ikj\ell} |ik\rangle\langle j\ell| \,.
\end{aligned} \tag{6.33}$$

We have finally found the subspace formed by the nonzero blocks of the invariant matrices under $(\Pi_S \oplus U_\perp) \otimes (\Pi_S \oplus U_\perp)$. To get the Kraus operators we first convert this subspace into the subspace of allowed Choi matrices, which will be the partial transpose of the subspace we have now. We get the following blocks, eigenstates and Kraus matrices. Note that the eigenstates we obtain are not normalised, but such that the Choi matrix has the form: $C = \sum_\psi |\psi\rangle\langle\psi|$. This allows us to calculate the Kraus matrices from them by un-vectorising.

i) From the first and second terms, in the blocks generated by $\{|i\alpha\rangle\langle j\alpha|\}$ and $\{|\alpha i\rangle\langle \alpha j|\}$ we obtain the blocks $C_S \otimes \sum_\alpha |\alpha\rangle\langle\alpha|$ and $\sum_\alpha |\alpha\rangle\langle\alpha| \otimes C_S$, which diagonalise as $\sum_i c_i |i\rangle \otimes \sum_\alpha c_\alpha |\alpha\rangle$ and $\sum_\alpha c_\alpha |\alpha\rangle \otimes \sum_i c_i |i\rangle$, respectively. Note that the part associated to $\sum_\alpha |\alpha\rangle\langle\alpha|$ must be a unit vector, that is $\sum_\alpha |c_\alpha|^2 = 1$. The associated Kraus forms are

$$\sum_{i\alpha} c_i c_\alpha |\alpha\rangle\langle i| \,, \quad \sum_{i\alpha} c_i c_\alpha |i\rangle\langle\alpha| \,. \tag{6.34}$$

ii) The other blocks form a related set of blocks, which in basis $\{|ij\rangle, |\alpha\alpha\rangle, |\alpha\gamma\rangle \;\alpha \neq \gamma\}$ is

$$\left[\begin{array}{ccccc|ccc|c}
 & & & & & A_{11} & \cdots & A_{11} & \\
 & & & & & \vdots & \cdots & \vdots & \\
 & & M + C\mathbb{1} - C\mathbb{1} & & & A_{ij} & \cdots & A_{ij} & 0 \\
 & & & & & \vdots & \cdots & \vdots & \\
 & & & & & A_{nn} & \cdots & A_{nn} & \\
\hline
A^*_{11} & \cdots & A^*_{ij} & \cdots & A^*_{nn} & & & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & & D|\Phi_+\rangle\langle\Phi_+| + C\mathbb{1} & & 0 \\
A^*_{11} & \cdots & A^*_{ij} & \cdots & A^*_{nn} & & & & \\
\hline
 & & 0 & & & & 0 & & C\mathbb{1}
\end{array}\right], \tag{6.35}$$

where $|\Phi_\perp\rangle$ is the unnormalised maximally entangled state restricted to the orthogonal subspace $|\Phi_\perp\rangle = \sum_\alpha |\alpha\alpha\rangle$. In its block, it translates to the matrix that has 1 in every matrix element. These blocks diagonalise as follows: $C\mathbb{1}$ in the subspace generated by $\{|\alpha\gamma\rangle \;\; \alpha \neq \gamma\}$ will have $(d-n)(d-n-1)$ mutually orthogonal vectors $\sqrt{C}\sum_{\alpha\neq\gamma} c_{\alpha\gamma} |\alpha\gamma\rangle$, with $\sum_{\alpha\neq\gamma} c_{\alpha\gamma} |\alpha\gamma\rangle$ a unit vector. The subspace generated by $\{|\alpha\alpha\rangle\}$ will have $d-n-1$ orthogonal vectors $\sqrt{C}\sum_\alpha c_\alpha |\alpha\alpha\rangle$ with $\sum_\alpha c_\alpha |\alpha\alpha\rangle$ unital and such that $\sum_\alpha c_\alpha = 0$. The associated Kraus matrices are

$$\sqrt{C}\sum_{\alpha\neq\gamma} c_{\alpha\gamma} |\gamma\rangle\langle\alpha| \,, \quad \sqrt{C}\sum_{\alpha} c_{\alpha} |\alpha\rangle\langle\alpha| \,. \tag{6.36}$$

Finally, consider the matrix

$$\left[\begin{array}{ccccc|c} & & & & & A_{11} \\ & & & & & \vdots \\ & & M - C\mathbb{1} & & & A_{ij} \\ & & & & & \vdots \\ & & & & & A_{nn} \\ \hline A_{11}^* & \cdots & A_{ij}^* & \cdots & A_{nn}^* & \mathrm{D} \end{array}\right] , \tag{6.37}$$

A set of $n+1$ eigenvalues and unit eigenvectors of this matrix,

$$\left\{ (\lambda, \sum_{ij} c_{ij} |ij\rangle + c\,|00\rangle) \right\} , \tag{6.38}$$

can be used to obtain the following set of n^2+1 eigenvectors of the block generated by $\{|ij\rangle\,, |\alpha\alpha\rangle\}$:

$$\sqrt{\lambda + C}\left(\sum_{ij} c_{ij} |ij\rangle + \frac{c}{d-n}\sum_\alpha |\alpha\alpha\rangle \right) , \tag{6.39}$$

which in turn yield the Kraus matrices

$$\sqrt{\lambda + C}\left(\sum_{ij} c_{ij} |j\rangle\langle i| + \frac{c}{d-n}\sum_\alpha |\alpha\rangle\langle\alpha| \right) . \tag{6.40}$$

Before moving forward, note we can simplify some expressions. Because the coefficients M_{ijkl}, A_{ij}, and D are unrelated, the coefficients of the Kraus operators in Eq. (6.40) are arbitrary, so we can simplify the expression to only show the structure of the Kraus operators: $\sum_{ij} c_{ij} |j\rangle\langle i| + c\sum_\alpha |\alpha\rangle\langle\alpha|$, where we abused notation by reusing the coefficients c_{ij} and c. We already did a similar thing in the first set of Kraus operators, where the eigenvalue is absorbed into the coefficients c_i. For the remaining Kraus operators, $\sqrt{C}\sum_{\alpha\neq\gamma} c_{\alpha\gamma} |\gamma\rangle\langle\alpha|$ and $\sqrt{C}\sum_\alpha c_\alpha |\alpha\rangle\langle\alpha|$, we can absorb the eigenvalue into the c_i as well, leaving the complete list:

$$\sum_{i\alpha} c_i c_\alpha |\alpha\rangle\langle i|\,, \tag{6.41}$$

$$\sum_{i\alpha} c_i c_\alpha |i\rangle\langle \alpha|\,, \tag{6.42}$$

$$\sum_{\alpha\neq\gamma} c_{\alpha\gamma} |\gamma\rangle\langle \alpha|\,, \tag{6.43}$$

$$\sum_{\alpha} c_\alpha |\alpha\rangle\langle \alpha|\,, \tag{6.44}$$

$$\sum_{ij} c_{ij} |j\rangle\langle i| + c\sum_\alpha |\alpha\rangle\langle\alpha|\,. \tag{6.45}$$

So far we have found 5 possible forms for the Kraus operators of optimal symmetric channels. It remains to see that 4 of these forms are either not allowed by the constraints of the problem or can be simulated by the last one, as explained at the beginning of the proof.

First, note that a channel $\mathcal{E}$ that contains the Kraus operators of the form $K = \sum_{i\alpha} c_i c_\alpha |\alpha\rangle\langle i|$ usually violates the condition $\mathcal{E}(\rho) = \sigma$. That is because

$$K\rho K^* = \left(\sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle\right)\sum_{\alpha\gamma} c_\alpha^* c_\gamma |\gamma\rangle\langle\alpha|\,. \tag{6.46}$$

If $\sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle \neq 0$ we obtain a positive term outside the support of σ that can not be eliminated with other Kraus operators, because their operation is positive. Thus we will only consider $K = \sum_i c_i c_\alpha |\alpha\rangle\langle i|$ with $\sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle = 0$ in the remainder of the proof.

Let K be a Kraus operator of either of the forms in Eq. (6.42) through Eq. (6.44). Then $\mathrm{Tr}\,[K\rho] = 0$, $K\rho K^* = 0$ and $K^*K \in \mathcal{B}(\mathcal{H}_\perp)$. Therefore we can ignore these operators and consider instead Kraus matrices $\{K'\}$ with form as in Eq. (6.41) and Eq. (6.45) with the conditions $\sum_{K'} K'^* K' = \Pi_S \oplus \Omega$, with $0 \leq \Omega \leq \Pi_\perp$ and $\sum_{K'} K'\rho K'^* = \sigma$. Note that $\Pi_\perp \sum_{K'} K'^* K' \Pi_\perp = \varepsilon \Pi_\perp$, $0 \leq \varepsilon \leq 1$, because the contribution to the orthogonal subspace is zero for Kraus operators as in Eq. (6.41) and proportional to the identity for Kraus operators as in Eq. (6.45). We can complete this set of Kraus operators with $K_\perp = \sqrt{1-\varepsilon}\sum_\alpha |\alpha\rangle\langle\alpha|$. This operator also fulfils $\mathrm{Tr}\left[K_\perp \rho\right] = 0$ and $K_\perp \rho K_\perp^* = 0$, so it does not affect the rest of the optimisation problem.

Finally, we need to see that if an optimal solution has a Kraus operator $K = \sum_{i\alpha} c_i c_\alpha |\alpha\rangle\langle i|$ such that $\sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle = 0$, we can replace it with Kraus operator with form as in Eq. (6.45), keeping every equation in the optimisation the same. Note that $K\rho K^* = 0$, $K^*K = \sum_{ij} c_i c_j^* |j\rangle\langle i|$ and $\mathrm{Tr}\left[K\rho\right] = 0$. Let

$$K' = |\psi\rangle\langle\psi| \quad \text{with} \quad |\psi\rangle = \frac{1}{\sqrt[4]{\sum_i |c_i|^2}} \sum_i c_i^* |i\rangle\,. \tag{6.47}$$

Note that K' has the form as in Eq. (6.45). It also has the same properties as K with respect to all three conditions:

$$\begin{aligned} K'\rho K'^* &= \frac{1}{\sum_i |c_i|^2} \sum_{ki} c_k^* c_i |k\rangle\langle i|\,\rho \sum_{j\ell} c_j^* c_\ell |j\rangle\langle\ell| \\ &= \left(\sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle\right) \frac{1}{\sum_i |c_i|^2} \sum_{k\ell} c_k^* c_\ell |k\rangle\langle\ell| = 0 \end{aligned} \tag{6.48}$$

$$\mathrm{Tr}\left[K'\rho\right] = \langle\psi|\rho|\psi\rangle = \frac{1}{\sqrt{\sum_i |c_i|^2}} \sum_{ij} c_i c_j^* \langle i|\,\rho\,|j\rangle = 0 \tag{6.49}$$

$$\begin{aligned} K'^* K' &= |\psi\rangle\langle\psi|\psi\rangle\langle\psi| = \frac{1}{\sum_i |c_i|^2} \left(\sum_{ij} c_i c_j^* \langle i|j\rangle\right)\left(\sum_{ij} c_j^* c_i |j\rangle\langle i|\right) \\ &= \sum_{ij} c_j^* c_i |j\rangle\langle i|\,. \end{aligned} \tag{6.50}$$

We have seen that the optimisation can be taken over Kraus matrices of the form $\sum_{ij} c_{ij} |j\rangle\langle i| + c\sum_\alpha |\alpha\rangle\langle\alpha|$, which is equivalent to an arbitrary matrix on the joint support of ρ and σ and a matrix proportional to the identity in the orthogonal space, that is $K_S \oplus c\Pi_\perp$. □

As seen in Example 6.4, the cost is a function of the dimension even if the joint support is not the full space. Theorem 6.6 shows the relation between the cost on the joint support and the cost in the full space.

Theorem 6.6
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

$$\mathcal{K}(\rho,\sigma) = 1 - \frac{1}{d}\max_{\{E_k\}}\left(\mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[E_k\rho\right]\mathrm{Tr}\left[E_k^*\right]\right) + d_\perp\sqrt{\sum_k |\mathrm{Tr}(E_k\rho)|^2}\right) \tag{6.51}$$

$$= 1 - \frac{1}{d}\max_{\mathcal{E}}\left(\mathfrak{Re}\left(\sum_{ij} p_i \langle i|\,\mathcal{E}(|i\rangle\langle j|)\,|j\rangle\right) + d_\perp\sqrt{\sum_{ij} p_i p_j \langle i|\,\mathcal{E}(|i\rangle\langle j|)\,|j\rangle}\right), \tag{6.52}$$

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

$$\mathcal{K}(\rho,\sigma) = 1 - \frac{1}{d}\max_{C_S}\left(\mathfrak{Re}(\langle\rho|\,C_S\,|\Phi_S\rangle) + d_\perp\sqrt{\langle\rho|\,C_S\,|\rho\rangle}\right) \tag{6.53}$$
$$\text{s.t. } C_S \geq 0,\ \mathrm{Tr}_B\, C_S = \mathbb{1}_S,\ \mathrm{Tr}\left[\rho^T C_S\right] = \sigma,$$

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 $|\rho\rangle = \sum_{ii} p_i\,|ii\rangle$.

Note that $\max_{\{E_k\}}\mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[E_k\rho\right]\mathrm{Tr}\left[E_k\right]\right)$ is the nontrivial part of the cost in $\mathcal{H}_S$. Because of the extra term $d_\perp\sqrt{\langle\rho|\,C_S\,|\rho\rangle}$ within the maximisation, the optimal channel is in general not the same for different dimension of the whole space d.

Proof. From Lemma 6.5, 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. (6.13), then is

$$\mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[\rho E_k\right]\mathrm{Tr}\left[E_k^*\right]\right) = \mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[\rho E_{kS}\right]\mathrm{Tr}\left[E_{kS}^*\right] + d_\perp\sum_k c_k^*\,\mathrm{Tr}\left[\rho E_{kS}\right]\right). \tag{6.54}$$

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 $c_k^*\,\mathrm{Tr}\left[\rho E_{kS}\right] = |c_k||\mathrm{Tr}\left[\rho E_{kS}\right]|$, 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 $\mathbf{v} = (\mathrm{Tr}[\rho E_{kS}])$. 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

$$\mathfrak{Re}\left(\sum_k \mathrm{Tr}[\rho E_{kS}]\,\mathrm{Tr}[E_{kS}^*] + d_\perp \sqrt{\sum_k |\mathrm{Tr}[\rho E_{kS}]|^2}\right), \tag{6.55}$$

where $\sqrt{\sum_k |\mathrm{Tr}[\rho E_k]|^2} = \sqrt{\langle \mathbf{v}|\mathbf{v}\rangle}$ is the norm of $\mathbf{v}$ and $c_k = \frac{\mathrm{Tr}[\rho E_{kS}]}{\sqrt{\sum_{k'} |\mathrm{Tr}[\rho E_{k'S}]|^2}}$. Finally, we can take the maximum over all admissible channels to obtain the optimal channel. The equivalent equations Eq. (6.52) and Eq. (6.53) follow immediately from the general expressions Eq. (6.12) and Eq. (6.10), respectively. ☐

To finish this section we show that for pure states the optimal map is unitary.

Proposition 6.7
Consider two pure states in a d-dimensional Hilbert space $\mathcal{H}$, WLOG $\rho = |0\rangle\langle 0|$ and $\sigma = (\alpha|0\rangle + \sqrt{1-\alpha^2}|1\rangle)(\alpha\langle 0| + \sqrt{1-\alpha^2}\langle 1|)$, with $\alpha \in \mathbb{R}$. Then

$$\mathcal{K}(\rho,\sigma) = (1-\alpha)(d+2\alpha) \tag{6.56}$$

and the optimal channel is given by conjugation with the unitary

$$U = \begin{bmatrix} \alpha & -\sqrt{1-\alpha^2} \\ \sqrt{1-\alpha^2} & \alpha \end{bmatrix} \oplus \mathbb{1}_{d-2}. \tag{6.57}$$

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 6.5 and Theorem 6.6, we can write E_k as follows

$$E_k = \sqrt{p_k}\left[\begin{array}{cc|c} \alpha & \gamma_k & 0 \\ \sqrt{1-\alpha^2} & \beta_k & \\ \hline \multicolumn{2}{c|}{0} & \Pi_\perp \end{array}\right], \tag{6.58}$$

with $c_k = \frac{\mathrm{Tr}[\rho E_k]}{\sqrt{\sum_{k'} |\mathrm{Tr}[\rho E_k]^2|}} = \frac{\sqrt{p_k}\alpha}{\alpha} = \sqrt{p_k}$. We can take β_k, γ_k to be positive, since $\mathrm{Tr}[\rho E_k] = \sqrt{p_k}\alpha$ is and we are maximising the real part of the product with $\sqrt{p_k}(\alpha + \beta_k^*)$. If we calculate $\sum_k E_k^* E_k$ we obtain

$$\mathbb{1} = \sum_k p_k \left[\begin{array}{cc|c} 1 & \alpha\gamma_k + \beta_k\sqrt{1-\alpha^2} & 0 \\ \alpha\gamma_k + \beta_k\sqrt{1-\alpha^2} & \beta_k^2 + \gamma_k^2 & \\ \hline \multicolumn{2}{c|}{0} & \Pi_\perp \end{array}\right]. \tag{6.59}$$

For the cost, we want to maximise $\sum_k p_k (\beta)_k$ constrained to

$$0 = \sum_k p_k(\alpha\gamma_k + \beta_k\sqrt{1-\alpha^2}) \tag{6.60}$$

$$1 = \sum_k p_k(\beta_k^2 + \gamma_k^2). \tag{6.61}$$

This problem has as variables: $p_k,\ \gamma_k,\ \beta_k$, and even the size of the index set of k, $|I|$. To simplify, fix $|I|$ and p_k. We can then define the Lagrangian

$$\begin{aligned}\mathcal{L}(\vec{\beta},\vec{\gamma},\mu,\nu) = &\sum_k p_k\beta_k + \mu\left(\sum_k p_k(\alpha\gamma_k + \beta_k\sqrt{1-\alpha^2})\right)\\ &+\nu\left(\sum_k p_k(\beta_k^2+\gamma_k^2) - 1\right)\end{aligned} \tag{6.62}$$

and optimise with its gradient

$$\begin{aligned}0 &= \nabla_{\vec{\beta},\vec{\gamma},\mu,\nu}\mathcal{L}(\vec{\beta},\vec{\gamma},\mu,\nu)\\ &= \left(p_k + \mu p_k\sqrt{1-\alpha^2} + \nu p_k 2\beta_k,\ \mu p_k\alpha + 2\nu p_k\gamma_k,\right.\\ &\qquad \left.\sum_k p_k(\alpha\gamma_k + \beta_k\sqrt{1-\alpha^2}), \sum_k p_k(\beta_k^2+\gamma_k^2) - 1\right)\\ &= \left(1 + \mu\sqrt{1-\alpha^2} + \nu 2\beta_k,\ \mu\alpha + 2\nu\gamma_k, \sum_k p_k(\alpha\gamma_k + \beta_k\sqrt{1-\alpha^2}),\right.\\ &\qquad \left.\sum_k p_k(\beta_k^2+\gamma_k^2) - 1\right).\end{aligned} \tag{6.63}$$

We see that the values of β_k and γ_k do not depend on k. This simplifies the equations to maximising β such that

$$0 = \alpha\gamma + \beta\sqrt{1-\alpha^2}, \tag{6.64}$$

$$1 = \beta^2 + \gamma^2. \tag{6.65}$$

It is clear that the optimal will be $\beta = \alpha$, $\gamma = -\sqrt{1-\alpha^2}$. The nontrivial part of the cost associated to each Kraus operator will be $p_k\alpha(\alpha+\beta+(d-2)) = p_k\alpha(2\alpha+d-2)$.

Thus the total cost is

$$\begin{aligned}\mathcal{K}(\rho,\sigma) &= d - \sum_k p_k(\alpha(2\alpha+d-2)) = d - (\alpha(2\alpha+d-2)) \\ &= (1-\alpha)(d+2\alpha),\end{aligned} \tag{6.66}$$

and we know from Eq. (6.22) in Example 6.4 that this cost is attained by the unitary, finishing the proof. □

6.3 Commuting states

This section takes the unitary invariant cost and in the case of commuting states. We want to compare this case to the arbitrary case. We also show that this cost is can be calculated analytically and that the optimal map is purely quantum. Finally, we compare the restriction to also classical channels, we show it is much larger numerically than the general quantum transport and we show its relation to the total variation distance, the equivalent to the unitary invariant case in classical optimal transport.

First, we show how we can lower bound the cost between any pair of states with a cost between diagonal states.

Proposition 6.8
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, $\mathcal{E}_\rho(x) = \sum_i \langle i|x|i\rangle\, |i\rangle\langle i|$. Then

$$\mathcal{K}(\rho,\sigma) \geq \mathcal{K}(\rho,\mathcal{E}_\rho(\sigma)). \tag{6.67}$$

Sketch of the proof. This proof relatively short, but its logic is somewhat subtle. The idea of the proof is that every channel between ρ and σ induces a classical stochastic map between ρ and the diagonal of σ in the diagonal basis of ρ, that is a classical stochastic map between ρ and $\mathcal{E}_\rho(\sigma)$. We show that the cost associated to this map is a lower bound to the cost associated to the original map (Eqs. (6.70), (6.77)). Because we can do it for every map we can do it for the optimal one, thus obtaining a lower bound. □

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 in the product basis, which need to be positive, are $\langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle$ and fulfil

$$\sum_j \langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle = \mathrm{Tr}\left[\mathbb{1}\mathcal{E}(|i\rangle\langle i|)\right] = 1. \tag{6.68}$$

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 $\sigma = \mathrm{Tr}_A\left[\rho C\right] = \sum_i p_i \mathcal{E}(|i\rangle\langle i|)$. We can apply the pinching $\mathcal{E}_\rho$ to this equation to obtain

$$\mathcal{E}_\rho(\sigma) = \sum_i p_i \mathcal{E}_\rho\left(\mathcal{E}(|i\rangle\langle i|)\right) = \sum_{ij} p_i \langle j|\mathcal{E}(|i\rangle\langle i|)|j\rangle\, |j\rangle\langle j| = \sum_{ij} p_i p(j|i)\, |j\rangle\langle j| \tag{6.69}$$

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 non 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,

$$\langle i|\,\mathcal{E}(|i\rangle\langle j|)\,|j\rangle \le \sqrt{\langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle\ \langle j|\mathcal{E}(|j\rangle\langle j|)|j\rangle} = \sqrt{p(i|i)p(j|j)}. \tag{6.70}$$

Finally, with Eq. (6.70) we obtain 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:

$$\begin{aligned}
\mathrm{Tr}\left[|\Phi_+\rangle\langle\Phi_+|\,(\rho\star C)\right] &= \sum_{i'j'ij} \langle i'i'|\,\frac{p_i+p_j}{2}\,\left(|i\rangle\langle j|\otimes\mathcal{E}(|i\rangle\langle j|)\right)|j'j'\rangle \\
&= \sum_{ij}\frac{p_i+p_j}{2}\,\langle i|\,\mathcal{E}(|i\rangle\langle j|)\,|j\rangle \\
&= \sum_i p_i\,\langle i|\mathcal{E}(|i\rangle\langle i|)|i\rangle + \sum_{i\neq j}\frac{p_i+p_j}{2}\,\langle i|\,\mathcal{E}(|i\rangle\langle j|)\,|j\rangle \\
&\le \sum_i p_i p(i|i) + \sum_{i\neq j}\frac{p_i+p_j}{2}\sqrt{p(i|i)p(j|j)} \\
&= \sum_i p_i p(i|i) + \sum_{i\neq j} p_i\sqrt{p(i|i)p(j|j)}.
\end{aligned} \tag{6.71}$$

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

$$\mathrm{Tr}_B\left[C_p\right] = \mathrm{Tr}_B\left[|\phi\rangle\langle\phi| + \sum_{i\neq j} p(j|i)\,|ij\rangle\langle ij|\right] \tag{6.72}$$

$$= \sum_i p(i|i) |i\rangle\langle i| + \sum_{i\neq j} p(j|i) |i\rangle\langle i| = \mathbb{1}, \tag{6.73}$$

$$\mathrm{Tr}_A\left[\rho^T C_p\right] = \mathrm{Tr}_A\left[\sum_k p_k(|k\rangle\langle k| \otimes \mathbb{1})\left(|\phi\rangle\langle\phi| + \sum_{i\neq j} p(i|j)\,|ij\rangle\langle ij|\right)\right] \tag{6.74}$$

$$= \sum_i p_i p(i|i) |i\rangle\langle i| + \sum_{i\neq j} p_i p(j|i) |j\rangle\langle j| = \sum_{ij} p(j|i) p_i |j\rangle\langle j| \tag{6.75}$$

$$= \mathcal{E}_\rho(\sigma), \tag{6.76}$$

where the last equality was seen in Eq. (6.69). 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

$$\begin{aligned}\langle i| \mathcal{E}_{C_p}(|i\rangle\langle j|) |j\rangle &= \langle i| \mathrm{Tr}_A\left[(|j\rangle\langle i| \otimes \mathbb{1})\,|\phi\rangle\langle\phi| + \sum_{i'\neq j'} p(j'|i')\,|i'j'\rangle\langle i'j'|\right]|j\rangle \\ &= \langle i| \mathrm{Tr}_A\left[(|j\rangle\langle i| \otimes \mathbb{1})\,|\phi\rangle\langle\phi|\right]|j\rangle = \sqrt{p(i|i)p(j|j)},\end{aligned} \tag{6.77}$$

which is the tight version of Eq. (6.70). This means the bound (6.71) 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

$$\mathcal{K}(\rho,\sigma) \geq d - \sum_i p_i p(i|i) + \sum_{i\neq j} p_i \sqrt{p(i|i)p(j|j)} \tag{6.78}$$

$$= \mathrm{Tr}\left[K_0(\rho \star C_p^{T_A})\right] \geq \mathcal{K}(\rho, \mathcal{E}_\rho(\sigma)), \tag{6.79}$$

finalising the proof. □

The main value of Proposition 6.8 comes from its relation to the following result, which allows us to analytically calculate the cost between commuting states, thus allowing us to always have a lower bound for the cost.

Proposition 6.9
Let ρ and σ commute. In a common diagonal basis they can be written as $\rho = \sum_i p_i |i\rangle\langle i|$, $\sigma = \sum_i q_i |i\rangle\langle i|$. Then

$$\mathcal{K}(\rho,\sigma) = \frac{1}{d}\left(d - \sum_{ij} p_i \sqrt{\min\{1, \tfrac{q_i}{p_i}\} \min\{1, \tfrac{q_j}{p_j}\}}\right) \tag{6.80}$$

Proof. We have seen in the proof of Proposition 6.8 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. (6.71) tight. Therefore the problem is equivalent to the following optimisation over classical stochastic maps:

$$\min_{p(j|i)} d - \sum_i p_i p(i|i) - \sum_{i \neq j} p_i \sqrt{p(i|i)p(j|j)}, \tag{6.81}$$

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 [Vil08]. 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 succintly $p(i|i) = \min\{1, \frac{q_i}{p_i}\}$. This will maximise the amount of weight the map leaves in place. When $p(i|i) < 1$ the transport plan $p(j|i)$ can be completed by distributing the remaining weight among $j \neq i$ arbitrarily such that the map is admissible, as these weights do not contribute to the cost. Hence the proof is finished. ☐

The proof gives us information not only on the optimal cost but also on the optimal channel that attains this cost. This is explained in the following remark:

Remark 6.10
The optimal channel associated to the unitary invariant quantum optimal transport problem between commuting states with common basis $\{|i\rangle\}$ will be, as we have seen in Eq. (6.77) and the proof of Proposition 6.9, in Choi matrix form:

$$C = |\phi\rangle\langle\phi| + \sum_{i \neq j} p(j|i) |ij\rangle\langle ij|, \tag{6.82}$$

with $|\phi\rangle = \sum_i \sqrt{p(i|i)} |ii\rangle$ and $p(i|i)$ as previously defined in the proofs of Proposition 6.8 and Proposition 6.9, 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:

$$\{E_k\} = \left\{ \sum_i \sqrt{p(i|i)} |i\rangle\langle i| ; \sqrt{p(j|i)} |j\rangle\langle i| , i \neq j \right\}. \tag{6.83}$$

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) [SAP17].*

Finally, we compare the cost between commuting states to the cost attained by purely classical channels.

Remark 6.11
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. Without loss of generality, we let* ρ *and* σ *be diagonal in the canonical basis, with eigenvalues denoted by* p_i *and* q_j *respectively, and the Jamiołkowski matrix be diagonal in the product of canonical basis. As seen in Remark 4.8 a state over time in this case will be of the form* $Q_{\text{cl}} = \sum_{ij} p_{ij}\,|ij\rangle\langle ij|$, *with* p_{ij} *a joint probability distribution, that is a classical coupling, with marginals* p_i *and* q_j. *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:*

$$\mathcal{K}_{\text{classical}}(\rho,\sigma) = \min_{Q_{\text{cl}}} \operatorname{Tr}\left[(\mathbb{1} - \frac{1}{d}\mathcal{S})Q_{\text{cl}}\right] = \tag{6.84}$$

$$= \min_{\{p_{ij}\}} \sum_{ij} p_{ij} - \frac{1}{d}\sum_i p_{ii} = 1 - \frac{1}{d}\max_{\{p_{ij}\}}\sum_i p_{ii} =$$

$$= 1 - \frac{1}{d}(1 - \frac{1}{2}|\rho - \sigma|) \geq 1 - \frac{1}{d}. \tag{6.85}$$

Where in the optimization of the last equality we used again that maximum value of $0 \leq p(i|i) \leq 1$ *subject to* $q_i = \sum_j p(i|j)p_j$ *is* $p(i|i) = \min\{1, \frac{q_i}{p_i}\}$ *or equivalently* $p_{ii} = p(i|i)p_i = \min\{p_i, q_i\}$, *and that* $|\rho - \sigma| = \sum_i |p_i - q_i| = \sum_i (p_i + q_i - 2\min\{p_i, q_i\})$

We obtain the cost in Proposition 6.9 without the term $-\frac{1}{d}\sum_{i\neq j} p_i \sqrt{p(i|i)p(j|j)}$[4]. *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.*

[4] We were using $K_0 = d\tilde{K}_0$, so the term $\frac{1}{d}$ was not there.

6.4 Limit $d \to \infty$

We have seen in Theorem 6.6 that the cost between two states on a Hilbert space $\mathcal{H}_S \subset \mathcal{H}$ depends on the dimension of $\mathcal{H}$ even if the support of these states does not explore the full space. In this section we aim to remove this dependence by taking the limit of the dimensions of the larger spaces to ∞. We see the limit exists and how to calculate it.

The commuting case can be calculated analytically and contains all the ideas used for the general result. We compute it as an example.

Example 6.12
Let us rewrite Eq. (6.80). We can split the sum into a sum over i and a sum over j as

$$\begin{aligned} &\frac{1}{d}\left(d - \sum_i p_i p(i|i) - \sum_{i\neq j} p_i \sqrt{p(i|i)p(j|j)}\right) \\ &\quad = \frac{1}{d}\left(d - \left(\sum_i p_i \sqrt{p(i|i)}\right)\left(\sum_j \sqrt{p(j|j)}\right)\right). \end{aligned} \tag{6.86}$$

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 and all the $p(i|i)$ have a minus sign in the minimisation, so we want them as large as possible. Therefore, if $p_i = 0$, we take $p(i|i) = 1$.

With this we can calculate the limit. Let ρ, σ in a n dimensional Hilbert space commute. 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:

$$\begin{aligned} \mathcal{K}_d(\rho,\sigma) &= \frac{1}{d}\left(d - N\left(M + d - n\right)\right) = 1 - \frac{NM}{d} - N + \frac{Nn}{d} \\ &\xrightarrow[d\to\infty]{} 1 - N = 1 - \sum_i p_i \sqrt{p(i|i)}. \end{aligned} \tag{6.87}$$

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:*

$$\begin{aligned}\mathcal{K}_\infty(\rho,\sigma) &= 1 - \sum_i p_i \sqrt{p(i|i)} = \sum_i p_i (1 - \sqrt{p(i|i)}) \\ &= \sum_i \sqrt{p_i}(\sqrt{p_i} - \sqrt{p(i|i)p_i}) = \sum_{q_i < p_i} \sqrt{p_i}(\sqrt{p_i} - \sqrt{q_i}),\end{aligned} \tag{6.88}$$

where the last equality comes from the definition of the optimal $p(i|i) = \min\{1, \frac{q_i}{p_i}\}$ in Eq. (6.80).

The general case uses the same ideas starting from Theorem 6.6.

Theorem 6.13
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,

$$\mathcal{K}_\infty(\rho,\sigma) = \lim_{d\to\infty} \mathcal{K}_d(\rho,\sigma) = 1 - \max_{\{E_k\}} \sqrt{\sum_k \left|\mathrm{Tr}\left[\rho E_k\right]\right|^2}, \tag{6.89}$$

where the maximisation is over all sets of admissible Kraus operators in $\mathcal{H}_S$.

Before we give the proof, note that the channel in the theorem is not necessarily the optimal channel for the problem defined in $\mathcal{H}_S$. That said, numerical evidence suggests that these channels are equal or at least very close.

Proof. Using Eq. (6.53) in Theorem 6.6 we can immediately obtain the result. Note that, like in the example, for every set of Kraus operators $\{E_k\}$,

$$\mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[E_k\rho\right]\mathrm{Tr}\left[E_k^*\right]\right) \quad \text{and} \quad \sqrt{\sum_k \left|\mathrm{Tr}\left[\rho E_k\right]\right|^2} \tag{6.90}$$

are fixed regardless of total dimension d. We can take the limit of Eq. (6.53) immediately:

$$\begin{aligned}\mathcal{K}_\infty(\rho,\sigma) = \lim_{d\to\infty} 1 - \frac{1}{d}\max_{\{E_k\}} &\left(\mathfrak{Re}\left(\sum_k \mathrm{Tr}\left[E_k\rho\right]\mathrm{Tr}\left[E_k^*\right]\right)\right. \\ &\left.+(d-n)\sqrt{\sum_k \left|\mathrm{Tr}\left[\rho E_k\right]\right|^2}\right) \\ =& 1 - \max_{\{E_k\}}\sqrt{\sum_k \left|\mathrm{Tr}\left[\rho E_k\right]\right|^2},\end{aligned} \tag{6.91}$$

as claimed. □

Finally, some remarks on the computability of $\mathcal{K}_\infty$ and the proof that the general theorem yields the example when ρ, σ commute.

Remark 6.14 ($\mathcal{K}_\infty$ can be calculated with an SDP)
consider the following SDP:

$$\begin{aligned} \max_J \quad & \mathrm{Tr}\left[\rho\mathcal{S}\rho J\right] \\ \textit{s.t.} \quad & \begin{cases} \mathrm{Tr}_A\left[\rho J\right] = \sigma \\ \mathrm{Tr}_B\, J = \mathbb{1} \\ J^{T_A} \geq 0 \end{cases}, \end{aligned} \tag{6.92}$$

where we simplified $(\rho\otimes\mathbb{1})$ to ρ. We can write the objective function as a function of the Kraus operators instead of the Jamiołkowski matrix. Recall that $J = \sum_k(\mathbb{1}\otimes E_k)\mathcal{S}(\mathbb{1}\otimes E_k^*)$. Then:

$$\begin{aligned}\mathrm{Tr}\left[\rho\mathcal{S}\rho J\right] &= \mathrm{Tr}\left[\rho\mathcal{S}\rho\sum_k(\mathbb{1}\otimes E_k)\mathcal{S}(\mathbb{1}\otimes E_k^*)\right] = \sum_k \mathrm{Tr}\left[\mathcal{S}(\rho\otimes\rho)(E_k^*\otimes E_k)S\right] \\ &= \sum_k \mathrm{Tr}\left[\rho E_k^*\right]\mathrm{Tr}\left[\rho E_k\right] = \sum_k \left|\mathrm{Tr}\left[\rho E_k\right]\right|^2.\end{aligned} \tag{6.93}$$

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

$$\mathcal{K}_\infty(\rho,\sigma) = 1 - \sqrt{\max_J \mathrm{Tr}\left[\rho\mathcal{S}\rho J\right]}. \tag{6.94}$$

Remark 6.15 (Theorem 6.13 reduces to Example 6.12)
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.* (6.83)*. Furthermore,* $\mathrm{Tr}\left[\rho\sqrt{p(j|i)}\,|j\rangle\langle i|\right]=0$ *for all* i,j *because* ρ *is diagonal and*

$$\mathrm{Tr}\left[\rho\sum_i\sqrt{p(i|i)}\,|i\rangle\langle i|\right]=\sum_i p_i\sqrt{p(i|i)}. \tag{6.95}$$

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

$$\mathcal{K}_\infty(\rho,\sigma)=1-\sqrt{\left(\sum_i p_i\sqrt{p(i|i)}\right)^2}=1-\sum_i p_i\sqrt{p(i|i)}, \tag{6.96}$$

equal to Eq. (6.87)*.*

6.5 Asymmetry and discontinuity of the cost function

We provide an example now with some interesting properties. Consider the UI cost on a finite dimensional Hilbert space $\mathcal{H}$. Let $\rho=|0\rangle\langle 0|$ and $\sigma=(1-p_\sigma)\rho+p_\sigma\mathbb{1}/d$. Fig. 6.2 shows the *symmetry gap* between ρ and σ:

$$\mathcal{K}(\rho,\sigma)-\mathcal{K}(\sigma,\rho). \tag{6.97}$$

The cost is symmetric if the symmetry gap is zero.

Clearly, Fig. 6.2 shows that the cost is not symmetric for the unitary invariant cost matrix K_0. Also clearly, $\mathcal{S}K_0\mathcal{S}=K_0$. Therefore this example answers the question posed in Section 5.5: we see that a symmetric cost matrix does not yield a symmetric transport cost.

Fig. 6.2 also shows a discontinuity when $p_\sigma\searrow 0$. 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

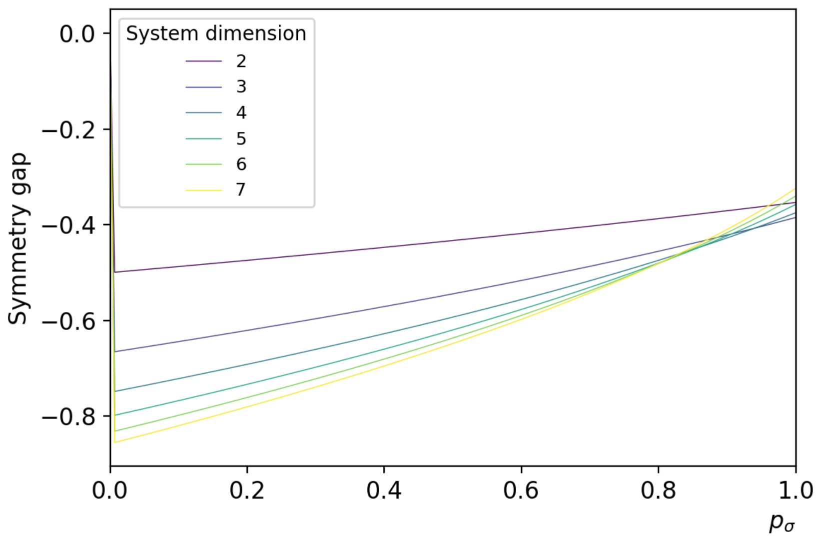

Figure 6.2: Symmetry gap between for $\tilde{K}_0 = \mathbb{1} - \frac{1}{d}\mathcal{S}$, $\rho = |0\rangle\langle 0|$ and $\sigma = (1 - p_\sigma)\rho + p_\sigma \mathbb{1}/d$.

a discontinuity in $\mathcal{K}(\sigma, \rho)$, which translates to the symmetry gap. We see how the discontinuity in the set of admissible maps translates to the discontinuity of the cost that analytically in the following example.

Example 6.16
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: $Q = \rho \otimes \sigma$. The associated cost is

$$\begin{aligned}\mathcal{K}(\rho(\varepsilon), \sigma) &= 1 - \frac{1}{d}\operatorname{Tr}\left[\mathcal{S}(\rho \otimes \sigma)\right] = 1 - \operatorname{Tr}\left[\rho\sigma\right] \\ &= 1 - \frac{1}{d}\left(\frac{1}{2}(1-\varepsilon) + \varepsilon\frac{1}{d}\right) \xrightarrow[\varepsilon \to 0]{} 1 - \frac{1}{2d}.\end{aligned} \tag{6.98}$$

If we consider $\varepsilon = 0$, ρ and σ are pure and we can use Proposition 6.7 to obtain

$$\mathcal{K}(\rho(0), \sigma) = \frac{1}{d}\left(1 - \frac{1}{\sqrt{2}}\right)\left(d + 2\frac{1}{\sqrt{2}}\right) = 1 - \frac{d + 2 - 2\sqrt{2}}{2d}. \tag{6.99}$$

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

$$\mathcal{K}(\rho(0), \sigma) < \lim_{\varepsilon \searrow 0} \mathcal{K}(\rho(\varepsilon), \sigma). \tag{6.100}$$

6.6 Toward a Hamiltonian based quantum optimal transport

The aim of this section is to construct a cost matrix associated with a Hamiltonian H on a d-dimensional Hilbert space $\mathcal{H}$, which quantifies, in some sense, the energy variations on states ρ to σ.

The guiding principle is that any meaningful notion of "energy variation" should depend only on the intrinsic structure defined by the Hamiltonian, and not on the particular representation of the energy eigenbasis. In other words, if we perform a unitary transformation that commutes with H, that is, one that does not alter the energy eigenspaces, the corresponding cost should remain invariant. This motivates us to require that, as in Eq. (6.1), the cost function be invariant under all such unitaries. Formally, we impose the Hamiltonian-invariance condition

$$\kappa(\rho, \mathcal{E}) = \kappa(U\rho U^*, \mathcal{U} \circ \mathcal{E} \circ \mathcal{U}^{-1}) \quad \forall U \in \{U \in U(d) |\, [U, H] = 0\}\,, \tag{6.101}$$

where $\mathcal{U}(X) = UXU^*$ is the corresponding unitary channel. This condition ensures that the cost, and consequently the optimal transport cost, are independent of the choice of energy basis.

To analyse the consequences of this invariance, let us introduce the relevant algebraic structures. We write the spectral decomposition of the Hamiltonian as

$$H = \sum_i \lambda_i P_i, \quad \lambda_i \in \mathbb{R},\ \lambda_i \neq \lambda_j \text{ for } i \neq j, \tag{6.102}$$

where each P_i is an orthogonal projection.

We define the set $\mathcal{C}_H$ as the collection of tensor products of identical H-commuting operators acting on both subsystems:

$$\mathcal{C}_H = \{A \otimes A \in \mathcal{B}(\mathcal{H} \otimes \mathcal{H}) \,|\, [H, A] = 0\}\,. \tag{6.103}$$

The unitaries that commute with H act as symmetries of the Hamiltonian, and the corresponding transformations $A \mapsto UAU^*$ leave the structure of $\mathcal{C}_H$ invariant.

In full analogy with the unitary-invariant cost, in particular following the proof in Theorem 6.1, we see that the cost matrix must commute with all unitaries $U \otimes U$ such that $[U, H] = 0$, or in other words it has to commute with the all the unitary elements of C_H. Using Lemma 2.7, this is equivalent to the commutation with $\mathcal{C}_H$:

$$(\mathcal{C}_H \cap U(d^2))' \subseteq (\mathcal{C}_H \cap U(d^2)'')' = (\mathcal{C}_H' \cup U(d^2)')'' = (\mathcal{C}_H' \cup \{\mathbb{1}\})'' = \mathcal{C}_H''' = \mathcal{C}_H'.$$

For the first inclusion we used that if $A \subseteq B$, then $A' \supseteq B'$ with $A = U(d^2)$ and $B = U(d^2)''$. The converse inclusion is clear because of this equation as well, thus

$$(\mathcal{C}_H \cap U(d^2))' = \mathcal{C}'_H. \tag{6.104}$$

Therefore, the appropriate cost matrices K are precisely those that commute with $\mathcal{C}_H$, ensuring that $\kappa(\rho, \mathcal{E})$ remains invariant under all transformations that preserve the eigenspace structure of the Hamiltonian H.

In Section 6.6.1 we calculate the commutant of $\mathcal{C}_H$, then in Section 6.6.2 we explore this set for good cost matrix candidates.

6.6.1 Finding the commutant

In this section we want to characterise the commutant of $\mathcal{C}_H$, which we do in Theorem 6.19. Before that we show two results that aid in this proof, together with some of the results from Section 2.1.2.

We start by showing that the eigenprojections of H, $\{P_i\}$ are in the bicommutant of H.

Proposition 6.17
Let $H = \sum_i \lambda_i P_i$ be a self adjoint operator on a Hilbert space $\mathcal{H}$. Let $\mathbb{C}[H]$ be the complex polynomials evaluated on H. Then

$$\mathbb{C}[H] = \{H\}'' = C^*(\{P_i\}). \tag{6.105}$$

Proof. First, let us show that $C^*(\{P_i\}) \subseteq \{H\}''$. Let $A \in \{H\}'$ and $|\phi\rangle$ be an eigenstate of H with eigenvalue λ_i. Then $P_j |\phi\rangle = \delta_{ij} |\phi\rangle$ and:

$$\begin{aligned} [H, A] |\phi\rangle = HA |\phi\rangle - AH |\phi\rangle \;\Rightarrow\; H(A |\phi\rangle) = h_i(A |\phi\rangle) \\ \Rightarrow\; P_j A |\phi\rangle = \delta_{ij} A |\phi\rangle \,. \end{aligned} \tag{6.106}$$

Now let $|\psi\rangle$ be any state. It can be decomposed in terms of non normalised eigenvalues of H as:

$$|\psi\rangle = |\psi_i\rangle + \sum_{i \neq j} |\psi_j\rangle \,. \tag{6.107}$$

Finally, we can see that

$$\begin{aligned} P_i A |\psi\rangle &= P_i A \left(|\psi_i\rangle + \sum_{i \neq j} |\psi_j\rangle \right) = P_i A |\psi_i\rangle + \sum_{i \neq j} P_i A |\psi_j\rangle \\ &= A |\psi_i\rangle = A P_i |\psi_i\rangle = A P_i \left(|\psi_i\rangle + \sum_{j \neq i} |\psi_j\rangle \right) = A P_i |\psi\rangle \,. \end{aligned} \tag{6.108}$$

Therefore

$$[A, P_i] = 0 \quad \Rightarrow \quad \{P_i\} \subseteq \{H\}''\,. \tag{6.109}$$

To see that $\{H\}'' = \mathbb{C}[H]$ consider that, clearly, $\{H\}' = \mathbb{C}[H]'$, which also implies $\{H\}'' = \mathbb{C}[H]''$. But because $\mathbb{C}[H]$ is a unital algebra, $\mathbb{C}[H]'' = \mathbb{C}[H]$ by Theorem 2.5.

$\mathbb{C}[H] = C^*(\{P_i\})$ follows form a similar reasoning. We have seen earlier in the proof that $\{P_i\} \subseteq \{H\}'' = \mathbb{C}[H]$. Because $\mathbb{C}[H]$ can be generated by $\{P_i\}$ and is a unital algebra, $C^*(\{P_i\}) = \mathbb{C}[H]$. □

Since our condition for cost zero of the identity is linear, we are interested in expressing $\mathcal{C}'_H$ as a vector space. In the following proposition we show the basis of the algebra that we will later see is equal to $\mathcal{C}_H$.

Proposition 6.18
The algebra generated by $\{P_i \otimes \mathbb{1}\ \forall i \in I, \mathcal{S}\}$ *is a vector space generated by the orthogonal basis*

$$\mathcal{B} = \{P_i \otimes P_j\ i \neq j, (P_i \otimes P_j)\mathcal{S}, (P_i \otimes P_i) - \frac{1}{\mathrm{Tr}[P_i]}(P_i \otimes P_i)\mathcal{S}\}. \tag{6.110}$$

Note that if P_i is of rank 1 for a given i, then $(P_i \otimes P_i) - \frac{1}{\mathrm{Tr}[P_i]}(P_i \otimes P_i)S = 0$. Therefore the dimension of this vector space is $dim_{\mathbb{C}} = 2|I|^2 - k$, where k is the number of non-degenerate subspaces.

Proof. We can use Lemma 2.6 to show this sets are equal. Also, instead of $\mathcal{B}$ we use

$$\mathcal{B}' = \{P_i \otimes P_j, (P_i \otimes P_j)\mathcal{S}\}\,, \tag{6.111}$$

which generates the same vector space. It is not orthogonal, but it does not matter for the proof.

For i), note that $P_i \otimes P_j = (P_i \otimes \mathbb{1})\mathcal{S}(P_j \otimes \mathbb{1})\mathcal{S}$ and $(P_i \otimes P_j)\mathcal{S} = (P_i \otimes \mathbb{1})\mathcal{S}(P_j \otimes \mathbb{1})$. On the left hand side of the equation we have the elements of $\mathcal{B}'$, and in the right hand side we have products of elements in gen $\mathscr{A}$. Thus $\mathcal{B}' \subseteq \mathscr{A}$.

For ii), we calculate all the possibilities:

- $(P_i \otimes P_j)(P_k \otimes P_l) = \delta_{ik}\delta_{jl}(P_i \otimes P_j) \in \mathcal{B}' \cup \{0\} \subseteq V$.
- $(P_i \otimes P_j)(P_k \otimes P_l)\mathcal{S} = \delta_{ik}\delta_{jl}(P_i \otimes P_j)\mathcal{S} \in \mathcal{B}' \cup \{0\} \subseteq V$.
- $(P_i \otimes P_j)\mathcal{S}(P_k \otimes P_l) = \delta_{il}\delta_{jk}(P_i \otimes P_j)\mathcal{S} \in \mathcal{B}' \cup \{0\} \subseteq V$.

- $(P_i \otimes P_j)\mathcal{S}(P_k \otimes P_l)\mathcal{S} = \delta_{il}\delta_{jk}(P_i \otimes P_j) \in \mathcal{B}' \cup \{0\} \subseteq V$

Therefore $\mathcal{B}' \cdot \mathcal{B}' \subseteq V$.

For iii), note that

$$\begin{aligned} P_i \otimes \mathbb{1} &= \sum_j P_i \otimes P_j \\ \mathcal{S} &= \left(\left(\sum_i P_i\right) \otimes \left(\sum_j P_j\right)\right)\mathcal{S} = \sum_{ij}(P_i \otimes P_j)\mathcal{S}, \end{aligned} \tag{6.112}$$

thus gen $\mathscr{A} \subseteq V$.

With this we have shown that $V = \mathscr{A}$, the proof that $\mathcal{B}$ is orthogonal is an immediate calculation.

□

Finally, we are ready to show the main theorem of this section. The theorem characterises $\mathcal{C}_H$ as an algebra and as a vector space.

Theorem 6.19
Let H be a self adjoint operator in a finite operator algebra. Let $\mathcal{C}_H$ be defined as in Eq. (6.103). Then

$$\mathcal{C}'_H = C^*(\{P_i \otimes \mathbb{1}, \mathcal{S}\}) = \langle\{P_i \otimes P_j, (P_i \otimes P_j)\mathcal{S}\}\rangle\,, \tag{6.113}$$

where $\langle S\rangle$ denotes the linear span of the set S.

Proof. $\mathcal{C}_H$ is not an algebra, since it is not closed under addition, but it is closed under multiplication and product with scalars. By Proposition 6.17 and Theorem 2.8, the algebra generated by $\mathcal{C}_H$ is the symmetric subset of the tensor product of the commutator of $\mathbb{C}[H]$ with itself. We can write this algebra, $C^*(\mathcal{C}_H)$, as

$$C^*(\mathcal{C}_H) = \mathbb{C}[H]' \otimes \mathbb{C}[H]' \cap \{x \in \mathcal{B}(\mathcal{H}^{\otimes 2}) \mid [x, \mathcal{S}] = 0\}. \tag{6.114}$$

Using Lemma 2.7, Theorem 2.8, Theorem 2.5 and Proposition 6.17 we can find this commutant:

$$\begin{aligned} \mathcal{C}'_H &= (\mathbb{C}[H]' \otimes \mathbb{C}[H]' \cap \{x \in \mathcal{B}(\mathcal{H}^{\otimes 2}) \mid [x, \mathcal{S}] = 0\})' \\ &= ((\mathbb{C}[H]' \otimes \mathbb{C}[H]')' \cup \{x \in \mathcal{B}(\mathcal{H}^{\otimes 2}) \mid [x, \mathcal{S}] = 0\}')'' \\ &= (\mathbb{C}[H]'' \otimes \mathbb{C}[H]'' \cup C^*(\{\mathcal{S}, \mathbb{1}\}))'' \\ &= C^*(\mathbb{C}[H] \otimes \mathbb{C}[H] \cup \{\mathcal{S}, \mathbb{1}\}) \\ &= C^*(\{P_i \otimes P_j, \mathcal{S}\}) = C^*(\{P_i \otimes \mathbb{1}, \mathcal{S}\}). \end{aligned} \tag{6.115}$$

The second equality in the statement of the theorem is immediate from Proposition 6.18. □

We have found the set of operators which commute with $\mathcal{C}_H$, which gives us a good basis to search for cost matrices. As mentioned in the discussion at the beginning of Section 6.6, this is a generalisation of Theorem 6.1. Before checking for properties, we show that it is indeed a generalisation of Theorem 6.1.

Example 6.20
Let $H = \mathbb{1}$. Then $\{P_i\} = \{\mathbb{1}\}$ and

$$\mathcal{C}'_I = \langle\{\mathbb{1} \otimes \mathbb{1}, (\mathbb{1} \otimes \mathbb{1})\mathcal{S}\}\rangle = \{\mathcal{S}, \mathbb{1}\}\,, \tag{6.116}$$

which is what we obtain at the beginning of the proof of Theorem 6.1.

6.6.2 Exploring the commutant

This section reports ongoing work toward a full characterization of the cost matrices that yield zero cost for the identity channel, and toward identifying those that produce a positive cost.

This section contains Proposition 6.21, which applies Theorem 5.13 to Theorem 6.19. Then we have three technical results that lead in to Corollary 6.26, the main result of this section.

Proposition 6.21
Let $\mathcal{C}'_H$ be as in Theorem 6.19. The subspace of $\mathcal{C}'_H$ that fulfils the condition for zero cost of the identity is

$$\langle\{\mathbb{P}_{ij}, P_i \otimes P_j \text{ s.t. } i \neq j\}\rangle\,, \tag{6.117}$$

with

$$\mathbb{P}_{ij} = (P_i \otimes P_j)\mathcal{S} - \frac{1}{2}\left(\mathrm{Tr}\left[P_j\right](P_i \otimes P_i) + \mathrm{Tr}\left[P_i\right](P_j \otimes P_j)\right). \tag{6.118}$$

Proof. We want to apply Theorem 5.13. We take the basis of $\mathcal{C}'_H$ in Theorem 6.19, $\{P_i \otimes P_j, (P_i \otimes P_j)\mathcal{S}\}$, and calculate $\mathrm{Tr}_B\left[\mathcal{S} \star \cdot\right]$ for each element:

$$\begin{aligned}
&\mathrm{Tr}_B\left[\mathcal{S} \star (P_i \otimes P_j)\right] = \delta_{ij} P_i,\\
&\mathrm{Tr}_B\left[\mathcal{S} \star (P_i \otimes P_j\mathcal{S})\right] = \frac{1}{2}\left(\mathrm{Tr}_B\left[\mathcal{S}^2 P_j \otimes P_i\right] + \mathrm{Tr}_B\left[\mathcal{S}^2 P_i \otimes P_j\right]\right)\\
&\quad = \frac{1}{2}\left(\mathrm{Tr}\left[P_i\right]P_j + \mathrm{Tr}\left[P_j\right]P_i\right).
\end{aligned} \tag{6.119}$$

From Theorem 5.13, the valid subspace for the cost function is the one generated by all linear combinations of elements whose corresponding sums of the terms on the right-hand side of Eq. (6.119) are equal to zero. Clearly, the contributions of $P_i \otimes P_j$ for $i \neq j$ can always be included, as each yields zero cost individually. In addition, we can suitably combine the contributions of $(P_i \otimes P_j \mathcal{S})$ with those of $P_i \otimes P_i$ and $P_j \otimes P_j$ to obtain

$$\mathbb{P}_{ij} = (P_i \otimes P_j)\mathcal{S} - \frac{1}{2}\left(\mathrm{Tr}\left[P_j\right](P_i \otimes P_i) + \mathrm{Tr}\left[P_i\right](P_j \otimes P_j)\right). \tag{6.120}$$

Finally, we can subtract $(P_i \otimes P_j \mathcal{S})$ and $(P_j \otimes P_i \mathcal{S})$, but the result is the same as $\mathbb{P}_{ij} - \mathbb{P}_{ji}$. Therefore we can express the vector space of matrices that yield cost one for the identity as

$$\langle \{\mathbb{P}_{ij}, P_i \otimes P_j \text{ s.t. } i \neq j\} \rangle. \tag{6.121}$$

□

Before presenting the lower bounds to the cost, we need the following two lemmas:

Lemma 6.22
Let $\mathcal{H}$ be a Hilbert space of finite dimension d. Let P be an orthogonal projection in $\mathcal{B}(\mathcal{H})$ different to the identity. Let $\mathcal{J} \subset \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ be the set of Jamiołkowski matrices. Then

$$\mathrm{Tr}_B\left[\mathcal{J}(\mathbb{1} \otimes P)\right] = \{\mu \in \mathcal{B}(\mathcal{H}) \,|\, 0 \leq \mu \leq \mathbb{1}\}. \tag{6.122}$$

Proof. We have an equality between two sets so we will show that they contain each other.

i) Starting with $\subseteq$. Let $J \in \mathcal{J}$. First note that because P is a projection and Tr_B fulfils the cyclic property of the trace for objects of the form $\mathbb{1} \otimes a$, $\mathrm{Tr}_B\left[J(\mathbb{1} \otimes P)\right] = \mathrm{Tr}_B\left[(\mathbb{1} \otimes P)J(\mathbb{1} \otimes P)\right]$. We need to see that given a Jamiołkowski matrix, the element $\mathrm{Tr}_B\left[J(\mathbb{1} \otimes P)\right]$ is positive and is dominated by the identity.

   To see the positivity, first consider the partial transpose:

$$((\mathbb{1} \otimes P)J(\mathbb{1} \otimes P))^{T_A} = (\mathbb{1} \otimes P)J^{T_A}(\mathbb{1} \otimes P), \tag{6.123}$$

   which is positive because J^{T_A} is. Thus its partial trace is positive and since $(\mathrm{Tr}_B\, M)^T = \mathrm{Tr}_B\, M^{T_A}$, we have that

$$\left(\mathrm{Tr}_B\left[(\mathbb{1} \otimes P)J(\mathbb{1} \otimes P)\right]\right)^T = \mathrm{Tr}_B\left[(\mathbb{1} \otimes P)J^{T_A}(\mathbb{1} \otimes P)\right] \geq 0, \tag{6.124}$$

which implies that $\mathrm{Tr}_B\left[(\mathbb{1}\otimes P)J(\mathbb{1}\otimes P)\right]\geq 0$ because the trace is a positive[5] map.

For the domination by the identity, consider the object $J^{T_A}-(\mathbb{1}\otimes P)J^{T_A}(\mathbb{1}\otimes P)\geq 0$, which is positive because J^{T_A} is positive and $(\mathbb{1}\otimes P)$ is an orthogonal projection. Then

$$\begin{aligned}\mathrm{Tr}_B\left[J^{T_B}-(\mathbb{1}\otimes P^T)J^{T_B}(\mathbb{1}\otimes P^T)\right]\geq 0\\ \Rightarrow\quad \mathbb{1}\geq \mathrm{Tr}_B\left[(\mathbb{1}\otimes P)J(\mathbb{1}\otimes P)\right].\end{aligned}\tag{6.125}$$

ii) Now, to show $\supseteq$, let $0\leq\mu\leq\mathbb{1}$ and $0\leq\tau=\mathbb{1}-\mu\leq\mathbb{1}$. Let $P^\perp=\mathbb{1}-P$ and $k=\mathrm{Tr}\left[P\right]$. Finally, define the following Jamiołkowski matrix:

$$J_\tau=\frac{1}{k}(\mathbb{1}\otimes P)+\frac{1}{d-k}(\tau\otimes P^\perp)-\frac{1}{k}(\tau\otimes P).\tag{6.126}$$

Note that this is a Jamiołkowski matrix because it is partial transpose positive, since $\mathbb{1}\otimes P^T-\tau\otimes P^T=(\mathbb{1}-\tau)\otimes P^T=\mu\otimes P^T\geq 0$ and $\tau\otimes {P^\perp}^T\geq 0$; and $\mathrm{Tr}_B\left[J_\tau\right]=\mathbb{1}$. Now we compute the partial trace of the projection:

$$\begin{aligned}\mathrm{Tr}_B\left[J_\tau(\mathbb{1}\otimes P)\right]&=\mathrm{Tr}_B\left[\left(\frac{1}{k}(\mathbb{1}\otimes P)+\frac{1}{d-k}(\tau\otimes P^\perp)-\frac{1}{k}(\tau\otimes P)\right)(\mathbb{1}\otimes P)\right]\\ &=\mathrm{Tr}_B\left[\frac{1}{k}(\mathbb{1}\otimes P)-\frac{1}{k}(\tau\otimes P)\right]=\mathbb{1}-\tau=\mu.\end{aligned}\tag{6.127}$$

□

Before writing the final lemma, we point out the following remark.

Remark 6.23
Let P be an orthogonal projection on a a finite dimension Hilbert space and H a hermitian operator. Then there exists a basis such that

$$P=\begin{pmatrix}\mathbb{1}&0\\0&0\end{pmatrix},\quad H=\begin{pmatrix}A&C\\C^*&B\end{pmatrix},\tag{6.128}$$

with A and B Hermitian. Moreover, $C=0\Leftrightarrow[H,P]=0$. The Jordan product of P and H written in this basis is

$$H\star P=\begin{pmatrix}A&\frac{C}{2}\\\frac{C^*}{2}&0\end{pmatrix}.\tag{6.129}$$

[5] But famously not completely positive [Per96].

Lemma 6.24
Let ρ be a state on a finite dimensional Hilbert space $\mathcal{H}$ and P an orthogonal projection. Then

$$\operatorname{Tr}\left[(\rho \star P)^{-}\right] \leq \frac{1}{8}, \tag{6.130}$$

where for an Hermitian A, $A^{\pm}$ denotes its projections onto its positive/negative eigenspaces.

Proof. Let $|\varphi\rangle$ be a pure state on a finite dimensional Hilbert space and P a rank k orthogonal projection. If $|\varphi\rangle\langle\varphi|$ is not orthogonal to P (else it is trivial), then we can define a pure state $|p\rangle = \frac{P|\varphi\rangle}{\|P|\varphi\rangle\|}$ and an orthonormal basis $\{|\alpha_i\rangle\}$ that includes $|p\rangle$ such that

$$|\varphi\rangle = a\,|p\rangle + b\,|\alpha_0\rangle \quad \text{and} \quad P = |p\rangle\langle p| + \sum_{i=1}^{k-1} |\alpha_i\rangle\langle\alpha_i|\,, \tag{6.131}$$

where, without loss of generality, a is real. Moreover, in this basis

$$|\varphi\rangle\langle\varphi| \star P = a^2\,|p\rangle\langle p| + \frac{1}{2}ab\,|\alpha_0\rangle\langle p| + \frac{1}{2}a\bar{b}\,|p\rangle\langle\alpha_0|\,, \tag{6.132}$$

which reduces any case to the 2 dimensional one. Moreover, in the 2 dimensional case the smallest (and only negative) eigenvalue of $|\varphi\rangle\langle\varphi| \star P$ is $\geq -1/8$, when the angle between the state and the projector is $\theta = 2\pi/3$ in the Bloch sphere. Finally, we can show that a convex combination of hermitian matrices has non increasing negative eigenspace. Let H_1, H_2 be hermitian and decomposed in 2 positive operators as $H_i = H_i^+ - H_i^-$. Then $(pH_1 + (1-p)H_2)^- \leq pH_1^- + (1-p)H_2^-$. To see this last inequality, note that there exists an orthogonal projection Π such that $(pH_1 + (1-p)H_2)^- = -\Pi\,(pH_1 + (1-p)H_2)\,\Pi$. then

$$\begin{aligned}
(pH_1 + (1-p)H_2)^- &= -\Pi\,(pH_1 + (1-p)H_2)\,\Pi \\
&= -\Pi\left(pH_1^+ + (1-p)H_2^+\right)\Pi \\
&\quad + \Pi\left(pH_1^- + (1-p)H_2^-\right)\Pi \\
&\leq \Pi\left(pH_1^- + (1-p)H_2^-\right)\Pi \\
&\leq pH_1^- + (1-p)H_2^-.
\end{aligned} \tag{6.133}$$

With this we can show that

$$\begin{aligned}
\operatorname{Tr}\left[(pH_1 + (1-p)H_2)^-\right] &\leq p\operatorname{Tr}\left[H_1^-\right] + (1-p)\operatorname{Tr}\left[H_2^-\right] \\
&\leq \max_{i\in\{1,2\}} \operatorname{Tr}\left[H_i^-\right],
\end{aligned} \tag{6.134}$$

which completes the proof. □

Corollary 6.25
Let $\mathcal{H}$ be a Hilbert space of dimension d. Let P_A, P_B be orthogonal projections in $\mathcal{B}(\mathcal{H})$ different to the identity. Let $\mathcal{J} \subset \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ be the set of Jamiołkowski matrices. Then,

$$\operatorname{Tr}_B\left[\mathcal{J} \star (P_A \otimes P_B)\right] = \{\mu \in \mathcal{B}(\mathcal{H}) \,|\, 0 \leq \mu \leq \mathbb{1}\} \star P_A. \tag{6.135}$$

Proof. The proof is clear from the properties of the Jordan product, the cyclic property of the trace and Lemma 6.22. □

The following corollary is the main result of this section. This corollary gives us a lower bound for the cost associated to the elements $P_i \otimes P_j$.

Corollary 6.26
Let $\mathcal{H}$ be a Hilbert space of dimension d. Let P_A, P_B be orthogonal projections in $\mathcal{B}(\mathcal{H})$ different to the identity. Let $J \in \mathcal{J} \subset \mathcal{B}(\mathcal{H} \otimes \mathcal{H})$ be a Jamiołkowski matrix and $\rho \in \mathcal{S}(\mathcal{H})$. Then

$$1 \geq \operatorname{Tr}\left[(\rho \star J)(P_A \otimes P_B)\right] \geq -\frac{1}{8}. \tag{6.136}$$

Moreover, these bounds are tight.

Proof. The upper bound is trivial. Let us show the lower bound. We use the properties of the Jordan product, Lemma 6.22 Lemma 6.24 and Corollary 6.25. Recall that $\operatorname{Tr}\left[(A \star B)C\right] = \operatorname{Tr}\left[A(B \star C)\right]$, then

$$\begin{aligned}
\operatorname{Tr}\left[(\rho \star J)(P_A \otimes P_B)\right] &= \operatorname{Tr}\left[\rho\,(J \star P_A \otimes P_B)\right] \\
&= \operatorname{Tr}\left[\rho \operatorname{Tr}_B\left[J \star P_A \otimes P_B\right]\right] \\
&= \operatorname{Tr}\left[\rho\,(P_A \star \mu)\right] = \operatorname{Tr}\left[(\rho \star P_A)\,\mu\right] \\
&= \operatorname{Tr}\left[(\rho \star P_A)^+ \mu\right] - \operatorname{Tr}\left[(\rho \star P_A)^- \mu\right] \\
&\geq -\operatorname{Tr}\left[(\rho \star P_A)^- \mu\right] \geq -\operatorname{Tr}\left[(\rho \star P_A)^- \mathbb{1}\right] \\
&\geq -\frac{1}{8}.
\end{aligned} \tag{6.137}$$

To show it is tight, choose J and P_B such that $\mu = \Pi_{\operatorname{supp}(\rho \star P_A)^-}$. Then, the first two inequalities in the previous derivation are tight. The proof of Lemma 6.24 contains an example of ρ and P_A such that the last inequality is tight, thus ensuring the tightness of the overall inequality.

To show the upper bound is tight, choose $\rho(\varepsilon) = (1-\varepsilon) \star P_A / \operatorname{Tr}\left[P_A\right] + \varepsilon(\mathbb{1} - P_A)/\operatorname{Tr}\left[\mathbb{1} - P_A\right]$, for $\varepsilon \geq 0$ and J and P_B such that $\mu = P_A$. Then

$\mathrm{Tr}\left[(\rho(\varepsilon) \star J)(P_A \otimes P_B)\right] \to 1$ when $\varepsilon \to 0$. We showed the limit version to demonstrate that this limit can be arbitrarily approximated with full rank states ρ. □

We have bounded the cost associated to $P_i \otimes P_j$. There is another set of elements, $\mathbb{P}_{ij}$ that need to be bounded. Moreover, it is possible that both elements have negative lower bounds, which would require us to further study potential convex combinations that yield positive costs. It would also be interesting to incorporate the eigenvalues of H, or rather the differences of the eigenvalues, into the cost matrix but it is unclear how to do it at the current moment.

7

Single-letter chain rules for the quantum relative entropy

We have already introduced the relative entropy in Section 2.4. This chapter works on quantum generalisations of the classical chain rule for the KL divergence[1]:

$$D(p\|q) - D(Mp\|Nq) \geq -\mathbb{E}_p\left[D(M\delta_j\|N\delta_j)\right], \tag{7.1}$$

where M, N are stochastic maps. This inequality decomposes the global decrease of relative entropy into an average of local divergences of the point distributions. It provides a finer description of how distinguishability flows through channels and underpins structural properties such as joint convexity of relative entropy.

The classical chain rule motivates analogous results in the quantum setting, which naturally lead to the study of recoverability. Petz introduced the concept of a recovery map—see Section 2.4 or [Pet86; Pet88]—that, given a quantum channel, attempts to reverse its effect for a given pair of states. Fawzi and Renner [FR15] gave the first quantitative refinement of Petz's result, proving that the conditional mutual information of a tripartite state lower bounds the fidelity of recovery:

$$I(A:C|B)_\rho \;\geq\; -2\log F(\rho_{ABC}, \mathcal{R}_{B\to BC}(\rho_{AB}))\,. \tag{7.2}$$

[1] We provide a proof of this chain rule in Proposition 7.1.

with $\mathcal{R}_{B\to BC}$ a recovery map depending only on ρ_{BC}. This result demonstrates that small conditional mutual information guarantees the existence of a recovery channel that restores the global state with high fidelity. Later, [BHO+15] strengthened this result by bounding the measured relative entropy instead of the fidelity.

These ideas were then extended to the quantum relative entropy. Wilde [Wil15] and Sutter *et al.* [STH16] developed inequalities bounding the difference between a relative entropy and its channel-processed version using rotated Petz maps:

$$\mathcal{R}^{(t)}_{\sigma,\mathcal{M}}(X) = \sigma^{\frac{1+it}{2}}\,\mathcal{M}^{\dagger}\Big(\mathcal{M}(\sigma)^{-\frac{1+it}{2}}\,X\,\mathcal{M}(\sigma)^{-\frac{1-it}{2}}\Big)\,\sigma^{\frac{1-it}{2}}. \tag{7.3}$$

Following the steps of [Wil15], [STH16] uses this map to obtain:

$$\begin{aligned} D(\rho\|\sigma) - D(\mathcal{M}(\rho)|\mathcal{M}(\sigma)) &\geq \sup_{M} D_M(\rho\|\mathcal{R}_{\sigma,\mathcal{M}}\circ\mathcal{M}(\rho)) \\ &\geq -2\log F(\rho,\ \mathcal{R}_{\sigma,\mathcal{M}}\circ\mathcal{M}(\rho))\,. \end{aligned} \tag{7.4}$$

In this result the recovery map $\mathcal{R}_{\sigma,\mathcal{M}}$ is a ρ-dependent convex combination of rotated Petz maps. This naturally raises the question of whether a universal recovery map, independent of ρ and depending only on σ, exists.

Junge *et al.* [JRS+18] and Sutter *et al.* [SBT16] answered this positively, by constructing the universal recovery channel,

$$\mathcal{R}_{\sigma,\mathcal{M}} = \int_{-\infty}^{\infty} d\beta_0\,(t)\,\mathcal{R}^{(t)}_{\sigma,\mathcal{M}}, \tag{7.5}$$

where

$$\beta_0(t) = \frac{\pi}{2}\,(\cosh \pi t + 1)^{-1}\,. \tag{7.6}$$

In addition, [JRS+18; SBT16] show that this map satisfies Eq. (7.4). These results quantitatively relate entropy loss to recoverability for arbitrary channels.

Notably, it is known that the measured entropy in Eq. (7.4) cannot be in general replaced by the quantum relative entropy: explicit counterexamples confirm the necessity of the measured relative entropy in such bounds [FF18; Hir18]. This fact marks a sharp departure from the classical case, where exact decompositions are available.

Despite these advances, there is no state-level quantum analogue of the classical chain inequality in Eq. (7.1). In the quantum setting, noncommutativity obstructs a decomposition of relative entropy into local divergences of point distributions. Fang *et al.* [FFR+20] proved that for channels $\mathcal{M},\ \mathcal{N}$ and states $\rho,\ \sigma$, the naive inequality with the single-letter channel divergence

$$D(\mathcal{M}\|\mathcal{N}) = \sup_{\rho_{RA}} D\Big((\mathrm{id}_R\otimes\mathcal{M})(\rho_{RA})\,\Big\|\,(\mathrm{id}_R\otimes\mathcal{N})(\rho_{RA})\Big) \tag{7.7}$$

does not hold. They furthermore showed that a chain rule can be recovered in the many-copy setting by introducing the regularized channel relative entropy

$$D^{\text{reg}}(\mathcal{M}\|\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} D(\mathcal{M}^{\otimes n}\|\mathcal{N}^{\otimes n}). \tag{7.8}$$

With this regularization, they established the chain rule

$$D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \;\leq\; D(\rho\|\sigma) + D^{\text{reg}}(\mathcal{M}\|\mathcal{N}). \tag{7.9}$$

Thus, in contrast to the classical case, a quantum chain rule exists only in an asymptotic, many-copy sense, and it is precisely the regularization that guarantees its validity.

This chapter is organized as follows. In Section 7.1, we present the classical chain rule for the KL divergence; in Section 7.2 we establish a complementary form of chain rule that holds already at the single-copy level, with a proof centred around the LS state over time, and in Section 7.3 we provide a generic entropy inequality and, as a corollary, a conditional chain rule.

7.1 Classical result

For the sake of completeness, in this section we show a simple proof of the classical chain rule using Jensen's Theorem [Jen06; Dur19].

Proposition 7.1 (Chain rule of the relative entropy)
Let p, q be probability distributions and M, N stochastic maps. Then

$$D(p\|q) - D(Mp\|Nq) \geq -\mathbb{E}_p\left[D(M\delta_j\|N\delta_j)\right], \tag{7.10}$$

Where δ_j is the delta probability distribution at point j.

Proof. Start by considering two classical probabilities on finite dimension sets and two stochastic maps acting on them. Let p, q be probability distributions and M, N stochastic maps. We consider the output probability distributions

$$\tilde{p}_i = (Mq)_i = \sum_j M_{ij}p_j, \quad \tilde{q}_i = (Nq)_i = \sum_j N_{ij}q_j. \tag{7.11}$$

With this definition we can find the following identity. Consider the quantity

$$\exp\left\{-\log\frac{M_{ij}p_j}{N_{ij}q_j} + \log\frac{\tilde{p}_i}{\tilde{q}_i}\right\}. \tag{7.12}$$

We calculate the average of this quantity over $M_{ij}p_j$:

$$\begin{aligned}\mathbb{E}_{Mp}\text{exp}\left\{-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right\} &= \sum_{ij} M_{ij}p_j\exp\left\{-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right\}\\ &= \sum_{ij} M_{ij}p_j\frac{\tilde{p}_i}{\tilde{q}_i}\frac{N_{ij}q_j}{M_{ij}p_j} = \sum_{ij}\frac{\tilde{q}_i}{\tilde{q}_i}\tilde{p}_i = 1.\end{aligned} \tag{7.13}$$

With this we acquire the following identity for these processes

$$\mathbb{E}_{Mp}\text{exp}\left\{-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right\} = 1. \tag{7.14}$$

From this identity and using Jensen's inequality we can obtain a classical entropy inequality.

From Eq. (7.14) we apply Jensen's inequality:

$$1 = \mathbb{E}_{Mp}\text{exp}\left\{-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right\} \geq \exp\left\{\mathbb{E}_{Mp}-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right\}. \tag{7.15}$$

Note that since M is a stochastic map, for a fixed j M_{ij} is a probability distribution on i and therefore $\sum_i M_{ij} = 1$. With this we can calculate the result

$$\begin{aligned}0 &\geq \mathbb{E}_{Mp}\left(-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right) = \sum_{ij} M_{ij}p_j\left(-\log\frac{M_{ij}p_j}{N_{ij}q_j}+\log\frac{\tilde{p}_i}{\tilde{q}_i}\right)\\ &= -\sum_{ij} M_{ij}p_j\left[\log\frac{M_{ij}}{N_{ij}}+\log\frac{p_j}{q_j}\right]+\sum_i \tilde{p}_i\log\frac{\tilde{p}_i}{\tilde{q}_i}\\ &= -\sum_j p_j\left(\sum_i M_{ij}\log\frac{M_{ij}}{N_{ij}}\right)-\sum_j\left(\sum_i M_{ij}\right)p_j\log\frac{p_j}{q_j}+D(Mp\|Nq)\\ &= -\mathbb{E}_p D(M\delta_j\|N\delta_j)-D(p\|q)+D(Mp\|Nq).\end{aligned} \tag{7.16}$$

□

7.2 Chain rule from states over time

In this section we obtain our first chain rule. We first recall the LS state over time and new derived objects we use in the proof. Then we show the main result and some remarks. Finally we show some applications.

7.2.1 Ensemble partitions from the LS state over time

The proof of our main result in this section relies on the use of the LS state over time defined in Section 4.2. Moreover, the result will be given as a function of the *ensemble partitions* of a state given a POVM, which we now introduce.

We begin by defining ensemble partitions. For any state τ and POVM $G = \{G_j\}$, let

$$\tau_j = \frac{\sqrt{\tau}\, G_j^{T_\tau} \sqrt{\tau}}{\mathrm{Tr}[G_j \tau]}, \qquad P_\tau^G(j) = \mathrm{Tr}[G_j \tau], \tag{7.17}$$

where T_τ denotes the transpose in a eigenbasis of τ and $P_\tau^G(j) = \mathrm{Tr}\,[G_j \tau]$ is the probability distribution associated with measuring G on τ.

Given a state τ and a quantum channel ε, we also consider the LS state over time

$$\omega_\tau^\varepsilon = (\mathrm{id} \otimes \varepsilon)(\|\tau\rangle\!\rangle\langle\!\langle\tau\|), \tag{7.18}$$

where $\|\tau\rangle\!\rangle = \sum_k \sqrt{\tau_k}\, |\tau_k, \tau_k\rangle$ is the canonical purification of τ.

The ensemble partition was first introduced by Fuchs [Fuc01] in the study of quantum conditional probabilities and later developed by Leifer and collaborators [Lei06; LS13] as part of a program to generalize Bayes' theorem to the quantum setting. Later these efforts would develop into the field of states over time, as we have seen in Chapter 4. In this framework, the unnormalised τ_j are understood as ensemble partitions of τ, providing the analogue of the Bayesian update rule for classical distributions under measurement. Building on these constructions, we obtain Theorem 7.2.

The proof of our main result, Theorem 7.2, starts from the LS states over time $\omega_\rho^{\mathcal{M}}$ and $\omega_\sigma^{\mathcal{N}}$ and performs a bipartite measurement to obtain classical probability distributions. These distributions can naturally be interpreted as measurements on $\mathcal{M}(\rho_i)$ and $\mathcal{N}(\sigma_i)$. The result then follows from using the classical chain rule and removing the measurements with Uhlmann's inequality and an extension to many copies. In the end we obtain a fully quantum, single-letter equation.

7.2.2 Main result

The main result of this section, Theorem 7.2, establishes a chain rule based on partitions of states, rather than basis elements as in Corollary 7.13. The remainder of the section explores consequences of this theorem, including an alternative proof of the data processing inequality and a semi-classical chain rule (see Corollary 7.8).

Theorem 7.2
Let ρ, σ be quantum states on a finite dimensional Hilbert space $\mathcal{H}_A$. Let $G = \{G_j\}$ be a POVM. Let $\mathcal{M}, \mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be completely positive trace preserving maps. Then

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -\mathbb{E}_{P_\rho^G} D(\mathcal{M}(\rho_j)\|\mathcal{N}(\sigma_j)). \tag{7.19}$$

Proof. We first construct classical states and maps from our quantum objects. Consider first a POVM $F = \{F_i\}$. We consider the following probability distributions:

$$p_j = P_\rho^G(j) = \mathrm{Tr}\left[G_j\rho\right], \quad q_j = P_\sigma^G(j) = \mathrm{Tr}\left[G_j\sigma\right]; \tag{7.20}$$

and classical stochastic maps

$$M_{ij} = \frac{\mathrm{Tr}\left[(G_j \otimes F_i)\omega_\rho^{\mathcal{M}}\right]}{\mathrm{Tr}\left[G_j\rho\right]} = \mathrm{Tr}\left[F_i\mathcal{M}(\rho_j)\right] = P^F_{\mathcal{M}(\rho_j)}(i), \tag{7.21}$$

$$N_{ij} = \frac{\mathrm{Tr}\left[(G_j \otimes F_i)\omega_\sigma^{\mathcal{N}}\right]}{\mathrm{Tr}\left[G_j\sigma\right]} = \mathrm{Tr}\left[F_i\mathcal{N}(\sigma_j)\right] = P^F_{\mathcal{N}(\sigma_j)}(i). \tag{7.22}$$

The output distributions then will be

$$\begin{aligned} \tilde{p}_i &= \sum_j M_{ij}p_j = \mathrm{Tr}\left[F_i\mathcal{M}(\rho)\right] = P^F_{\mathcal{M}(\rho)}(i), \\ \tilde{q}_i &= \sum_j N_{ij}q_j = \mathrm{Tr}\left[F_i\mathcal{N}(\sigma)\right] = P^F_{\mathcal{N}(\sigma)}(i). \end{aligned} \tag{7.23}$$

We plug these objects into Proposition 7.1:

$$D(P^F_{\mathcal{M}(\rho)}\|P^F_{\mathcal{N}(\sigma)}) - D(P^G_\rho\|P^G_\sigma) \leq \mathbb{E}_{P^G_\rho} D(P^F_{\mathcal{M}(\rho_j)}\|P^F_{\mathcal{N}(\sigma_j)}), \tag{7.24}$$

which in Hayashi's [Hay01] notation for the measured relative entropy is

$$D_F(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) - D_G(\rho\|\sigma) \leq \mathbb{E}_{P^G_\rho} D_F(\mathcal{M}(\rho_j)\|\mathcal{N}(\sigma_j)). \tag{7.25}$$

We can use Uhlmann inequality [Uhl77] to remove two of the measurements:

$$D_F(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) - D(\rho\|\sigma) \leq \mathbb{E}_{P^G_\rho} D(\mathcal{M}(\rho_j)\|\mathcal{N}(\sigma_j)). \tag{7.26}$$

We will now use [HP91, Theorem 2.3] to remove the last measurement. Consider n copies of the same system. We will apply (with a small abuse of notation) Eq. (7.26) to $\rho^{\otimes n}$, $\sigma^{\otimes n}$, $G^{\otimes n}$, $\mathcal{M}^{\otimes n}$ and $\mathcal{N}^{\otimes n}$:

$$\begin{aligned}\frac{1}{n}D_F(\mathcal{M}(\rho)^{\otimes n}\|\mathcal{N}(\sigma)^{\otimes n}) - \frac{1}{n}D(\rho^{\otimes n}\|\sigma^{\otimes n})&\\ &\leq \mathbb{E}_{P^{G^{\otimes n}}_{\rho^{\otimes n}}}\frac{1}{n}D\left(\bigotimes_{k=1}^{n}\mathcal{M}(\rho_{j_k})\middle\|\bigotimes_{k=1}^{n}\mathcal{N}(\sigma_{j_k})\right)\\ &\leq \frac{1}{n}\mathbb{E}_{P^{G^{\otimes n}}_{\rho^{\otimes n}}}\sum_{k=1}^{n}D(\mathcal{M}(\rho_{j_k})\|\mathcal{N}(\sigma_{j_k}))\\ &= \mathbb{E}_{P^G_\rho}D(\mathcal{M}(\rho_j)\|\mathcal{N}(\sigma_j)).\end{aligned} \tag{7.27}$$

Note that $D(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD(\rho\|\sigma)$. We take the limit $n\to\infty$, then by [HP91, Theorem 2.3] $\frac{1}{n}D_F(\mathcal{M}(\rho)^{\otimes n}\|\mathcal{N}(\sigma)^{\otimes n}) \to D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma))$ and we obtain the result. □

Remark 7.3
An alternative definition of the state over time is the one given by Leifer and Spekkens [Lei06; LS13], as seen in Section 4.2.1:

$$\omega^{\varepsilon}_{\tau} = (\tau^{\frac{1}{2}}\otimes \mathbb{1})(\mathbb{1}\otimes\varepsilon)(|\Phi^+\rangle\langle\Phi^+|)(\tau^{\frac{1}{2}}\otimes\mathbb{1}). \tag{7.28}$$

where $|\Phi^+\rangle$ denotes the maximally entangled state in the canonical basis. With this definition, the conditional states take the form

$$\tau_j = \frac{(\sqrt{\tau}G_j\sqrt{\tau})^T}{\mathrm{Tr}\left[G_j\tau\right]}, \tag{7.29}$$

where the transpose is taken in the basis associated with $|\Phi^+\rangle$ (usually the canonical basis). Substituting these expressions into Theorem 7.2 leads to a simplified right-hand side of the inequality.

The inequality in Eq. (7.19) does not in general reduce to the standard data processing inequality when $\mathcal{M}=\mathcal{N}$. However, if the input states commute, i.e., $[\rho,\sigma]=0$, one can choose a suitable measurement such that Eq. (7.19) reduces to the data processing inequality when $\mathcal{M}=\mathcal{N}$:

Example 7.4
Let ρ, σ be commuting states on some Hilbert space $\mathcal{H}_A$ with common basis $\{\Pi_j\}$ and $\mathcal{M},\mathcal{N}:\mathcal{B}(\mathcal{H}_A)\to\mathcal{B}(\mathcal{H}_B)$ be CPTP maps. In the common basis, we write

$$\rho = \sum_j p_j\Pi_j,\quad \sigma=\sum_j s_j\Pi_j. \tag{7.30}$$

Let G be the projective measurement on the common basis of ρ, σ, we can write ρ_j, σ_j in Theorem 7.2 as

$$\rho_j = \frac{(\sqrt{\rho}\Pi_j\sqrt{\rho})^T}{\mathrm{Tr}\,[\Pi_j\rho]} = \frac{p_j\Pi_j^T}{p_j} = \Pi_j^T, \tag{7.31}$$

and similarly $\sigma_j = \Pi_j^T$. Moreover, $P_\rho^G(j) = p_j$. Therefore Eq. (7.19) becomes

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -\mathbb{E}_p D(\mathcal{M}(\Pi_j^T)\|\mathcal{N}(\Pi_j^T)). \tag{7.32}$$

Finally, because the transpose is taken in the eigenbasis of ρ, in this case the canonical basis, each projector is self-transpose, leaving the final equation as:

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -\mathbb{E}_p D(\mathcal{M}(\Pi_j)\|\mathcal{N}(\Pi_j)). \tag{7.33}$$

Eq. (7.33) has the nice property that the right hand side relative entropies are independent of ρ, σ. This means, for example, that the equation reduces to Uhlmann's inequality when $\mathcal{M} = \mathcal{N}$.

7.2.3 Applications

We can derive two consequences from Theorem 7.2. First, we can use to show properties of the relative entropy: joint convexity, strong subadditivity and monotonicity[2]. We can also derive an extension to joint convexity. Finally, we obtain a chain rule with basis elements in the right hand side, instead of partition states like in Eq. (7.19) and fully general, unlike Eq. (7.33).

Properties of the relative entropy

Corollary 7.5
Example 7.4 with $\rho = \sigma$ is equivalent to the joint convexity of the relative entropy.

Proof. (Example $\Rightarrow$ joint convexity) Consider the probability distribution p_j and collections of states $\{\tau_j\}$, $\{\mu_j\}$ on a finite-dimensional Hilbert space. Let $\{\Pi_j\}_j$ be any family of orthogonal rank-one projectors with $\sum_j \Pi_j = \mathbb{1}$, and set $\rho = \sum_j p_j\Pi_j$.

Let $\mathcal{M}, \mathcal{N}$ be the classical-quantum channels

$$\mathcal{M}(X) = \sum_i \tau_i\ \mathrm{Tr}(\Pi_i X), \qquad \mathcal{N}(X) = \sum_i \mu_i\ \mathrm{Tr}(\Pi_i X). \tag{7.34}$$

[2] Admittedly the most convoluted proof of these properties.

By Example 7.4 with $\rho = \sigma$ and given the decomposition of ρ in the basis $\{\Pi_j\}$, we have

$$-D(\mathcal{M}(\rho)\,\|\,\mathcal{N}(\rho)) \;\geq\; -\sum_j p_j\, D\big(\mathcal{M}(\Pi_j)\,\|\,\mathcal{N}(\Pi_j)\big). \tag{7.35}$$

Using (7.34), this becomes

$$-D\left(\sum_j p_j \tau_j \,\middle\|\, \sum_j p_j \mu_j\right) \;\geq\; -\sum_j p_j\, D(\tau_j \| \mu_j), \tag{7.36}$$

which is exactly joint convexity of the relative entropy.

(Joint convexity ⇒ example with $\rho = \sigma$) Conversely, assume joint convexity. Let ρ be a state with spectral decomposition $\rho = \sum_j p_j \Pi_j$ in an orthonormal eigenbasis $\{\Pi_j\}$. For arbitrary channels $\mathcal{M}$, $\mathcal{N}$, define

$$\tau_j = \mathcal{M}(\Pi_j), \qquad \mu_j = \mathcal{N}(\Pi_j). \tag{7.37}$$

Applying joint convexity to the ensembles $\{p_j, \tau_j\}$ and $\{p_j, \mu_j\}$ gives

$$D\left(\sum_j p_j \tau_j \,\middle\|\, \sum_j p_j \mu_j\right) \;\leq\; \sum_j p_j D(\tau_j \| \mu_j). \tag{7.38}$$

Since $\sum_j p_j \tau_j = \mathcal{M}(\rho)$ and $\sum_j p_j \mu_j = \mathcal{N}(\rho)$, this is precisely

$$-D(\mathcal{M}(\rho)\,\|\,\mathcal{N}(\rho)) \;\geq\; -\sum_j p_j\, D\big(\mathcal{M}(\Pi_j)\,\|\,\mathcal{N}(\Pi_j)\big), \tag{7.39}$$

which is the $\rho = \sigma$ instance of Example 7.4. Hence the two statements are equivalent. □

Remark 7.6

It was shown in [Rus07] that joint convexity is sufficient to prove the data processing inequality [Lin75] as well as strong subadditivity for quantum relative entropy [LR73]. Our proof of Theorem 7.2 relies only on Uhlmann's inequality [Uhl77], a weaker version of data processing, and Jensen's inequality[3]*. The result from Hiai and Petz [HP91] that we use also does not depend on any of the properties we obtain. Therefore Corollary 7.5 can be used to prove this properties from Theorem 7.2.*

[3] See Eq. (7.26) and Eq. (7.15).

Extension to joint convexity

The following corollary provides an extension to the joint convexity of the relative entropy for ensembles with different probability distributions.

Corollary 7.7
Let $\{\tau_i\}, \{\mu_i\} \subseteq \mathcal{S}(\mathcal{H}_B)$ be collections of states and p_i, q_i probability distributions. Then

$$D(p\|q) - D\left(\sum_j p_j \tau_j \middle\| \sum_j q_j \mu_j\right) \geq -\sum_j p_j D(\tau_j \| \mu_j). \tag{7.40}$$

Proof. Choose and arbitrary basis $\{\Pi_i\}$ (note that $\mathrm{Tr}\,[\Pi_i] = 1$ for all i) and let $\rho = \sum_i p_i \Pi_i$, $\sigma = \sum_i q_i \Pi_i$. Similarly to before, let

$$\mathcal{M}(x) = \sum_i \tau_i \,\mathrm{Tr}\,[\Pi_i x] \quad \text{and} \quad \mathcal{N}(x) = \sum_i \mu_i \,\mathrm{Tr}\,[\Pi_i x]\,. \tag{7.41}$$

Because $[\rho, \sigma] = 0$ we can use Eq. (7.33) with these objects to obtain

$$D(\rho\|\sigma) - D\left(\mathcal{M}\left(\sum_j p_j \Pi_j\right) \middle\| \mathcal{N}\left(\sum_j q_j \Pi_j\right)\right) \geq -\sum_j p_j D(\mathcal{M}(\Pi_j)\|\mathcal{N}(\Pi_j)), \tag{7.42}$$

which together with the definition and linearity of $\mathcal{M}, \mathcal{N}$ and $D(\rho\|\sigma) = D(p\|q)$[4] yields the result. □

Projector chain rule

Finally, we see our chain rule with projectors. Note that in this corollary the projectors are different for $\mathcal{M}$ and $\mathcal{N}$. Ideally they would be the same, which would yield a much cleaner result. Unfortunately, this is not true, as we see in Section 7.3.1.

Corollary 7.8
Let ρ, σ be quantum states on a finite dimensional Hilbert space $\mathcal{H}_A$; with spectra p, q and rank 1 eigenprojectors $\{\Pi_j\}, \left\{\tilde{\Pi}_j\right\}$, respectively. Let $\mathcal{M}, \mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be completely positive trace preserving maps. Then

$$D(p\|q) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -E_p D(\mathcal{M}(\Pi_i)\|\mathcal{N}(\tilde{\Pi}_i)). \tag{7.43}$$

[4] Because $[\rho, \sigma] = 0$.

Proof. Note we can write $\mathcal{N}(\sigma)$ as a function of the eigendecomposition of σ as follows:

$$\mathcal{N}(\sigma) = \sum_j q_j \mathcal{N}(\tilde{\Pi}_j). \tag{7.44}$$

Since there exists[5] a unitary U that maps the basis of σ to the basis of ρ, i.e. $\mathcal{U}(\tilde{\Pi}_i) = \Pi_j$, we can write

$$\sum_j q_j \mathcal{N}(\tilde{\Pi}_j) = \sum_j q_j \mathcal{N} \circ \mathcal{U}(\Pi_j) = \sum_j q_j \mathcal{N} \circ \mathcal{U}(\Pi_j) = \mathcal{F}(\sigma'), \tag{7.45}$$

where $\mathcal{F} = \mathcal{N} \circ \mathcal{U}$ and $\sigma' = \sum_j q_j \Pi_j$.

Note that $[\rho, \sigma'] = 0$ commute. The inequality of Theorem 7.2 applied to $\mathcal{M}$, $\mathcal{F}$, ρ, σ' and $\{\Pi_i\}$ becomes

$$D(p\|q) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -E_p[D(\mathcal{M}(\Pi_i)\|\mathcal{N}(\tilde{\Pi}_i))]. \tag{7.46}$$

□

Before ending this section we comment on the ensembles, since equivalent ensembles can yield different equations.

Remark 7.9

The results in Corollary 7.7 and Corollary 7.8 depend on an arbitrary choice of ordering for the ensembles in Corollary 7.7 and the basis in Corollary 7.8. In both cases the second term of the left hand side does not change, $D\left(\sum_j p_j \tau_j \middle\| \sum_j q_j \mu_j\right)$ *and* $D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma))$*, respectively. This is also the term we are most interested in bounding. These results could use an optimisation over the* $n!$ *possible orderings; with* n *being the number of elements of the ensemble or the dimension of the system, respectively.*

Note as well that Corollary 7.8 can be proven from Corollary 7.7. Choose $\{p_i, \tau_i\} = \{p_i, \mathcal{M}(\Pi_i)\}$ and $\{q_i, \tau_i\} = \{q_i, \mathcal{N}(\tilde{\Pi}_i)\}$ in Corollary 7.7 to immediately obtain Corollary 7.8.

[5]There are multiple non-equivalent unitaries that do that that will result in different values for $D(p\|q)$ and $E_p D(\mathcal{M}(\Pi_i)\|\mathcal{N}(\tilde{\Pi}_i))$. This is further discussed in Remark 7.9.

7.3 Conditional chain rules

This section contains entropy inequalities based around the maps

$$\mathcal{R}^{\alpha}_{\gamma,\sigma,\mathcal{M}}(X) = J^{\alpha}_{\sigma} \circ \mathcal{M}^{\dagger} \circ J^{-\alpha}_{\gamma}(X) = \sigma^{\alpha}\, \mathcal{M}^{\dagger}\big(\gamma^{-\alpha}\, X\, \gamma^{-\alpha^*}\big)\, \sigma^{\alpha^*}, \tag{7.47}$$

where $\alpha = (1 - \mathrm{i}t)/2$ and $J^{\alpha}_{\sigma}(X) = \sigma^{\alpha}\, X\, \sigma^{\alpha^*}$. In particular, we consider the following weighted sum:

$$\mathcal{R}_{\gamma,\sigma,\mathcal{M}} \int_{-\infty}^{\infty} d\beta_0\,(t)\, \mathcal{R}^{\alpha}_{\gamma,\sigma,\mathcal{M}}, \tag{7.48}$$

where

$$\beta_0\,(t) = \frac{\pi}{2}\,(\cosh \pi t + 1)^{-1}\,. \tag{7.49}$$

Note the similarity to the rotated Petz recovery channels from [Wil15; STH16; JRS+18; SBT16].

In our work, we start by considering some arbitrary states and a channel, and we obtain a general entropy inequality: Theorem 7.10. From Theorem 7.10 we can choose inputs in a clever way so that we obtain meaningful results.

Theorem 7.10
Let $\mathcal{M}, \mathcal{N}$ be quantum channels and ρ, σ, γ and ω states. Then

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\gamma) + D(\mathcal{M}(\rho)\|\omega) \geq D_{\Pi}(\rho\|\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)), \tag{7.50}$$

with

$$\mathcal{R}_{\gamma,\sigma,\mathcal{M}} = \int_{-\infty}^{+\infty} d\beta_0(t) J^{\alpha}_{\sigma} \circ \mathcal{M}^{\dagger} \circ J^{-\alpha}_{\gamma}, \tag{7.51}$$

$\beta_0(t) = \frac{\pi}{2}(\cosh(\pi t) + 1)^{-1}$ *and* $\alpha = \frac{1-\mathrm{i}t}{2}$.

In the case where $\omega = \mathcal{M}(\rho)$ this becomes

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\gamma) \geq D_{\Pi}(\rho\|\mathcal{R}_{\gamma,\sigma,\mathcal{M}} \circ \mathcal{M}(\rho)). \tag{7.52}$$

The proof of the theorem uses similar ideas to [KK19], which in turn takes from [SBT16].

Proof. Consider the quantity $\mathrm{Tr}\left[p_i^{-1}\Pi_i(J^{\alpha}_{\sigma} \circ \mathcal{M}^{\dagger} \circ J^{-\alpha}_{\gamma})(\omega)\right]$, where $\rho = \sum_i p_i \Pi_i$. Let U be the unitary that dilates the map $M(\cdot) = U \cdot \otimes \mathbb{1}_E U^{\dagger}$ and note that we will not write the tensor product with the identity on the ancillary system. Let $\alpha = \frac{1-i\theta}{2}$, then

$$\begin{aligned}
&\mathrm{Tr}\left[p_i^{-1}\Pi_i(J_\sigma^\alpha \circ \mathcal{M}^\dagger \circ J_\gamma^{-\alpha})(\omega)\right] \\
&= \mathrm{Tr}\left[\Pi_i(p_i\Pi_i)^{-1}\sigma^\alpha U^\dagger \gamma^{-\alpha} U U^\dagger \omega U U^\dagger \gamma^{-\alpha^*} U \sigma^{\alpha^*}\right] \\
&= \mathrm{Tr}\left[e^{\log \Pi_i} e^{-\log p_i \Pi_i} e^{\alpha \log \sigma} e^{-\alpha U^\dagger \log \gamma U} e^{U^\dagger \log \omega U} e^{-\alpha^* U^\dagger \log \gamma U} e^{\alpha^* \log \sigma}\right] \\
&= \left\| e^{\alpha \log \Pi_i} e^{-\alpha \log p_i \Pi_i} e^{\alpha \log \sigma} e^{-\alpha U^\dagger \log \gamma U} e^{\alpha U^\dagger \log \omega U} \right\|_2^2 \\
&= \left\| \prod_{k=1}^{5} e^{2\alpha H_k} \right\|_2^2 .
\end{aligned} \tag{7.53}$$

With $H_1 = \frac{1}{2}\log \Pi_i$, $H_2 = -\frac{1}{2}\log p_i\Pi_i$, $H_3 = \frac{1}{2}\log\sigma$, $H_4 = -\frac{1}{2}U^\dagger \log \gamma U$ and $H_5 = \frac{1}{2}U^\dagger \log \omega U$. We can now integrate over t with the weight $d\beta_0(t/2)$, take the logarithm on both sides of the equality and then average over the spectrum of ρ. The left hand side is

$$\begin{aligned}
&\sum_i p_i \log \int_{-\infty}^{\infty} d\beta_0(t)\, \mathrm{Tr}\left[p_i^{-1}\Pi_i(J_\sigma^\alpha \circ \mathcal{M}^\dagger \circ J_\gamma^{-\alpha})(\omega)\right] \\
&= \sum_i p_i \log \mathrm{Tr}\left[p_i^{-1}\Pi_i \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)\right] = \sum_i p_i \left(\log \mathrm{Tr}\left[\Pi_i \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)\right] - \log p_i\right) \\
&= -D_\Pi(\rho \| \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)),
\end{aligned} \tag{7.54}$$

where D_Π is the measured entropy on the basis of ρ.

We can now use concavity of the logarithm, the inequality from [SBT16, Corollary 3.3] and the Perels-Bogoliubov inequality [Ara75; KK19] to operate on the right hand side. The Perels-Bogoliubov inequality claims that $\mathrm{Tr}\left[e^{F+R}\right] \geq e^{\mathrm{Tr}\left[Fe^R\right]}$ For F, G self-adjoint and $\mathrm{Tr}\left[e^R\right] = 1$. We let $F = -\log\rho + \log\sigma - U^\dagger \log \gamma U + U^\dagger \log \omega U$, which is clearly self-adjoint, and $R = \log \Pi_i$, so that $e^R = \Pi_i$ has trace 1.

$$\begin{aligned}
&\sum_i p_i \log \int_{-\infty}^{\infty} d\beta_0(t) \left\| \prod_{k=1}^{5} e^{2\alpha H_k} \right\|_2^2 \\
&\geq 2 \sum_i p_i \int_{-\infty}^{\infty} d\beta_0(t) \log \left\| \prod_{k=1}^{5} e^{2\alpha H_k} \right\|_2 \\
&\geq 2 \sum_i p_i \log \left\| \exp\left(\sum_{k=1}^{5} H_k \right) \right\|_2 \\
&\geq \sum_i p_i \operatorname{Tr}\left[e^{\log \Pi_i} \left(-\log \Pi_i p_i + \log \sigma - U^\dagger \log \gamma U + U^\dagger \log \omega U \right) \right] \\
&\geq \sum_i p_i \operatorname{Tr}\left[\Pi_i \left(-\log \Pi_i p_i + \log \sigma - U^\dagger \log \gamma U \right.\right. \\
&\quad \left.\left. + U^\dagger \log \mathcal{M}(\rho) U - U^\dagger \log \mathcal{M}(\rho) U + U^\dagger \log \omega U \right) \right] \\
&= \sum_i p_i \operatorname{Tr}\left[\Pi_i \left(-\log \Pi_i p_i \right) \right] + \operatorname{Tr}\left[\sum_i p_i \Pi_i \left(\log \sigma - U^\dagger \log \gamma U \right.\right. \\
&\quad \left.\left. + U^\dagger \log \mathcal{M}(\rho) U - U^\dagger \log \mathcal{M}(\rho) U + U^\dagger \log \omega U \right) \right] \\
&= -\operatorname{Tr}\left[\rho \log \rho \right] + \operatorname{Tr}\left[\rho \left(\log \sigma - U^\dagger \log \gamma U + U^\dagger \log \mathcal{M}(\rho) U \right.\right. \\
&\quad \left.\left. - U^\dagger \log \mathcal{M}(\rho) U + U^\dagger \log \omega U \right) \right] \\
&= -D(\rho\|\sigma) + D(\mathcal{M}(\rho)\|\gamma) - D(\mathcal{M}(\rho)\|\omega).
\end{aligned} \tag{7.55}$$

Joining both sides together we obtain the result. □

We can immediately recover an improvement to the data processing inequality, as we see in the following:

Corollary 7.11
Let $\mathcal{M}, \mathcal{N}$ be quantum channels and ρ, σ states. Then

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq D_\Pi(\rho\|\mathcal{R}_{\mathcal{N}(\sigma),\sigma,\mathcal{M}}(\mathcal{M}(\rho))). \tag{7.56}$$

Proof. The result is immediate from Theorem 7.10 by letting $\gamma = \mathcal{N}(\sigma)$. □

Theorem 7.10 provides a general framework of entropy inequalities with a lot of degrees of freedom. By choosing particular cases we can find some interesting consequences. As stated at the beginning of the section, we are interested in effectively conditioning our results to when $\mathcal{R}_{\gamma,\sigma,\mathcal{M}}$ is trace non-increasing for a particular input. The following theorem provides the mathematical expression for this conditioning.

Lemma 7.12
Let $\mathcal{M}$ be a quantum channel and ρ, σ, γ states, and ω a positive semidefinite operator, and D_Π the measured relative entropy on the basis of ρ. Then

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\gamma) \geq -D(\mathcal{M}(\rho)\|\omega), \tag{7.57}$$

if $\mathrm{Tr}\left[\Pi_\rho \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)\right] \leq 1$.

Proof. Consider the setting of Theorem 7.10. We need to show that

$$T = \mathrm{Tr}\left[\Pi_\rho \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)\right] \leq 1 \quad \Rightarrow \quad D_\Pi(\rho\|\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)) \geq 0. \tag{7.58}$$

We can calculate it directly:

$$\begin{aligned} D_\Pi(\rho\|\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)) &= D_\Pi\left(\rho\middle\| T\frac{\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)}{T}\right) \\ &= D_\Pi\left(\rho\middle\|\frac{\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)}{T}\right) - \log T \geq 0. \end{aligned} \tag{7.59}$$

The first term is non-negative because it is the measured relative entropy between two states[6] and the second term is by the hypothesis. □

Observe that $\mathrm{Tr}\left[\Pi_\rho \mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)\right] \leq 1$ in Lemma 7.12 is akin to asking for the map to $\mathcal{R}_{\gamma,\sigma,\mathcal{M}}$ to yield a less than trace 1 output in the support of ρ on input ω. With this conditioning written precisely, we can state the main result of this section:

Corollary 7.13
Let $\mathcal{M}, \mathcal{N}$ be quantum channels and let $\rho = \sum_i p_i \Pi_i$, σ be states. Then

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -\mathbb{E}_p D(\mathcal{M}(\Pi_i)\|\mathcal{N}(\Pi_i)) \tag{7.60}$$

if $\mathrm{Tr}\left[\Pi_\rho \mathcal{R}_{\mathcal{N}(\sigma),\sigma,\mathcal{M}}(\mathcal{N}(\rho))\right] \leq 1$.

[6] $\mathcal{R}_{\gamma,\sigma,\mathcal{M}}(\omega)/T$ might not actually be normalised since the trace is over the support of ρ, but it is normalised when restricted to the support of ρ, which is what matters for the measured relative entropy in this basis.

Proof. Consider Lemma 7.12 and let $\omega = \mathcal{N}(\rho)$ and $\gamma = \mathcal{N}(\sigma)$. We obtain

$$\begin{aligned} D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) &\geq -D(\mathcal{M}(\rho)\|\mathcal{N}(\rho)) - \\ &= -D\left(\mathcal{M}\left(\sum_i p_i \Pi_i\right)\middle\|\mathcal{N}\left(\sum_i p_i\Pi_i\right)\right). \end{aligned} \tag{7.61}$$

We can now use the joint convexity of the quantum relative entropy to find the result:

$$\begin{aligned} D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) &\geq -D\left(\mathcal{M}\left(\sum_i p_i \Pi_i\right)\middle\|\mathcal{N}\left(\sum_i p_i\Pi_i\right)\right) \\ &\geq -\sum_i p_i D(\mathcal{M}(\Pi_i)\|\mathcal{N}(\Pi_i)). \end{aligned} \tag{7.62}$$

□

The condition $\operatorname{Tr}\left[\Pi_\rho \mathcal{R}_{\mathcal{N}(\sigma),\sigma,\mathcal{M}}(\mathcal{N}(\rho))\right] \leq 1$ is a bit weaker than one might expect a priori. The aforementioned works [JRS+18; SBT16] work with the recovery map $\mathcal{R}_{\mathcal{M}(\sigma),\sigma,\mathcal{M}}$, which is CPTP. Our two-map generalisation as it appears in Corollary 7.13, $\mathcal{R}_{\mathcal{N}(\sigma),\sigma,\mathcal{M}}$, is also CP, but in general not TP. One might expect the condition to require this map to be TP, but this in not the case. Instead, the condition requires that the map is TP for the relevant objects. Moreover, the map is still universal, since its dependencies are $\mathcal{M}$, $\mathcal{N}$ and σ, but not ρ. There is a ρ dependence in the condition, but we can relax Π_ρ to $\mathbb{1}$ to remove it.

7.3.1 Necessity of the condition in Corollary 7.13

In Corollary 7.13 we saw a sufficient condition for a chain rule. In this appendix we show an example that demonstrates that there are cases in which the chain rule is false, showing that there exists a necessary condition for the fulfilment of the chain rule. Consider the regularised version of the chain rule:

$$D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \geq -\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\rho_n}\left[D(\mathcal{M}^{\otimes n}(\Pi_k)\|\mathcal{N}^{\otimes n}(\Pi_k))\right], \tag{7.63}$$

where Π_k are the eigenprojectors of $\rho_n = \rho^{\otimes n}$. This version can be obtained by applying Corollary 7.13 to $\rho^{\otimes n}$, $\sigma^{\otimes n}$, $\mathcal{M}^{\otimes n}$ and $\mathcal{N}^{\otimes n}$, and then taking the limit $n \to \infty$. Therefore a violation of Eq. (7.63) implies a violation of Eq. (7.60).

We provide a simple class of counterexamples in Example 7.14 and a generalisation of this class in Example 7.15, which shows that the states that violate Eq. (7.63) are not some measure 0 set.

Example 7.14
We show a simple counterexample to Eq. (7.63). Let $d = 2$, $\rho = |0\rangle\langle 0|$, $\sigma = (1-\varepsilon)|+\rangle\langle+| + \varepsilon|-\rangle\langle-|$, $\mathcal{M}(x) = \mathrm{Tr}(x)|-\rangle\langle-|$ *and* $\mathcal{N} = \mathcal{E}_\sigma$, *where* $= \mathcal{E}_\sigma$ *denotes the pinching map on the basis of* σ *[Hay01]. Then*

$$\begin{aligned} D(\rho\|\sigma) &= -\langle 0|(\log(1-\varepsilon)|+\rangle\langle+| + \log\varepsilon|-\rangle\langle-|)|0\rangle \\ &= -\frac{1}{2}(\log(1-\varepsilon) + \log\varepsilon) \\ D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) &= D(|-\rangle\langle-|\,\|\sigma) \\ &= -\langle -|(\log(1-\varepsilon)|+\rangle\langle+| + \log\varepsilon|-\rangle\langle-|)|-\rangle \\ &= -\log\varepsilon \end{aligned} \tag{7.64}$$

Therefore $D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) = \frac{1}{2}(\log\varepsilon - \log(1-\varepsilon))$, *which approaches* $-\infty$ *when* ε *goes to 0. We now need to check that the right hand side of Eq. (7.63) is finite to finish the counterexample.*

Consider the projectors of $\rho_n = \rho^{\otimes n}$. *Because* ρ *is pure,* ρ_n *will have only 2 eigenvalues*[7]*, 0 and 1, with associated projectors:*

$$\Pi_1 = |0\rangle\langle 0|^{\otimes n}, \quad \Pi_0 = \mathbb{1} - |0\rangle\langle 0|^{\otimes n}. \tag{7.65}$$

We need to apply the pinching

$$\mathcal{E}_\sigma^{\otimes n}(x) = \sum_{i_1,\dots,i_n \in \{+,-\}} |i_1 \dots i_n\rangle\langle i_1 \dots i_n|\, x\, |i_1 \dots i_n\rangle\langle i_1 \dots i_n| \tag{7.66}$$

to these projectors. Because $\mathcal{M}^{\otimes n}(\Pi_k) = |-\rangle\langle-|^{\otimes n}$ *is pure and* ρ_n *has a single nonzero eigenvalue, we only care about the value of* $\langle -,\dots,-|\Pi_1|-,\dots,-\rangle$ *in*

$$\mathcal{N}^{\otimes n}(\Pi_k) = \mathcal{E}_\sigma^{\otimes n}(\Pi_k) = \sum_{i_1,\dots,i_n \in \{+,-\}} \langle i_1 \dots i_n|\Pi_k|i_1 \dots i_n\rangle\, |i_1 \dots i_n\rangle\langle i_1 \dots i_n|, \tag{7.67}$$

since the term associated to Π_1 *is the only one that matters and the diagonal element of* $\mathcal{N}^{\otimes n}(\Pi_1)$ *associated to* $|-,\dots,-\rangle\langle -,\dots,-|$ *will be the only one that survives in the calculation.*

We can calculate this value:

$$\langle -\cdots-|\Pi_1|-\cdots-\rangle = \langle -\cdots-||0\rangle\langle 0|^{\otimes n}|-\cdots-\rangle = \frac{1}{2^n}. \tag{7.68}$$

[7] This example also works with $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$, $p \neq \frac{1}{2}$, but the resulting projectors and the subsequent calculation are a bit more complicated, see Example 7.15.

Therefore $D(\mathcal{M}^{\otimes n}(\Pi_1)\|\mathcal{N}^{\otimes n}(\Pi_1)) = -\log\frac{1}{2^n} = n\log 2 = n$. *Finally, the right hand side is*

$$-\lim_{n\to\infty}\frac{1}{n}\mathbb{E}_{\rho_n}D(\mathcal{M}^{\otimes n}(\Pi_k)\|\mathcal{N}^{\otimes n}(\Pi_k)) = -\lim_{n\to\infty}\frac{1}{n}n = -1 > -\infty. \tag{7.69}$$

A quick calculation shows that $0 < \varepsilon < \frac{1}{5}$ *violates Eq.* (7.63).

In the following example we generalise by adding an angle to σ and a mixture to ρ. The resulting set can be seen in Fig. 7.1.

Example 7.15
Let $\rho = (1-p)|0\rangle\langle 0| + p|1\rangle\langle 1|$, $p \in (0, \frac{1}{2})$; $\sigma = (1-\varepsilon)|\tilde{+}\rangle\langle\tilde{+}| + \varepsilon|\tilde{-}\rangle\langle\tilde{-}|$ *with* $|\tilde{+}\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$ *and* $|\tilde{-}\rangle = \sin\frac{\theta}{2}|0\rangle - \cos\frac{\theta}{2}|1\rangle$; $\mathcal{M}(x) = \mathrm{Tr}(x)|\tilde{-}\rangle\langle\tilde{-}|$ *and* $\mathcal{N} = \mathcal{E}_\sigma$.

We will follow the same steps as in Example 7.14. First we calculate the relative entropies:

$$\begin{aligned} D(\rho\|\sigma) =& - S(p) - (1-p)\langle 0|\big(\log(1-\varepsilon)|\tilde{+}\rangle\langle\tilde{+}| + \log\varepsilon|\tilde{-}\rangle\langle\tilde{-}|\big)|0\rangle \\ & - p\langle 1|\big(\log(1-\varepsilon)|\tilde{+}\rangle\langle\tilde{+}| + \log\varepsilon|\tilde{-}\rangle\langle\tilde{-}|\big)|1\rangle \\ =& - S(p) - (1-p)\left[\cos^2\frac{\theta}{2}\log(1-\varepsilon) + \sin^2\frac{\theta}{2}\log\varepsilon\right] \\ & - p\left[\sin^2\frac{\theta}{2}\log(1-\varepsilon) + \cos^2\frac{\theta}{2}\log\varepsilon\right] \\ =& - S(p) - \log(1-\varepsilon)\left[(1-p)\cos^2\frac{\theta}{2} + p\sin^2\frac{\theta}{2}\right] \\ & - \log\varepsilon\left[(1-p)\sin^2\frac{\theta}{2} + p\cos^2\frac{\theta}{2}\right] \\ D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) =& D(|\tilde{-}\rangle\langle\tilde{-}|\,\|\sigma) = -\langle\tilde{-}|\big(\log(1-\varepsilon)|\tilde{+}\rangle\langle\tilde{+}| + \log\varepsilon|\tilde{-}\rangle\langle\tilde{-}|\big)|\tilde{-}\rangle \\ =& - \log\varepsilon \end{aligned} \tag{7.70}$$

Note that the coefficients of $\log(1-\varepsilon)$, $\log\varepsilon$ *in* $D(\rho\|\sigma)$ *add up to 1. Therefore*

$$\begin{aligned} & D(\rho\|\sigma) - D(\mathcal{M}(\rho)\|\mathcal{N}(\sigma)) \\ & = -S(p) + [\log\varepsilon - \log(1-\varepsilon)]\left[(1-p)\cos^2\frac{\theta}{2} + p\sin^2\frac{\theta}{2}\right]. \end{aligned} \tag{7.71}$$

Similarly to in Example 7.14 $[\log\varepsilon - \log(1-\varepsilon)]$ *can be infinitely negative for small* ε *and* $\left[(1-p)\cos^2\frac{\theta}{2} + p\sin^2\frac{\theta}{2}\right]$ *is a strictly positive coefficient, therefore the left hand side of Eq.* (7.63) *goes to* $-\infty$.

ρ_n will now have $n+1$ eigenspaces, each of dimension $\binom{n}{k}$, for $k \in \{0,\dots,n\}$ with eigenvalues $(1-p)^k p^{n-k}$. Let $\mathbb{X}_k = \{x \in \{0,1\}^n \text{ s.t. } |x| = k\}$, where $|x|$ is the number of ones a sequence has. The kth projector will be $\Pi_k = \sum_{x\in\mathbb{X}_k} |x\rangle\langle x|$. Similarly to Example 7.14,

$$\mathcal{N}^{\otimes n}(x) = \mathcal{E}_\sigma^{\otimes n}(x) = \sum_{i_1,\dots,i_n \in \{\tilde{+},\tilde{-}\}} |i_1\dots i_n\rangle\langle i_1\dots i_n| \, x \, |i_1\dots i_n\rangle\langle i_1\dots i_n| \tag{7.72}$$

and we are only concerned with the coefficients associated to $|\tilde{-}\rangle\langle\tilde{-}|^{\otimes n}$. This coefficients are

$$\begin{aligned}\langle\tilde{-}|^{\otimes n} \Pi_k |\tilde{-}\rangle^{\otimes n} = \langle\tilde{-}|^{\otimes n} \sum_{x\in\mathbb{X}_k} |x\rangle\langle x| \, |\tilde{-}\rangle^{\otimes n} &= \sin^{2k}\frac{\theta}{2}\cos^{2(n-k)}\frac{\theta}{2}|\mathbb{X}_k| \\ &= \binom{n}{k}\sin^{2k}\frac{\theta}{2}\cos^{2(n-k)}\frac{\theta}{2}.\end{aligned} \tag{7.73}$$

The eigenvalue associated to the kth eigenspace is $(1-p)^k p^{n-k}$. Due to $\mathrm{Tr}[\Pi_k] = \binom{n}{k}$, a coefficient $\binom{n}{k}$ appears outside the relative entropy, since $D(k\rho\|k\sigma) = kD(\rho\|\sigma)$. Therefore the expectation value in the right hand side is

$$\begin{aligned}&-\mathbb{E}_{\rho_n} D(\mathcal{M}^{\otimes n}(\Pi_k)\|\mathcal{N}^{\otimes n}(\Pi_k)) \\ &\qquad = \sum_{k=0}^{n}\binom{n}{k}(1-p)^k p^{n-k}\log\left(\sin^{2k}\frac{\theta}{2}\cos^{2(n-k)}\frac{\theta}{2}\right).\end{aligned} \tag{7.74}$$

The limit can be solved exactly. We can consider first the case $p=0$. While the discussion on the projectors is false for $p=0$ because all there is a single eigenspace and eigenvalue for $k<n$, because this eigenvalue is 0 we can still use Eq. (7.74). Only the term $k=n$ survives and it simplifies to $\log\left(\sin^{2n}\frac{\theta}{2}\right)$. The limit is then $\log\left(\sin^2\frac{\theta}{2}\right)$ We can find the values of ε and θ that violate Eq. (7.63) by setting the left hand side to be smaller than the right hand side.

$$[\log\varepsilon - \log(1-\varepsilon)]\cos^2\frac{\theta}{2} < \log\left(\sin^2\frac{\theta}{2}\right) \Leftrightarrow \varepsilon < \frac{(\sin^2\frac{\theta}{2})^{\frac{1}{\cos^2\frac{\theta}{2}}}}{1+(\sin^2\frac{\theta}{2})^{\frac{1}{\cos^2\frac{\theta}{2}}}} \tag{7.75}$$

These states are plotted in Fig. 7.1.

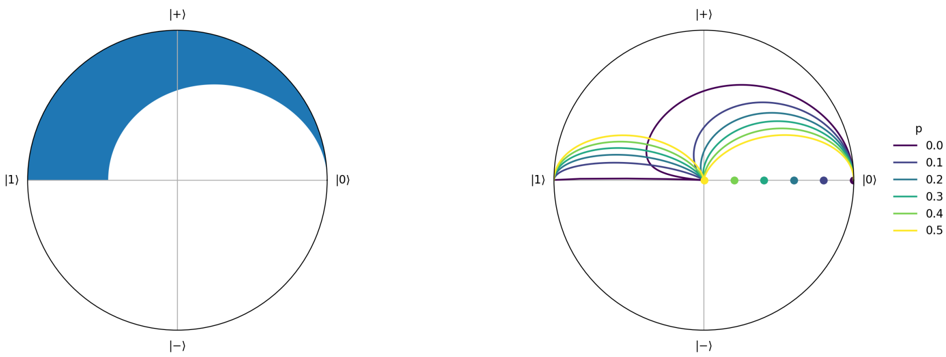

Figure 7.1: Cross section of the $x-z$ plane of the Bloch sphere. Left shows the states σ for which Eq. (7.63) is violated when $\rho = |0\rangle\langle 0|$ ($p=0$). All the boundaries of the coloured region are excluded from the set. Right shows the bound for mixed values of p. The dots represent the ρ for each value of p. Note that in the right picture the cases $p=0$ and $p=\frac{1}{2}$ are calculated for values very close to 0 and $\frac{1}{2}$ but not exactly 0 and $\frac{1}{2}$, since the bound is not continuous at $\frac{1}{2}$ and at 0 it has a very rapid change.

If we let $p \in (0, \frac{1}{2})$ *we can lower bound the right hand side. Note that*

$$\begin{aligned}
&\sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} \log\left(\sin^{2k}\frac{\theta}{2}\cos^{2(n-k)}\frac{\theta}{2}\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} \left(k\log\left(\sin^{2n}\frac{\theta}{2}\right) + (n-k)\log\left(\cos^{2n}\frac{\theta}{2}\right)\right) \\
&= n(1-p)\log\left(\sin^2\frac{\theta}{2}\right) + np\log\left(\cos^2\frac{\theta}{2}\right) = \log\left(\sin^{2n(1-p)}\frac{\theta}{2}\cos^{2np}\frac{\theta}{2}\right)
\end{aligned} \tag{7.76}$$

In the limit this will be $\log\left(\sin^{2(1-p)}\frac{\theta}{2}\cos^{2p}\frac{\theta}{2}\right)$*. Eq. (7.63) will be violated if the left* *hand side is smaller:*

$$-S(p) + [\log \varepsilon - \log(1-\varepsilon)] \left[(1-p)\cos^2\frac{\theta}{2} + p\sin^2\frac{\theta}{2}\right] < \log\left(\sin^{2(1-p)}\frac{\theta}{2}\cos^{2p}\frac{\theta}{2}\right)$$
$$\Updownarrow$$
$$\varepsilon < \frac{\exp\left\{\frac{(1-p)\log\left(\sin^2\frac{\theta}{2}\right)+p\log\left(\cos^2\frac{\theta}{2}\right)+S(p)}{\left[(1-p)\cos^2\frac{\theta}{2}+p\sin^2\frac{\theta}{2}\right]}\right\}}{1+\exp\left\{\frac{(1-p)\log\left(\sin^2\frac{\theta}{2}\right)+p\log\left(\cos^2\frac{\theta}{2}\right)+S(p)}{\left[(1-p)\cos^2\frac{\theta}{2}+p\sin^2\frac{\theta}{2}\right]}\right\}}. \tag{7.77}$$

Fig. 7.1 shows the bound for different values of p*. In the limit where* $p \to \frac{1}{2}$ *the bound simplifies to*

$$\varepsilon < \frac{\sin^2\theta}{1+\sin^2\theta}. \tag{7.78}$$

Conclusion

In this thesis, we address two distinct topics in quantum information theory, both unified by the use of states over time as a central conceptual tool. This notion, which aims to characterise quantum properties when systems evolve or are traversed across time—rather than across space as in standard quantum states—provides a natural framework to describe couplings between states and channels and to explore their informational and structural properties. Within this framework, we develop a theory of quantum optimal transport and introduce a quantum chain rule for the relative entropy in the single-letter regime.

Our results on quantum optimal transport provide the basic building blocks for a physically motivated, channel-based theory of transport. We initially explored a formulation based on joint quantum states as couplings, in analogy with the classical case, but this approach proved unsatisfactory: joint states do not compose naturally and fail to capture the dynamical structure of quantum processes. This motivated our adoption of the "states over time" formalism, in which couplings are defined bilinearly in both the input states and the channels. Within this framework, many mathematical questions arise—particularly those concerning the definition of a suitable distinguishability metric—which much of our analysis attempts, and partially succeeds, to address. The main gap in our theory lies in the characterisation of the dual of the set of states over time. A complete characterisation of this set would provide necessary and sufficient conditions for the positivity of the cost and for the fulfilment of the triangle inequality. Because of its importance, this problem has been a central part of our work over the last few years. The fact that it remains unsolved, as well as its similarity to known hard problems, suggests that its resolution may involve significant computational or structural complexity, as discussed in Section 5.3.

Given that a full characterisation of all valid cost functions appears intractable, we focus instead on a particular, physically meaningful family, namely those that are invariant under unitary transformations. This symmetry singles out a unique cost function within this class, allowing us to gain deeper insight into the structure and limitations of our quantum optimal transport functional. In this unitary-invariant setting, we are able to highlight genuinely quantum features and to further probe the mathematical and conceptual boundaries of our theory. These results also raise several open questions. As discussed in Section 3.5, the square root that appears in the joint-state formulation of the cost is both arbitrary and undesirable. One of our motivations for adopting the "states over time" framework was precisely that the coupling becomes linear in both the state and channel inputs, avoiding this asymmetry. Nonetheless, the closed formula we obtain for isospectral states in the unitary-invariant regime surprisingly behaves again as the square of a distance. This observation leads us to conjecture a deeper connection between quantum transport and the emergence of such square-root structures.

A particularly remarkable feature of the unitarily invariant optimal transport cost is its sensitivity to the behaviour of the channel outside the joint support of the input states. This sharply contrasts with the classical case, where the cost depends only on regions with nonzero input probability and is entirely unaffected by the transport plan elsewhere. This nonlocal sensitivity is reminiscent of the quantum Aharonov–Bohm effect [AB59], in which a magnetic field confined to a distant region can still influence the interference pattern of a charged particle's wave function. In Section 6.4, we explore the limit of increasingly large ambient Hilbert spaces and show that it is possible to define a renormalised cost that depends only on the action of the channel over the joint support of the states. The mathematical and physical properties of this renormalised quantity remain to be fully understood and deserve further investigation.

We next extended the notion of unitarily invariant cost to a framework based on an energy metric, corresponding to invariance under unitaries that commute with a fixed Hamiltonian. Despite progress in formalising this class of costs, we have not yet achieved a complete characterisation of the valid cost matrices within it. Beyond the Hamiltonian-based cost, other quantum cost functions remain open for exploration—for instance, a quantum generalisation of the Hamming distance, which could find applications in quantum information theory.

Finally, we turn to a different line of results concerning quantum entropy inequalities. We establish, to the best of our knowledge, the first single-letter chain rule inequality for the quantum relative entropy, derived using the formalism of states over time and quantum Bayes' theorem. In our chain rule inequality, a residual degree of freedom appears in the choice of measurement defining the ensemble partitions. In principle, one could optimise over all measurements to obtain a tighter bound, and it would be highly desirable to find an analytical procedure to determine the optimal measurement as a function of the underlying states and channels.

We also investigate the limitations of the most natural quantum extension of the classical chain rule, namely the use of rank-one orthogonal projectors. We identify a sufficient condition under which this extension holds, related to the recoverability of the corresponding ensemble, and provide a counterexample showing that the condition is not generally necessary. This points to the existence of a sharper—ideally necessary and sufficient——criterion yet to be found.

Beyond their foundational interest, these entropy inequalities could have practical implications in quantum cryptography, error correction, and channel estimation.

A

SDPs in practice

In Section 5.2 we saw that our quantum optimal transport approach an be computed with an SDP. SDPs appear as well in Section 3.3 and Section 4.4.2. In this appendix we briefly comment on the practicality of these SDPs and how we computed them.

This appendix is divided in four parts, we first discuss the SDP related to our main quantum optimal transport in Section 5.2 and we comment on our calculations of this SDP, then we comment on the SDP in Section 4.4.2 and finally we briefly discuss the relation between an SDP and its square root, due to the square root that we see appear in Section 3.3.1.

A.1 SDP for the Jordan product QOT

We first recall the SDP associated to our definition of the quantum optimal transport cost, Definition 5.4:

$$\begin{aligned} \min_{J} \quad & \mathrm{Tr}\left[(K \star \rho) J\right] \\ \text{s.t.} \quad & \begin{cases} \mathrm{Tr}_B\, J = \mathbb{1} \\ \mathrm{Tr}_A\left[\rho J\right] = \sigma \\ J^{T_A} \geq 0 \end{cases} . \end{aligned} \tag{A.1}$$

In our experience, the condition $J^{T_A} \geq 0$ is hard to implement due to the partial transpose and J being on our variable. Due to the properties of ^{T_A}, we can shift it to the fixed objects and instead optimise over the Choi matrices, obtaining

$$\begin{aligned} \min_{C} \quad & \mathrm{Tr}\left[(K \star \rho)^{T_A} C\right] \\ \text{s.t.} \quad & \begin{cases} \mathrm{Tr}_B\, C = \mathbb{1} \\ \mathrm{Tr}_A\left[\rho^T C\right] = \sigma \\ C \geq 0 \end{cases} . \end{aligned} \tag{A.2}$$

In terms of the scaling of this problem, SDPs scale efficiently. Note that the computation is performed in the tensor space, which scales as n^2, where n is the dimension of the Hilbert space. Using a personal computer, we could produce graphics[1] for up to $n = 7$ for reasonable (a couple hours) time. $n = 8$ was attempted once and took several days.

We present the code for this SDP. It relies on functions we defined for the Jordan product (`jordanP`); a standard cost matrix, in case none is given (`genCP`); as well as function for the partial transpose defined in by Murray in [Mur18]. The calculations were performed using CVXPY [DB16] and the solver MOSEK [MOS24]. We also used the QuTiP package [JNN12; JNN13] for manipulation of quantum objects. The rest of the code, as well as figures produced can be found in [Hoo25].

```
import cvxpy as cp
import numpy as np
import qutip as qt
from cvxpy_Ptrace import (partial_trace,
                          np_partial_trace)

def solvePrimalTransport(rho, sigma, C=None):

    #Buld rho and sigma as numpy arrays, and set the
       ↪ dimension
    if type(rho)==type(qt.Qobj()):
        rho = rho.full()
    d = len(rho)
    if type(sigma)==type(qt.Qobj()):
        sigma = sigma.full()
    if len(sigma)!=d:
```

[1]That is, solve the problem around one hundred times to get enough resolution.

```
        raise Exception('rho and sigma must have the
          ↪ same size')

    #Check/build the cost matrix:
    if type(C)==type(None):
        C=genCP(d,2)
    if type(C)==type(np.identity(d)):
        C=qt.Qobj(C)
        C.dims=[[d,d],[d,d]]
    if type(C)==type(np.matrix(0)):
        C=qt.Qobj(C)
        C.dims=[[d,d],[d,d]]

    #Operate on the inputs
    rho = qt.Qobj(np.kron(rho,np.identity(d)))
    rho.dims = [[d,d],[d,d]]
    C = jordanP(rho,C)
    C = qt.partial_transpose(C,[1,0]).full()
    rho = qt.partial_transpose(rho,[1,0])
    rho = rho.full()

    #Build the problem:
    E_1 = cp.Variable((d*d,d*d), hermitian=True)
    constraints = [partial_trace(E_1, [d,d], 1) == np.
      ↪ identity(d), partial_trace(rho @ E_1, [d,d],
      ↪ 0) == sigma, E_1>>0]
    objective = cp.Minimize(cp.real(cp.trace(C @ E_1)))
    prob = cp.Problem(objective, constraints)
    cost = prob.solve(solver='MOSEK')

    #print('Cost is ', cost)
    #print('E_1 is ', E_1.value)


    return cost, E_1.value
```

A.2 SDPs as feasibility solvers

Eq. (4.38) presents an SDP to invert the Jordan product as presented in Theorem 4.6. We recall this SDP:

$$\begin{array}{ll} \min_{J} & f(J) \\ \text{s.t.} & \begin{cases} \langle ik|\, J \,|j\ell\rangle = \dfrac{2}{p_i + p_j} \langle ik|\, \omega \,|j\ell\rangle\,, & \forall i,j \in B \mid p_i + p_j \neq 0,\ \forall k,\ell \in B \\ \mathrm{Tr}_B\, J = \mathbb{1} \\ J^{T_A} \geq 0 \end{cases} . \end{array} \tag{A.3}$$

In the SDP, 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 non-empty 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:

$$\begin{array}{ll} \min_{(x,J)} & -\,x \\ \text{s.t.} & \begin{cases} \langle ik|\, J \,|j\ell\rangle = \dfrac{2}{p_i + p_j} \langle ik|\, \omega \,|j\ell\rangle\,, & \forall i,j \in B \mid p_i + p_j \neq 0,\ \forall k,\ell \in B \\ \mathrm{Tr}_B\, J = \mathbb{1} \\ J^{T_A} \geq x\mathbb{1} \end{cases} . \end{array} \tag{A.4}$$

From this it is clear that if the output of the SDP is a non-negative x then the associated matrix J will be a Jamiołkowski matrix. This changes to a larger feasibility space guarantees that the SDP will be computable.

A.3 Square root SDP

We want to write the quantity we are interested in Section 3.3.1:

$$\sqrt{\mathcal{C}(\rho,\sigma)} = \sqrt{\inf_{\omega \in \Omega(\rho,\sigma)} \mathrm{Tr}\,[K\omega]} \tag{A.5}$$

as an SDP. Note that we were originally interested in this because the variables of the problem might change in some interesting way[2]. Our original SDP is:

$$\begin{aligned} \min_{\omega} \quad & \mathrm{Tr}\left[K\omega\right] \\ \text{s.t.} \quad & \begin{cases} \mathrm{Tr}_B\left[\omega\right] = \rho \\ \mathrm{Tr}_A\left[\omega\right] = \sigma \\ \omega \geq 0 \end{cases}, \end{aligned} \tag{A.6}$$

and its dual:

$$\begin{aligned} \max \quad & \mathrm{Tr}\left[A\rho\right] - \mathrm{Tr}\left[B\sigma\right] \\ \text{s.t.} \quad & \begin{cases} A \otimes \mathbb{1} - \mathbb{1} \otimes B \leq K \\ A, B \geq 0 \end{cases}. \end{aligned} \tag{A.7}$$

Based on the method to write an SDP for the square of an SDP shown in [VB96] we modify the dual[3,4] to find the square root of the original problem. The new dual formulation we obtain is

$$\begin{aligned} \max \quad & t \\ \text{s.t.} \quad & \begin{cases} A \otimes \mathbb{1} - \mathbb{1} \otimes B \leq K \\ \begin{bmatrix} \mathrm{Tr}\left[A\rho\right] - \mathrm{Tr}\left[B\sigma\right] & t \\ t & 1 \end{bmatrix} \geq 0 \\ A, B \geq 0 \end{cases}. \end{aligned} \tag{A.8}$$

We next find the dual to this problem to recover the primal formulation for the

[2] They do not, but the exercise itself seems like useful knowledge.

[3] We are using a dual formulation that forces A, B to be psd. This is not necessary and we first undo this step since it adds unnecessary variables to the final result. We will use both formulations interchangeably depending on which one is more convenient.

[4] We can not, to my knowledge, modify the primal problem straight away.

square root of our original problem. The resulting problem is:

$$
\begin{aligned}
\min \quad & \mathrm{Tr}\left[K\omega\right] + \mathrm{Tr}\left[\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix} X\right] \\
\text{s.t.} \quad & \begin{cases}
\mathrm{Tr}_B\left[\omega\right] = \rho\,\mathrm{Tr}\left[\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} X\right] \\
\mathrm{Tr}_A\left[\omega\right] = \sigma\,\mathrm{Tr}\left[\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} X\right] \\
\quad -1 = \mathrm{Tr}\left[\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} X\right] \\
\quad \omega, X \geq 0
\end{cases},
\end{aligned}
\tag{A.9}
$$

where X is a 2×2 positive semidefinite matrix.

These formulations have been tested numerically to yield the correct result.

A.3.1 Analytical solution

Pure state case

Similarly to the original problem, we can solve the case where either ρ or σ is a pure state. Then we know that $\omega = x_{11}\rho\otimes\sigma$. The problem becomes

$$
\begin{aligned}
\min \quad & Cx_{11} + x_{22} \\
\text{s.t.} \quad & \begin{cases}
x_{11}x_{22} \geq \frac{1}{4} + b^2 \\
x_{11}, x_{22} \geq 0 \\
\quad b \in \mathbb{R}
\end{cases},
\end{aligned}
\tag{A.10}
$$

where $C = \mathrm{Tr}\left[K\rho\otimes\sigma\right]$. A little basic optimisation shows that the optimal solution is

$$
\begin{aligned}
b &= 0 \\
x_{11} &= \frac{1}{2\sqrt{C}} \\
x_{22} &= \frac{\sqrt{C}}{2}
\end{aligned}
\tag{A.11}
$$

General mixed states

We can do something similar in the general case, which will show that solving the new problem is equivalent to solving the original problem.

We see from the partial trace conditions that there is a bijective map between Ω and Ω_{old}:

$$\omega = x_{11}\omega_{old}.$$

thus we can write our problem as (now ignoring b, since it always trivially goes to 0)

$$\begin{aligned} \min \quad & \mathrm{Tr}\left[K\omega_{old}\right] x_{11} + x_{22} \\ \text{s.t.} \quad & \begin{cases} x_{11}x_{22} \geq \frac{1}{4} \\ x_{11}, x_{22} \geq 0 \end{cases}. \end{aligned} \tag{A.12}$$

We can see that $\mathrm{Tr}\left[K\omega_{old}\right]$ acts on the function to be minimised in a way such that

$$\mathrm{Tr}\left[K\omega_{old}\right] < \mathrm{Tr}\left[K\omega'_{old}\right] \Rightarrow \mathrm{Tr}\left[K\omega_{old}\right] x_{11} + x_{22} < \mathrm{Tr}\left[K\omega'_{old}\right] x_{11} + x_{22}. \tag{A.13}$$

Therefore we want to take the minimum $C = \mathrm{Tr}\left[K\omega_{old}\right]$ possible, which is just the solution to the original problem. Now that this quantity is fixed this becomes equivalent to the pure state case we just discussed, thus:

$$\begin{aligned} x_{11} &= \frac{1}{2\sqrt{C}} \\ x_{22} &= \frac{\sqrt{C}}{2} \\ \omega &= x_{11}\omega_{old}. \end{aligned} \tag{A.14}$$

We have seen that, in both cases, X is

$$X = \frac{1}{2}\begin{bmatrix} \frac{\sqrt{C}}{C} & -1 \\ -1 & \sqrt{C} \end{bmatrix}. \tag{A.15}$$

Bibliography

[Agr13] Agredo, J., "A Wasserstein-type Distance to Measure Deviation from Equilibrium of Quantum Markov Semigroups", Open Systems & Information Dynamics **20**, 350009 (2013).

[AB59] Aharonov, Y. and Bohm, D., "Significance of Electromagnetic Potentials in the Quantum Theory", Physical Review **115**, 485–491 (1959).

[Ara75] Araki, H., "Relative Hamiltonian for states of von Neumann algebras", Trudy Mat. Inst. Steklov. **135**, 18–25, 256 (1975).

[Arg20] Argerami, M., *Generated von Neumann algebra and commutant*, (2020) https://math.stackexchange.com/questions/3726289/generated-von-neumann-algebra-and-commutant (visited on 05/24/2022).

[Arv76] Arveson, W., *An Invitation to C*-Algebras* (Springer New York, 1976).

[BW11] Bakonyi, M. and Woerdeman, H., *Matrix Completions, Moments, and Sums of Hermitian Squares* (Princeton University Press, Princeton NJ, 2011).

[Bea25] Beatty, E., "Wasserstein Distances on Quantum Structures: an Overview", arXiv:2506.09794v1 (2025).

[BS25] Beatty, E. and Stilck França, D., "Order p Quantum Wasserstein Distances from Couplings", Annales Henri Poincaré (2025).

[BB00] Benamou, J.-D. and Brenier, Y., "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem", Numerische Mathematik **84**, 375–393 (2000).

[BŻ06] Bengtsson, I. and Życzkowski, K., *Geometry of Quantum States, An introduction to quantum entanglement* (Cambridge University Press, 2006).

[BK08] Bertlmann, R. A. and Krammer, P., "Bloch vectors for qudits", Journal of Physics A: Mathematical and Theoretical **41**, 235303 (2008).

[Bia20] Biamonte, J., "Lectures on Quantum Tensor Networks", arXiv:1912.10049v2 (2020).

[BEŻ23] Bistroń, R., Eckstein, M., and Życzkowski, K., "Monotonicity of a quantum 2-Wasserstein distance", Journal of Physics A: Mathematical and Theoretical **56**, 095301 (2023).

[BV04] Boyd, S. and Vandenberghe, L., *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[BHO+15] Brandao, F. G., Harrow, A. W., Oppenheim, J., and Strelchuk, S., "Quantum conditional mutual information, reconstructed states, and state redistribution", Physical Review Letters **115**, 050501 (2015).

[Bre15] Brenier, Y., "Optimal transportation of particles, fluids and currents", in *Variational methods for evolving objects. Papers based on the conference, Hokkaido University, Sapporo, Japan, July 30 – August 3, 2012* (Tokyo: Mathematical Society of Japan (MSJ), 2015), pp. 59–85.

[BPT+25] Bunth, G., Pitrik, J., Titkos, T., and Virosztek, D., "Wasserstein distances and divergences of order p by quantum channels", arXiv:math-ph/2501.08066 (2025).

[CM14] Carlen, E. A. and Maas, J., "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).

[CM17] Carlen, E. A. and Maas, J., "Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance", Journal of Functional Analysis **273**, 1810–1869 (2017).

[CM19] Carlen, E. A. and Maas, J., "Non-commutative Calculus, Optimal Transport and Functional Inequalities in Dissipative Quantum Systems", Journal of Statistical Physics **178**, 319–378 (2019).

[CDP09] Chiribella, G., D'Ariano, G., and Perinotti, P., "Theoretical framework for quantum networks", Physical Review A **80**, 022339 (2009).

[Cho75] Choi, M.-D., "Completely positive linear maps on complex matrices", Linear Algebra and Its Applications **10**, 285–290 (1975).

[CEF+23] Cole, S., Eckstein, M., Friedland, S., and Życzkowski, K., "On Quantum Optimal Transport", Mathematical Physics, Analysis and Geometry **26** (2023).

[DMT+21] De Palma, G., Marvian, M., Trevisan, D., and Lloyd, S., "The Quantum Wasserstein Distance of Order 1", IEEE Transactions on Information Theory **67**, 6627–6643 (2021).

[DT21] De Palma, G. and Trevisan, D., "Quantum Optimal Transport with Quantum Channels", Annales Henri Poincaré **22**, 3199–3234 (2021).

[DT23] De Palma, G. and Trevisan, D., "The Wasserstein Distance of Order 1 for Quantum Spin Systems on Infinite Lattices", Annales Henri Poincaré **24**, 4237–4282 (2023).

[DB16] Diamond, S. and Boyd, S., "CVXPY: A Python-embedded modeling language for convex optimization", Journal of Machine Learning Research **17**, 1–5 (2016).

[Dur19] Durrett, R., *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 2019).

[Duv21] Duvenhage, R., "Optimal quantum channels", Physical Review A **104** (2021).

[EK95] Ekert, A. and Knight, P., "Entangled quantum systems and the Schmidt decomposition", American Journal of Physics **63**, 415 (1995).

[Eve57] Everett, H., ""Relative State" Formulation of Quantum Mechanics", Rev. Mod. Phys. **29**, 454–462 (1957).

[FFR+20] Fang, K., Fawzi, O., Renner, R., and Sutter, D., "Chain Rule for the Quantum Relative Entropy", Physical Review Letters **124** (2020).

[FF18] Fawzi, H. and Fawzi, O., "Efficient optimization of the quantum relative entropy", Journal of Physics A: Mathematical and Theoretical **51**, 154003 (2018).

[FR15] Fawzi, O. and Renner, R., "Quantum conditional mutual information and approximate Markov chains", Communications in Mathematical Physics **340**, 575–611 (2015).

[FJV15] Fitzsimons, J. F., Jones, J. A., and Vedral, V., "Quantum correlations which imply causation", Scientific Reports **5**, 1–7 (2015).

[FEC+22] Friedland, S., Eckstein, M., Cole, S., and Życzkowski, K., "Quantum Monge-Kantorovich Problem and Transport Distance between Density Matrices", Physical Review Letters **129**, 110402 (2022).

[Fuc01] Fuchs, C. A., "Quantum Foundations in the Light of Quantum Information", arXiv:quant-ph/0106166v1 (2001).

[Ful23] Fullwood, J., "Quantum dynamics as a pseudo-density matrix", arXiv: quant-ph/2304.03954 (2023).

[FP22] Fullwood, J. and Parzygnat, A. J., "On quantum states over time", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **478**, 20220104 (2022).

[FP25] Fullwood, J. and Parzygnat, A. J., "On Dynamical Measures of Quantum Information", Entropy **27**, 331 (2025).

[GH25] Gasbarri, G. and Hoogsteder-Riera, M., "Chain rules for quantum relative entropy in the single-letter regime", in preparation, 2025.

[GMP16] Golse, F., Mouhot, C., and Paul, T., "On the Mean Field and Classical Limits of Quantum Mechanics", Communications in Mathematical Physics **343**, 165–205 (2016).

[Gur03] Gurvits, L., "Classical deterministic complexity of Edmonds' problem and Quantum Entanglement", Conference Proceedings of the Annual ACM Symposium on Theory of Computing **10**, 10–19 (2003).

[Ham50] Hamming, R. W., "Error detecting and error correcting codes", The Bell System Technical Journal **29**, 147–160 (1950).

[Hay01] Hayashi, M., "Asymptotics of quantum relative entropy from a representation theoretical viewpoint", Journal of Physics A: Mathematical and General **34**, 3413–3419 (2001).

[Hia20] Hiai, F., "Concise lectures on selected topics of von Neumann algebras", arXiv: 2004.02383v1 (2020).

[HP91] Hiai, F. and Petz, D., "The proper formula for relative entropy and its asymptotics in quantum probability", Communications in Mathematical Physics **143**, 99–114 (1991).

[Hir18] Hirche, C., "From asymptotic hypothesis testing to entropy inequalities", Available at `https://arxiv.org/abs/1812.05142`, PhD thesis (Universitat Autònoma de Barcelona, 2018).

[Hoo18] Hoogsteder-Riera, M., "A transport approach to distances in quantum systems", Bachelor's Thesis (Universitat Autònoma de Barcelona, 2018).

[Hoo25] Hoogsteder-Riera, M., *Quantum optimal tranport*, `https://doi.org/10.5281/zenodo.17244966`, 2025.

[HCW25] Hoogsteder-Riera, M., Calsamiglia, J., and Winter, A., "Approach to optimal quantum transport via states over time", arXiv:2504.04856 (2025).

[HHH96] Horodecki, M., Horodecki, P., and Horodecki, R., "Separability of mixed states: necessary and sufficient conditions", Physics Letters A **223**, 1–8 (1996).

[HHP+17] Horsman, D., Heunen, C., Pusey, M. F., Barrett, J., and Spekkens, R. W., "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).

[Hus40] Husimi, K., "Some Formal Properties of the Density Matrix", Proceedings of the Physico-Mathematical Society of Japan. 3rd Series **22**, 264–314 (1940).

[Jam72] Jamiołkowski, A., "Linear transformations which preserve trace and positive semidefiniteness of operators", Reports on Mathematical Physics **3**, 275–278 (1972).

[Jen06] Jensen, J. L. W. V., "Sur les fonctions convexes et les inégalités entre les valeurs moyennes", Acta Mathematica **30**, 175–193 (1906).

[JNN12] Johansson, J. R., Nation, P. D., and Nori, F., "QuTiP: An open-source Python framework for the dynamics of open quantum systems", Computer Physics Communications **183**, 1760–1772 (2012).

[JNN13] Johansson, J. R., Nation, P. D., and Nori, F., "QuTiP 2: A Python framework for the dynamics of open quantum systems", Computer Physics Communications **184**, 1234–1240 (2013).

[Jon15] Jones, V. F. R., "Von Neumann Algebras", (2015).

[Jor32] Jordan, P., "Über eine Klasse nichtassoziativer hyperkomplexer Algebren", Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse **1932**, 569–575 (1932).

[Jor33] Jordan, P., "Über die Multiplikation quantenmechanischer Größen", Zeitschrift für Physik **80**, 285–291 (1933).

[Jor34] Jordan, P., "Über die Multiplikation quantenmechanischer Größen. II", Zeitschrift für Physik **87**, 505–512 (1934).

[JvW35] Jordan, P., von Neumann, J., and Wigner, E., "On an Algebraic Generalization of the Quantum Mechanical Formalism", Annals of Mathematics **35**, 29–64 (1935).

[JRS+18] Junge, M., Renner, R., Sutter, D., Wilde, M. M., and Winter, A., "Universal Recovery Maps and Approximate Sufficiency of Quantum Relative Entropy", Annales Henri Poincaré **19**, 2955–2978 (2018).

[Kan48] Kantorovich, L. V., "On a problem of Monge", Uspekhi Mat. Nauk. **3**, English translation: J. Math. Sciences, vol. 133, p. 1383 (2006), 225–226 (1948).

[Kan60] Kantorovich, L. V., "Mathematical Methods of Organizing and Planning Production", Management Science **6**, 366–422 (1960).

[Kan58] Kantorovitch, L. V., "On the Translocation of Masses", Management Science **5**, 1–4 (1958).

[Kel84] Kellerer, H., "Duality theorems for marginal problems", Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **67**, 399–432 (1984).

[Kra83] Kraus, K., "States, Effects and Operations", in *Lecture notes in physics*, Vol. 190 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1983), pp. 1–12.

[KK19] Kwon, H. and Kim, M. S., "Fluctuation Theorems for a Quantum Channel", Physical Review X **9** (2019).

[LP08] Leifer, M. and Poulin, D., "Quantum Graphical Models and Belief Propagation", Annals of Physics **323**, 1899–1946 (2008).

[Lei06] Leifer, M. S., "Quantum dynamics as an analog of conditional probability", Physical Review A **74**, 042310 (2006).

[LS13] Leifer, M. S. and Spekkens, R. W., "Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference", Physical Review A **88**, 052130 (2013).

[LN24] Lie, S. H. and Ng, N. H. Y., "Quantum state over time is unique", Physical Review Research **6**, 033144 (2024).

[LR73] Lieb, E. H. and Ruskai, M. B., "Proof of the strong subadditivity of quantum-mechanical entropy", Journal of Mathematical Physics **14**, 1938–1941 (1973).

[Lin75] Lindblad, G., "Completely positive maps and entropy inequalities", Communications in Mathematical Physics **40**, 147–151 (1975).

[Mon81] Monge, G., *Mémoire sur la théorie des déblais et des remblais* (De l'Imprimerie Royale, Paris, 1781).

[MOS24] MOSEK ApS, *MOSEK Optimizer API for Python 10.2.5* (2024).

[Mur18] Murray, R., *Native function for partial trace*, `https://github.com/cvxpy/cvxpy/issues/563`, Commented on Aug 20, 2018.

[NC10] Nielsen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[OH01] Ogawa, T. and Hayashi, M., "A New Proof of the Direct Part of Stein's Lemma in Quantum Hypothesis Testing", arXiv:quant-ph/0110125 (2001).

[Per96] Peres, A., "Separability Criterion for Density Matrices", Physical Review Letters **77**, 1413–1415 (1996).

[Pet86] Petz, D., "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra", Communications in mathematical physics **105**, 123–131 (1986).

[Pet88] Petz, D., "Sufficiency of channels over von Neumann algebras", The Quarterly Journal of Mathematics **39**, 97–108 (1988).

[PM17] Puchała, Z. and Miszczak, J. A., "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).

[RD17] Rouzé, C. and Datta, N., "Relating Relative Entropy, Optimal Transport and Fisher Information: A Quantum HWI Inequality", Annales Henri Poincaré **21** (2017).

[RD19] Rouzé, C. and Datta, N., "Concentration of quantum states from quantum functional and transportation cost inequalities", Journal of Mathematical Physics **60**, 012202 (2019).

[Rus07] Ruskai, M. B., "Another short and elementary proof of strong subadditivity of quantum entropy", Reports on Mathematical Physics **60**, 1–12 (2007).

[San18] Santambrogio, F., *A short story on optimal transport and its many applications*, edited by imaginary.org, `https://doi.org/10.14760/SNAP-2018-013-EN`, 2018.

[SAS+25] Scandi, M., Abiuso, P., Surace, J., and De Santis, D., "Quantum Fisher information and its dynamical nature", Reports on Progress in Physics **88**, 076001 (2025).

[Sch07] Schmidt, E., "Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener", Mathematische Annalen **63**, 433–476 (1907).

[SSŻ09] Skowronek, Ł., Størmer, E., and Życzkowski, K., "Cones of positive maps and their duality relations", Journal of Mathematical Physics **50**, 062106 (2009).

[Sti55] Stinespring, W. F., "Positive Functions on C*-Algebras", Proceedings of the American Mathematical Society **6**, 211–216 (1955).

[SAP17] Streltsov, A., Adesso, G., and Plenio, M. B., "Colloquium: Quantum coherence as a resource", Reviews in Modern Physics **89**, 041003 (2017).

[SBT16] Sutter, D., Berta, M., and Tomamichel, M., "Multivariate Trace Inequalities", Communications in Mathematical Physics **352**, 37–58 (2016).

[STH16] Sutter, D., Tomamichel, M., and Harrow, A. W., "Strengthened monotonicity of relative entropy via pinched Petz recovery map", IEEE Transactions on Information Theory **62**, 2907–2913 (2016).

[The75] The Royal Swedish Academy of Sciences, *The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1975*, `https://www.nobelprize.org/prizes/economic-sciences/1975/summary`, Accessed on 22/09/2025, 1975.

[Uhl77] Uhlmann, A., "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory", Communications in Mathematical Physics **54** (1977).

[Val09] Vallentin, F., "Symmetry in semidefinite programs", Linear Algebra and its Applications **430**, 360–369 (2009).

[VB96] Vandenberghe, L. and Boyd, S., "Semidefinite Programming", SIAM Review **38**, 49–95 (1996).

[Vas69] Vaseršteĭn, L., "Markov processes over denumerable products of spaces describing large system of automata", Problems Inform. Transmission, 47–52 (1969).

[Vil08] Villani, C., *Optimal Transport: Old and New*, Vol. 338, Grundlehren der mathematischen Wissenschaften (Springer Verlag, Berlin, Paris, Dakar, 2008).

[Wer89] Werner, R. F., "Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model", Physical Review A **40**, 4277–4281 (1989).

[Wil15] Wilde, M. M., "Recoverability in quantum information theory", Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471**, 20150338 (2015).

[Wir21] Wirth, M., "A Noncommutative Transport Metric and Symmetric Quantum Markov Semigroups as Gradient Flows of the Entropy", arXiv:1808.05419 (2021).

[Woo87] Wootters, W. K., "A Wigner-function formulation of finite-state quantum mechanics", Annals of Physics **176**, 1–21 (1987).

[ŻS98] Życzkowski, K. and Słomezyński, W., "The Monge distance between quantum states", Journal of Physics A: Mathematical and General **31**, 90–95 (1998).

[ŻS01] Życzkowski, K. and Słomezyński, W., "The Monge metric on the sphere and geometry of quantum states", Journal of Physics A: Mathematical and General **34**, 6689–6722 (2001).