A concrete example of matrix tuples optimally coupled in a larger matrix algebra

Junchen Zhao
Advisor: Professor Todd Kemp and Dr. David Jekel

June 1, 2022

Contents

1 Introduction

1.1 Main results

In this thesis, we study optimal couplings in the non-commutative probability space $(M_n(\mathbb{C}), \mathrm{tr})$ extending from discussions in papers [GJNS21], [HM11], and [MR19]. We will provide motivations and background information, connect their results by filling in intermediate steps, and come up with a new concrete example of their existence claims.

Non-commutativity requires us to use more abstract tools to describe random variables and their joint distributions. Therefore, optimal coupling in non-commutative probability spaces is formulated in more algebraic terms compared to in classical probability spaces, where we seek a product probability measure with fixed marginals that achieves optimal L^2 distance between the marginals. The problem of optimally coupling matrix tuples $X, Y \in M_n(\mathbb{C})^m_{\mathrm{sa}}$ requires us to find a tracial von Neumann algebra $(\mathcal{C}, \tau)$ and unital trace-preserving $*$-homomorphisms $\alpha, \beta : M_n(\mathbb{C}) \to \mathcal{C}$ that achieve

$$\inf \sum_{j=1}^{m} \|\alpha(X_j) - \beta(Y_j)\|_\tau^2, \text{ or equivalently, } \sup \sum_{j=1}^{m} \langle \alpha(X_j), \beta(Y_j) \rangle_\tau.$$

Moreover by a calculation in Proposition 2.2, it often reduces to achieve

$$\sup_{\Psi \in \mathcal{FM}(M_n(\mathbb{C}))} \sum_{j=1}^{m} \langle \Psi(X_j), Y_j \rangle_{\mathrm{tr}},$$

where $\mathcal{FM}(M_n(\mathbb{C}))$ is the set of factorizable maps $M_n(\mathbb{C}) \to M_n(\mathbb{C})$. Every $\Psi \in \mathcal{FM}(M_n(\mathbb{C}))$ is of the form $\beta^* \circ \alpha$, and we say it factorizes through $\mathcal{C}$. Details are given in Definition 2.5, 2.11, 2.16, and 2.17.

In [GJNS21] (Abstract), the authors point out that two non-commutative laws that can be realized in finite-dimensional algebras may still require an infinite-dimensional algebra to optimally couple. This is done by studying the existence of factorizable maps between finite-dimensional algebras that factorize through an infinite-dimensional algebra. In [HM11] (Theorem 5.6), the authors give concrete examples of factorizable maps that cannot be expressed by any convex combinations of automorphisms of $M_n(\mathbb{C})$, meaning those factorizable maps cannot factorize through the same matrix algebra $M_n(\mathbb{C})$, so we need to embed into a tracial von Neumann algebra of a larger dimension. Moreover in [MR19] (Theorem 4.1), the authors give concrete examples of factorizable maps that cannot factorize through any finite-dimensional C^*-algebra, and in fact they factorize through the hyperfinite type II_1 factor. However, none of those papers give a concrete example of a pair of tuples in $M_n(\mathbb{C})^m$ whose optimal coupling is achieved by the factorizable maps they describe.

In this thesis, we will construct a pair of matrix tuples $X, Y \in M_3(\mathbb{C})^9_{\mathrm{sa}}$ and prove that we must embed them into a larger tracial von Neumann algebra to achieve the infimum above, thus optimally coupling X and Y. More specifically, we have $X, Y \in M_3(\mathbb{C})^9_{\mathrm{sa}}$ and $\Phi \in \mathcal{FM}(M_3(\mathbb{C}))$, factorizing through $(M_3(\mathbb{C}) \otimes M_3(\mathbb{C}), \mathrm{tr} \otimes \mathrm{tr})$, that satisfy the following equality/inequality:

$$\begin{aligned}\sup_{U \in \mathrm{U}(3)} \sum_{j=1}^{9} \langle U X_j U^*, Y_j \rangle_{\mathrm{tr}} &= \sup_{\Psi \in \mathrm{conv}(\mathrm{Aut}(M_3(\mathbb{C})))} \sum_{j=1}^{9} \langle \Psi(X_j), Y_j \rangle_{\mathrm{tr}} \\ &< \sup_{\Psi \in \mathcal{FM}(M_3(\mathbb{C}))} \sum_{j=1}^{9} \langle \Psi(X_j), Y_j \rangle_{\mathrm{tr}} = \sum_{j=1}^{9} \langle \Phi(X_j), Y_j \rangle_{\mathrm{tr}}.\end{aligned}$$

Note the first line of the inequality corresponds to the optimal coupling by embedding into the same matrix algebra $M_3(\mathbb{C})$, and the second line of the inequality correspond to the optimal coupling by embedding into all possible tracial von Neumann algebras. Details are given in Section 4.2.

Spoiler alert: $X = (X_1, \dots, X_9)$ will be 9 matrices that form an orthonormal basis of $M_3(\mathbb{C})_{\mathrm{sa}}$, and $Y = (Y_1, \dots, Y_9)$ will be given by $(N(X_1), \dots, N(X_9))$, where N is a special linear transformation on the vector space $M_3(\mathbb{C})_{\mathrm{sa}}$ relating to a "normal vector" that witnesses a strict separation.

Why do we care about such an example? It is also shown in Section 4.1 of this thesis that the optimal coupling of two tuples in $M_2(\mathbb{C})_{\mathrm{sa}}$ is always achieved by embedding into the same algebra $M_2(\mathbb{C})_{\mathrm{sa}}$, thus given by conjugating some unitary $U \in \mathrm{U}(2)$. Therefore, one might suspect whether it suffices to couple the tuples in the algebra they come from. However, our example shows it is not even sufficient in rank $n = 3$. We will get a better coupling by embedding into a larger matrix algebra, and we will show it is the optimal coupling among all embeddings into tracial von Neumann algebras.

In section 2, we rigorously discuss von Neumann algebras (aka W^*-algebras), non-commutative probability spaces (and their analogies to classical probability spaces), and factorizable maps. We also supply a list of notations used extensively in this thesis at the end of this section.

In section 3, we study optimal couplings given by unitary conjugation. We present detailed computations in rank $n = 2$.

In section 4, we study optimal couplings given by more general factorizable maps, and in Theorem 4.7 we present our new example explicitly.

The tools we mainly use come from operator algebras and functional analysis. More specifically, these include Hilbert spaces, bounded linear operators, von Neumann algebras, conditional expectations, tensor products, homomorphisms of $*$-algebras, linear isomorphisms, Hilbert projection theorem, Hahn-Banach separation theorem, Skolem–Noether theorem, Schur–Weyl duality, etc.

1.2 Motivation

We explain the motivation of this study from two perspectives: why we study optimal coupling, and why we care about non-commutative probability spaces (NCPSs).

First we take a step back and introduce optimal transport. Optimal transport in mathematics, and related fields like economics and physics, is the study of efficiently allocating resources. We give a relatable example of an optimal transport problem in the COVID era: suppose we have N orders of food ready in N different restaurants, and N UberEats drivers across the city ready for delivery. We also have a numerical cost function $c(x, y)$ representing the cost (in time, distance, etc.) of sending driver x to restaurant y. The problem is to send those drivers to those restaurants (more precisely: pair each driver with a restaurant) so that the average cost is minimized. It is clear that optimal transport has important real-world applications. Mathematically, it has an abstract formulation borrowing language from probability theory: given two probability measures μ, ν on metric spaces X, Y respectively, and a measurable cost function $c : X \times Y \to [0, \infty]$, the optimal transport problem is to find a measureable transportation map $T : X \to Y$ that "transports" μ to ν ($\nu(B) = \mu(T^{-1}(B)) \quad \forall \nu$-measurable $B \subseteq Y$) with lowest average cost of transportation, meaning that it achieves the infimum

$$\inf \left\{ \int_X c(x, T(x))\, \mu(dx) \right\}.$$

The above formulation is due to French mathematician Monge in the late eighteenth century, and is a natural extension from the food delivery example we gave: X, Y are set to be integers $\{1, \ldots, N\}$ with μ, ν being the uniform probability measure. This will require the transportation map to be a bijection (pairing) between drivers and restaurants, and our integral degenerates to a sum. However natural this formulation is, it is technically ill-posed. For example if one probability measure has point mass and the other does not, then no transportation map exists, and the infimum would be infinity. One way to get around with this is to consider an alternative setup, the Kantorovich's formulation: the optimal transport problem is to find a probability measure γ on $X \times Y$ with marginals μ on X and ν on Y, also called a transportation plan, that achieves the infimum

$$\inf \left\{ \int_{X \times Y} c(x, y)\, \gamma(dx, dy) \right\}.$$

In this formulation, there always exists a transportation plan γ taken to be the product measure $\mu \otimes \nu$ on the product space $X \times Y$. Moreover, given a transportation map T, we can always translate it to a transportation plan γ_T defined by

$$\gamma_T(A \times B) = \mu(A \cap T^{-1}(B)).$$

It follows that $\gamma_T(A \times Y) = \mu(A)$, $\gamma_T(X \times B) = \nu(B)$, and by testing on indicators of product measurable sets,

$$\int_{X \times Y} 1_{A \times B}(x, y)\, \gamma_T(dx, dy) = \int_X 1_{A \times B}(x, T(x))\, \mu(dx).$$

By Dynkin's multiplicative systems theorem ([Kem21] Lecture 14.1),

$$\int_{X\times Y} c(x,y)\,\gamma_T(dx,dy) = \int_X c(x,T(x))\,\mu(dx),$$

and thus we have a more general formulation (transportation plan always exists) of the same problem and it is consistent (the integrals are equal) when a transportation map exists.

The reason we introduce optimal transport is due to its connection with optimal coupling of random variables. A coupling of two random variables on two different spaces is realizing them as a joint distribution in the product space, such that the two marginals of the joint distribution match the distributions of the two random variables, respectively. The problem of optimally coupling two random variables is just to find an optimal transportation plan between their distributions, with a specific cost function relating to L^p norms.

Now that we revealed the importance of optimal coupling by its connection with optimal transport, we explain our other motivation: what are NCPSs, and why do we care about them?

Recall in a classical probability space, we have the triple $(\Omega, \mathcal{F}, \mathbb{P})$, a space of outcomes, a space of events, and a probability measure. We study random variables as (Borel-)measurable (complex valued) functions $X : \Omega \to \mathbb{C}$. It is inherent in this measure-theoretic setup that the multiplication of our random variables is automatically commutative. As we witness plenty of instances of mathematicians generalizing mathematical concepts through many levels of abstraction in all branches of mathematics, it is also the case in probability theory. After John von Neumann founded the study of operator algebras through the von Neumann algebras in the 1930s, von Neumann algebras are considered to be the non-commutative version of measure theory, which becomes the tool we use to study non-commutative probability theory, just like their commutative counterparts. However, in contrast to classical probability theory where our basic constructs are events and probabilities, in non-commutative probability theory, our basic constructs are (a non-commutative algebra of) "random variables" and their "expectation", so the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is replaced by a tracial von Neumann algebra $(\mathcal{A}, \tau)$. Hence we can study how various concepts in classical probability theory transfer to non-commutative probability theory, including probability distribution, independence, conditional expectation, limit laws, and of course, (optimal) coupling of random variables.

One reason to study NCPSs is due to their significant applications in physics. This brings us to a historical review of the beginning of the study of quantum mechanics and quantum information theory. Uncoincidentally, von Neumann was also the first to establish a mathematically rigorous framework of quantum mechanics around 1930 in his papers [vN27a], [vN27c], and [vN27b]. In his formulation, the state of a quantum mechanical system is an element of a complex Hilbert space, and physical quantities of interest (position, momentum, etc) are represented by observables, which are self-adjoint linear operators acting on that Hilbert space. Among various others, the non-commutativity of linear operators is a precise reflection of the uncertainty principle in quantum mechanics: determination of the position of a particle prevents determination of its momentum, and vice versa. The rigorous study of von Neumann algebras was therefore carried out as a special kind of subalgebras of bounded linear operators on a Hilbert space (von Neumann was also the first one to come up with the abstract and axiomatic definition of a Hilbert space that we use today). In this way, non-commutative probability theory becomes the crucial tool in studying the information of the state of a quantum system (known as quantum information theory), again, just like their classical counterparts. Meanwhile, we will see in this thesis how coupling of non-commutative random variables relate to factorizable maps, a special type of quantum channels (a communication channel transmitting quantum information) in quantum information theory. Our construction of a concrete example of matrix tuples also uses previous results on quantum channels.

2 Background

In this section, we provide definitions and motivations of optimal coupling in a non-commutative probability space under the context of operator algebra, reveal their connection to factorizable maps and quantum information theory, as well as introduce relevant examples and notations that are used extensively in the rest of the thesis. The readers can refer to the appendix for more details (proofs), but we also include some excellent introductory texts of the subjects aforementioned. For background reference on C^*-algebras, see [Dix69]; for operator algebras, see [Zhu93], [KR83], [Tak02]; for W^*-algebras, see [Sak71]; for introduction to free probability, see [VDN92]; for introduction to quantum information theory, see [Wil13], [Wit20].

2.1 Non-commutative probability space (NCPS)

In essence, we want to extend the notion of a probability space into one that allows the random variables to be potentially non-commutative. Similar to using the language of measure theory to study probability theory, we use the language of tracial W^*-algebra, a non-commutative analogue of measure theory, to study non-commutative probability. We start with some definitions from [GJNS21] (Definition 2.1-2.11) or standard textbooks in functional analysis and operator algebra.

Definition 2.1 (unital $*$-algebra)**.** A unital $*$-algebra is a untial algebra $\mathcal{A}$ over $\mathbb{C}$ closed under $*$-operation. That is, for arbitrary $a, b \in \mathcal{A}, \alpha \in \mathbb{C}$,

- $(a^*)^* = a$;
- $(ab)^* = b^* a^*$;
- $(\alpha a + b)^* = \overline{\alpha} a^* + b^*$.

Definition 2.2 (weak operator topology)**.** Let $\{T_n\}_{n=1}^{\infty}$ and T be bounded linear operators on a Hilbert space $\mathcal{H}$, then $T_n \to T$ in the weak operator topology (WOT) if

$$\lim_{n\to\infty} \langle T_n \xi, \eta \rangle = \langle T\xi, \eta \rangle$$

for arbitrary $\xi, \eta \in \mathcal{H}$.

Definition 2.3 (W^*-algebra)**.** A W^*-algebra (von Neumann algebra) $\mathcal{A}$ is a unital $*$-algebra of bounded linear operators on a Hilbert space such that $\mathcal{A}$ is closed under WOT.

There are other analytical or abstract characterizations of W^*-algebras due to von Neumann's bicommutant theorem and a theorem of Sakai [Sak71], respectively.

Definition 2.4 (trace)**.** Let $\mathcal{A}$ be a W^*-algebra, then a faithful normal trace on $\mathcal{A}$ is a linear functional $\tau : \mathcal{A} \to \mathbb{C}$ such that

- unital: $\tau(1) = 1$;
- positive: $\tau(a^* a) \geq 0$ for all $a \in \mathcal{A}$;
- faithful: $\tau(a^* a) = 0$ if and only if $a = 0$;
- tracial: $\tau(ab) = \tau(ba)$ for all $a, b \in \mathcal{A}$;
- τ is weak-$*$ continuous.

Note that the last condition insisting on τ being continuous in the weak-$*$ topology is slightly weaker than in the weak operator topology, but it gives a non-commutative analogue of dominated convergence in measure theory (more about this analogue in Section 2.1.1).

Definition 2.5 (NCPS)**.** A non-commutative probability space is a tracial W^*-algebra $(\mathcal{A}, \tau)$, where $\mathcal{A}$ is a W^*-algebra and τ is a faithful normal trace.

Definition 2.6 (Non-commutative polynomial algebra)**.** We denote by $\mathbb{C}\langle x_1, \ldots, x_m \rangle$ the universal unital algebra generated by variables $x_1, \ldots, x_m$. As a vector space, $\mathbb{C}\langle x_1, \ldots, x_m \rangle$ has a basis consisting of all products $x_{i_1} \cdots x_{i_l}$ for $l \geq 0$ and $i_1, \ldots, i_l \in \{1, \ldots, d\}$. We equip $\mathbb{C}\langle x_1, \ldots, x_m \rangle$ with the unique $*$-operation such that $(x_{i_1} \cdots x_{i_l})^* = x_{i_l}^* \cdots x_{i_1}^*$.

Definition 2.7 (Non-commutative law)**.** the $*$-distribution λ_{a,a^*} of a is the collection of all joint moments of a and a^*. More specifically, λ_{a,a^*} is a linear functional on $\mathbb{C}\langle x, y \rangle$ satisfying $\lambda_{a,a^*}(p(x, y)) = \tau(p(a, a^*))$ for all $p \in \mathbb{C}\langle x, y \rangle$.

Definition 2.8 (non-commutative law of a self-adjoint m-tuple)**.** Let $(\mathcal{A}, \tau)$ be a NCPS. Let $X = (X_1, \ldots, X_m) \in L^\infty(\mathcal{A})^m_{\text{sa}}$. Then we define $\lambda_X : \mathbb{C}\langle x_1, \ldots, x_m\rangle \to \mathbb{C}$ by $\lambda_X(p) = \tau(p(X_1, \ldots, X_m))$ for $p \in \mathbb{C}\langle x_1, \ldots, x_m\rangle$.

Notation 2.9. In a NCPS $(\mathcal{A}, \tau)$, $\langle a, b\rangle_\tau = \tau(b^* a)$ gives an inner product on $\mathcal{A}$. Let $\|\cdot\|_\tau$ be the induced norm of the inner product space $(\mathcal{A}, \langle\cdot,\cdot\rangle_\tau)$. We denote the Hilbert space $\mathcal{H}_\tau = \overline{\mathcal{A}}^{\|\cdot\|_\tau}$, as the Cauchy completion of the metric induced by the norm $\|\cdot\|_\tau$. Abusing notations from measure theory, we also denote $L^2(\mathcal{A}) = \mathcal{H}_\tau$, and think of $L^\infty(\mathcal{A}) = \mathcal{A}$ (as an isomorphic copy) being a subspace of $L^2(\mathcal{A})$ due to the embedding in the Cauchy completion $\mathcal{A} \xrightarrow{\widehat{\ }} \mathcal{H}_\tau$.

Now that we have a Hilbert space, we can talk about subspaces (sub-W^*-algebras) and orthogonal projections (conditional expectations).

Definition 2.10 (conditional expectation)**.** Let $(\mathcal{A}, \tau)$ be a NCPS, and $\mathcal{B} \subseteq \mathcal{A}$ is a sub-W^*-algebra (so $(\mathcal{B}, \tau|_\mathcal{B})$ is a non-commutative probability subspace). Then there exists a map $E_\mathcal{B} : \mathcal{A} \to \mathcal{B}$ uniquely determined by $\tau(E_\mathcal{B}(a)b) = \tau(ab)$ for all $a \in \mathcal{A}, b \in \mathcal{B}$. Moreover, $E_\mathcal{B}$ satisfies the following properties:

- $E_\mathcal{B}$ is a $\mathcal{B}$-$\mathcal{B}$ bimodule map: $E_\mathcal{B}(a_1 b a_2) = a_1 E_\mathcal{B}(b) a_2$ for all $a_1, a_2 \in \mathcal{A}$ and $b \in \mathcal{B}$;
- $E_\mathcal{B}$ is unital: $E_\mathcal{B}(1) = 1$;
- $E_\mathcal{B}$ is trace-preserving: $\tau \circ E_\mathcal{B} = \tau$;
- $E_\mathcal{B}$ respects $*$-operation: $E_\mathcal{B}(a^*) = E_\mathcal{B}(a)^*$ for all $a \in \mathcal{A}$,
- let $P_\mathcal{B}$ denote the orthogonal projection between Hilbert (sub)spaces $L^2(\mathcal{A}) \to L^2(\mathcal{B})$, then the following diagram commutes:

$$\begin{array}{ccc} \mathcal{A} & \xrightarrow{\widehat{\ }} & L^2(\mathcal{A}) \\ \big\downarrow{\scriptstyle E_\mathcal{B}} & & \big\downarrow{\scriptstyle P_\mathcal{B}} \\ \mathcal{B} & \xrightarrow{\widehat{\ }} & L^2(\mathcal{B}) \end{array}$$

We call this map $E_\mathcal{B}$ the conditional expectation onto the non-commutative probability subspace $(\mathcal{B}, \tau|_\mathcal{B})$.

Next, we provide the definition of the NCPS inclusion map as an adjoint to a conditional expectation.

Definition 2.11 (NCPS embedding)**.** Given NCPSs $(\mathcal{A}, \tau_\mathcal{A})$ and $(\mathcal{C}, \tau_\mathcal{C})$, a NCPS embedding is a unital trace-preserving $*$-homomorphism $\iota_\mathcal{A} : \mathcal{A} \to \mathcal{C}$. More specifically:

- $\iota_\mathcal{A}(1_\mathcal{A}) = 1_\mathcal{C}$;
- $\tau_\mathcal{A} = \tau_\mathcal{C} \circ \iota_\mathcal{A}$;
- $\iota_\mathcal{A}(a^*) = \iota_\mathcal{A}(a)^*$ for all $a \in \mathcal{A}$;
- $\iota_\mathcal{A}$ is an algebra homomorphism.

We conclude this section with the following proposition relating conditional expectation and NCPS embedding ([GJNS21] Lemma 2.13) and provide its proof:

Proposition 2.1. *Any NCPS embedding $\iota_\mathcal{A} : \mathcal{A} \to \mathcal{C}$ extends to an isometry $L^2(\mathcal{A}) \to L^2(\mathcal{C})$, whose adjoint restricted to $\mathcal{C}$ is the conditional expectation $\mathcal{C} \to \mathcal{A}$.*

Proof. We prove $\iota_\mathcal{A}$ is an isometry $\mathcal{A} \to \mathcal{C}$, because $\iota_\mathcal{A}$ is a trace-preserving $*$-homomorphism:

$$\|\iota_\mathcal{A}(a)\|^2_{L^2(\mathcal{C})} = \tau_\mathcal{C}(\iota_\mathcal{A}(a)\iota_\mathcal{A}(a)^*) = \tau_\mathcal{C}(\iota_\mathcal{A}(a)\iota_\mathcal{A}(a^*)) = \tau_\mathcal{C}(\iota_\mathcal{A}(aa^*)) = \tau_\mathcal{A}(aa^*) = \|a\|^2_{L^2(\mathcal{A})}.$$

Therefore, $\iota_\mathcal{A}$ can extend uniquely to $L^2(\mathcal{A})$ since $\mathcal{A}$ is dense in $L^2(\mathcal{A})$. Finally, we note

$$\langle \iota_\mathcal{A}(a), c\rangle_{L^2(\mathcal{C})} = \langle a, P_\mathcal{A}(c)\rangle_{L^2(\mathcal{A})} \qquad \forall a \in L^2(\mathcal{A}), c \in L^2(\mathcal{C}).$$

Hence $\iota_\mathcal{A} = P^*_\mathcal{A}|_\mathcal{A}$. $\square$

2.1.1 Analogies to classical probability spaces

Recall that a classical probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set of outcomes, $\mathcal{F} \subset 2^\Omega$ is a sigma-algebra, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. We study (complex-valued) random variables as (Borel-)measurable functions $X : (\Omega, \mathcal{F}) \to \mathbb{C}$. We have, based on an integratablity condition, spaces of random variables $L^p(\Omega, \mathcal{F}, \mathbb{P})$, two of which are particularly interesting: $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a Hilbert space, and $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is a W^*-algebra.

In fact, a classical probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be recovered by the pair $(\mathcal{A}, \tau)$, where $\mathcal{A} = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ and $\tau(X) = \mathbb{E}[X] = \int_\Omega X \, d\mathbb{P}$. The sigma-algebra $\mathcal{F}$ is encoded as indicators 1_E, which are projections (self-adjoint idempotents) in the algebra $\mathcal{A}$, and set union (addition), complement (additive inversion), intersection (multiplication) also transfer, although do not coincide with the usual operations in $\mathcal{A}$. The probability measure $\mathbb{P}$ of an event in $\mathcal{F}$ is given by τ evaluated at the corresponding projection in $\mathcal{A}$.

One can easily verify that, after identifying the $*$-operation as complex conjugation, the pair $(\mathcal{A}, \tau)$ is a commutative version of a NCPS defined above. τ being unital comes from the classical axiom that requires a probability measure $\mathbb{P}$ to have total mass 1; τ being positive faithful comes from classical integration theory applied to the linear functional $\mathbb{E}$; and τ being weak-$*$ continuous is like the dominated convergence theorem in classical probability theory.

Therefore conversely, we can think of a (unital faithful normal) tracial W^*-algebra as an extension of the theory of classical probability spaces that allows multiplication in the algebra of random variables to be potentially non-commutative.

2.1.2 Example: matrix algebra with normalized trace

In addition to $(L^\infty, \mathbb{E})$ of a classical probability space being a commutative example of a NCPS, we provide a non-degenerate example that we study in this thesis: $(M_n(\mathbb{C}), \mathrm{tr})$, where $M_n(\mathbb{C})$ denotes the n by n matrices with complex entries and tr denotes the normalized trace $\frac{1}{n}\operatorname{Tr} : X \mapsto \frac{1}{n}\sum_{j=1}^{n} X_{jj}$. In this case the $*$-operation is given by conjugate transpose (adjoint). One can verify $M_n(\mathbb{C}) \cong \mathcal{B}(\mathbb{C}^n)$, and $M_n(\mathbb{C})$ is closed under WOT, so $M_n(\mathbb{C})$ is a W^*-algebra. One can also verify tr is a unital positive faithful trace. Hence $(M_n(\mathbb{C}), \mathrm{tr})$ is indeed a NCPS.

2.2 Coupling of random variables

In this section we will define coupling of random variables. In fact, the more precise terminology would be coupling a pair of laws realized by a pair of random variables having those laws, respectively.

2.2.1 Motivation: coupling of two real-valued random variables

We recall coupling in classical probability space to motivate our discussion of coupling in a NCPS. Intuitively, coupling takes two probability measures as the input and gives a probability measure on the product probability space as the output, whose marginals are the inputs.

Definition 2.12 (coupling of probability measures)**.** Given two probability measures μ and ν on $(\Omega, \mathcal{F})$, a coupling of μ and ν is a pair of random variables (X, Y) on $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ whose marginals are μ and ν, respectively.

The intuition we get from this definition is that: a coupling of a pair of random variables requires a common probability space (in this case the product space) with our pair of random variables sitting inside. Now that we have a common probability space, we can talk about some quantities using their joint relation obtained from the coupling.

Definition 2.13 (2-Wasserstein distance)**.** The 2-Wasserstein distance between two probability measures μ and ν is $d_W^{(2)}(\mu, \nu) = \inf\{\|X - Y\|_2 : (X, Y) \textit{ is a coupling of } \mu \textit{ and } \nu\}$.

In the literature of optimal transport theory, the 2-Wasserstein distance relates to Kantorovich's formulation of the optimal transportation problem (with cost function being the L^2 norm $\|\cdot\|_2$), which amounts to finding a probability measure on the product space that achieves the infinum above. More precisely, given a pair of laws, we want a common probability space and a joint law that achieves minimum cost. This motivates our formulation of optimal coupling in a NCPS.

2.2.2 Optimal coupling of two tuples of random variables in a NCPS

The non-commutative analogue of optimal coupling was first studied in [BV01] (Section 1.1).

Definition 2.14 (coupling of non-commutative laws)**.** Let μ and ν be non-commutative laws. A coupling of μ and ν is a triple $(\mathcal{C}, X, Y)$ where $(\mathcal{C}, \tau)$ is a NCPS and $X, Y \in L^\infty(\mathcal{C})^m_{\mathrm{sa}}$ have laws μ, ν, respectively.

Why can we require our random variables realizing the non-commutative laws to be self-adjoint? For general $X = (X_1, \dots, X_m) \in L^\infty(\mathcal{C})^m$, we can write $X_j = A_j + iB_j$ where $A_j = (X_j + X_j^*)/2$ and $B_j = (X_j - X_j^*)/2i$. The law ($*$-distribution) of each X_j is described by joint moments of X_j and X_j^*, so we can equivalently consider the tuple $X' = (A_1, \dots, A_m, B_1, \dots, B_m) \in L^\infty(\mathcal{C})^{2m}_{\mathrm{sa}}$, because the tuples $(X_1, \dots, X_m)$, $(X_1, \dots, X_m, X_1^*, \dots, X_m^*)$, and $(A_1, \dots, A_m, B_1, \dots, B_m)$ all generate the same W^*-algebra.

Definition 2.15 (non-commutative 2-Wasserstein distance)**.** The non-commutative 2-Wasserstein distance between two non-commutative laws is

$$d_W^{(2)}(\mu, \nu) = \inf \left\{ \|X - Y\|_{L^2(\mathcal{C})^m_{\mathrm{sa}}} : (\mathcal{C}, X, Y) \textit{ is a coupling of } (\mu, \nu) \right\}.$$

Definition 2.16 (optimal coupling)**.** A coupling $(\mathcal{C}, X, Y)$ of (μ, ν) is optimal if $\|X - Y\|_{L^2(\mathcal{C})^m_{\mathrm{sa}}} = d_W^{(2)}(\mu, \nu)$.

Here is how we understand optimally coupling tuples of random variables (instead of non-commutative laws) as an optimization problem from the definitions above: given a pair of random variables in a pair of NCPSs: $X \in L^\infty(\mathcal{A}, \tau_{\mathcal{A}})^m_{\mathrm{sa}}$ and $Y \in L^\infty(\mathcal{B}, \tau_{\mathcal{B}})^m_{\mathrm{sa}}$, we want a common NCPS $(\mathcal{C}, \tau_{\mathcal{C}})$ and NCPS embeddings $\iota_{\mathcal{A}} : \mathcal{A} \to \mathcal{C}$ and $\iota_{\mathcal{B}} : \mathcal{B} \to \mathcal{C}$ such that $\|\iota_{\mathcal{A}}(X) - \iota_{\mathcal{B}}(Y)\|_{L^2(\mathcal{C})^m_{\mathrm{sa}}}$ is minimized (over all possible common NCPSs and embeddings) and thus equals $d_W^{(2)}(\mathrm{Law}(X), \mathrm{Law}(Y))$.

Proposition 2.2. *Equivalently, we can maximize the real-valued quantity*

$$\langle \iota_{\mathcal{A}}(X), \iota_{\mathcal{B}}(Y) \rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}} = \sum_{j=1}^m \langle \iota_{\mathcal{A}}(X_j), \iota_{\mathcal{B}}(Y_j) \rangle_{L^2(\mathcal{C})_{\mathrm{sa}}}.$$

Proof. Note, because X_j, Y_j are self-adjoint and $\iota_{\mathcal{A}}, \iota_{\mathcal{B}}$ are trace-preserving $*$-homomorphisms,

$$\begin{aligned}
&\|\iota_{\mathcal{A}}(X) - \iota_{\mathcal{B}}(Y)\|^2_{L^2(\mathcal{C})^m_{\mathrm{sa}}} = \sum_{j=1}^m \|\iota_{\mathcal{A}}(X_j) - \iota_{\mathcal{B}}(Y_j)\|^2_{L^2(\mathcal{C})_{\mathrm{sa}}} \\
=&\sum_{j=1}^m \|\iota_{\mathcal{A}}(X_j)\|^2_{L^2(\mathcal{C})_{\mathrm{sa}}} - \langle \iota_{\mathcal{A}}(X_j), \iota_{\mathcal{B}}(Y_j) \rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} - \langle \iota_{\mathcal{B}}(Y_j), \iota_{\mathcal{A}}(X_j) \rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} + \|\iota_{\mathcal{B}}(Y_j)\|^2_{L^2(\mathcal{C})_{\mathrm{sa}}} \\
=&\sum_{j=1}^m \|X_j\|^2_{L^2(\mathcal{A})_{\mathrm{sa}}} - 2\langle \iota_{\mathcal{A}}(X_j), \iota_{\mathcal{B}}(Y_j) \rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} + \|Y_j\|^2_{L^2(\mathcal{B})_{\mathrm{sa}}} \\
=&\|X\|^2_{L^2(\mathcal{A})^m_{\mathrm{sa}}} + \|Y\|^2_{L^2(\mathcal{B})^m_{\mathrm{sa}}} - 2\sum_{j=1}^m \langle \iota_{\mathcal{A}}(X_j), \iota_{\mathcal{B}}(Y_j) \rangle_{L^2(\mathcal{C})_{\mathrm{sa}}},
\end{aligned}$$

and $\|X\|^2_{L^2(\mathcal{A})^m_{\mathrm{sa}}}$ and $\|Y\|^2_{L^2(\mathcal{B})^m_{\mathrm{sa}}}$ are constants depending on $\mathrm{Law}(X)$ and $\mathrm{Law}(Y)$, respectively. $\square$

Proposition 2.3. *Suppose $X \in L^\infty(\mathcal{A}, \tau_{\mathcal{A}})^m_{\mathrm{sa}}$ and $Y \in L^\infty(\mathcal{B}, \tau_{\mathcal{B}})^m_{\mathrm{sa}}$ are optimally coupled in a common NCPS $(\mathcal{C}, \tau_{\mathcal{C}})$. Let $a = (a_1, \dots, a_m), b = (b_1, \dots, b_m) \in \mathbb{R}^m$. Then $X + a$ and $Y + b$ are optimally coupled.*

Proof. Again, because X_j, Y_j are self-adjoint, $\iota_{\mathcal{A}}, \iota_{\mathcal{B}}$ are trace-preserving $*$-homomorphisms, and $\tau_{\mathcal{C}}$ is a unital linear functional,

$$\begin{aligned}
&\langle \iota_{\mathcal{A}}(X+a), \iota_{\mathcal{B}}(Y+b)\rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}} = \sum_{j=1}^m \langle \iota_{\mathcal{A}}(X_j+a_j), \iota_{\mathcal{B}}(Y_j+b_j)\rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} \\
=&\sum_{j=1}^m \langle \iota_{\mathcal{A}}(X_j)+a_j, \iota_{\mathcal{B}}(Y_j)+b_j\rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} \\
=&\sum_{j=1}^m \langle \iota_{\mathcal{A}}(X_j), \iota_{\mathcal{B}}(Y_j)\rangle_{L^2(\mathcal{C})_{\mathrm{sa}}} + \tau_{\mathcal{C}}(a_j \iota_{\mathcal{B}}(Y_j)^*) + \tau_{\mathcal{C}}(\iota_{\mathcal{A}}(X_j)\overline{b_j}) + \tau_{\mathcal{C}}(a_j\overline{b_j}) \\
=&\langle \iota_{\mathcal{A}}(X), \iota_{\mathcal{B}}(Y)\rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}} + \sum_{j=1}^m a_j \tau_{\mathcal{C}}(\iota_{\mathcal{B}}(Y_j^*)) + b_j\tau_{\mathcal{C}}(\iota_{\mathcal{A}}(X_j)) + a_j b_j \\
=&\langle \iota_{\mathcal{A}}(X), \iota_{\mathcal{B}}(Y)\rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}} + \sum_{j=1}^m a_j \tau_{\mathcal{B}}(Y_j) + b_j\tau_{\mathcal{A}}(X_j) + a_j b_j.
\end{aligned}$$

Note the summation over j on the last line is just a constant depending on X, Y and a, b, so maximizing $\langle \iota_{\mathcal{A}}(X+a), \iota_{\mathcal{B}}(Y+b)\rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}}$ is equivalent to maximizing $\langle \iota_{\mathcal{A}}(X), \iota_{\mathcal{B}}(Y)\rangle_{L^2(\mathcal{C})^m_{\mathrm{sa}}}$. By the previous Proposition 2.2, the statement is proved. □

2.3 Factorizable maps

Factorizable maps were first introduced and studied in [AD06] (Definition 2.6).

Definition 2.17 (factorizable map)**.** Let $(\mathcal{A}, \tau_{\mathcal{A}}), (\mathcal{B}, \tau_{\mathcal{B}})$ be NCPSs. A linear map $\Phi : \mathcal{A} \to \mathcal{B}$ is factorizable if there exists a NCPS $(\mathcal{C}, \tau_{\mathcal{C}})$ and NCPS embeddings $\iota_{\mathcal{A}} : \mathcal{A} \to \mathcal{C}$ and $\iota_{\mathcal{B}} : \mathcal{B} \to \mathcal{C}$ such that $\Phi = \iota_{\mathcal{B}}^* \circ \iota_{\mathcal{A}}$, where $\iota_{\mathcal{B}}^*$ is the conditional expectation adjoint to $\iota_{\mathcal{B}}$. We say Φ factorizes through $(\mathcal{C}, \tau_{\mathcal{C}})$ if there exists $\iota_{\mathcal{A}}$ and $\iota_{\mathcal{B}}$ as above.

Notation 2.18. We denote the space of factorizable maps from the NCPS $(\mathcal{A}, \tau_{\mathcal{A}})$ to $(\mathcal{B}, \tau_{\mathcal{B}})$ by $\mathcal{FM}(\mathcal{A}, \mathcal{B})$. We denote by $\mathcal{FM}_{\mathrm{mat}}(\mathcal{A}, \mathcal{B})$ the set of maps that factorize through a matrix algebra $\mathcal{C}$, denote by $\mathcal{FM}_{\mathrm{fin}}(\mathcal{A}, \mathcal{B})$ the set of maps that factorize through a finite-dimensional C^*-algebra $\mathcal{C}$, and denote by $\mathcal{FM}_{\mathcal{C}}(\mathcal{A}, \mathcal{B})$ the set of maps that factorize through the algebra $\mathcal{C}$. When $\mathcal{A}, \mathcal{B}$ refer to the same NCPS, we simply use the notation $\mathcal{FM}(\mathcal{A})$.

A simple yet non-trivial example of a factorizable map is $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C}), X \mapsto UXU^*$ where $U \in \mathrm{U}(n)$. In this case we start from NCPSs $(\mathcal{A}, \tau_{\mathcal{A}}) = (\mathcal{B}, \tau_{\mathcal{B}}) = (M_n(\mathbb{C}), \mathrm{tr}_n)$, and embed them into the same space $(\mathcal{C}, \tau_{\mathcal{C}}) = (M_n(\mathbb{C}), \mathrm{tr}_n)$. Using the notations introduced above, we say $\Phi \in \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C}))$. This Φ comes from the identity NCPS embedding and a $*$-isomorphism as the other NCPS embedding. We will rigorously see this in Proposition 3.1.

Proposition 2.4. *We summarize some facts about factorizable maps stated in [GJNS21] (Proposition 5.3), [MR19] (Theorem 4.1):*

- $\mathcal{FM}(\mathcal{A}, \mathcal{B})$ *and* $\mathcal{FM}_{\mathrm{fin}}(\mathcal{A}, \mathcal{B})$ *are convex sets;*
- $\mathcal{FM}_{\mathrm{mat}}(\mathcal{A}, \mathcal{B}) \subsetneq \mathcal{FM}_{\mathrm{fin}}(\mathcal{A}, \mathcal{B})$*, and* $\mathrm{conv}(\mathcal{FM}_{\mathrm{mat}}(\mathcal{A}, \mathcal{B})) = \mathcal{FM}_{\mathrm{fin}}(\mathcal{A}, \mathcal{B})$*;*
- $\mathcal{FM}_{\mathrm{fin}}(M_d(\mathbb{C}), M_d(\mathbb{C})) \subsetneq \overline{\mathcal{FM}_{\mathrm{fin}}(M_d(\mathbb{C}), M_d(\mathbb{C}))}^{\textit{weak-}*}$ *for dimensions* $d \geq 11$*;*
- $\mathcal{FM}(\mathcal{A}, \mathcal{B})$ *is weak-$*$ closed;*
- $\mathcal{FM}(\mathcal{A})$ *is closed under composition.*

The reason why factorizable maps are relevant and important to optimal coupling is the following: given a pair of random variables in a pair of NCPS: $X \in L^\infty(\mathcal{A}, \tau_{\mathcal{A}})^m_{\mathrm{sa}}$ and $Y \in L^\infty(\mathcal{B}, \tau_{\mathcal{B}})^m_{\mathrm{sa}}$, asking for an optimal coupling of X and Y is equivalent to finding $\Phi \in \mathcal{FM}(\mathcal{A}, \mathcal{B})$ such that

$$\langle \Phi(X), Y\rangle_{L^2(\mathcal{B})^m_{\mathrm{sa}}} = \sup_{\Psi \in \mathcal{FM}(\mathcal{A}, \mathcal{B})} \langle \Psi(X), Y\rangle_{L^2(\mathcal{B})^m_{\mathrm{sa}}}$$

by Proposition 2.2.

We conclude this section with a lemma that will be useful later.

Lemma 2.5. *If $\Phi \in \mathcal{FM}(\mathcal{A},\mathcal{B})$, then $\Phi(a^*) = \Phi(a)^*$ for all $a \in \mathcal{A}$. In particular, this means $a = a^*$ implies $\Phi(a) = \Phi(a)^*$.*

Proof. Write $\Phi = \iota_{\mathcal{B}}^* \circ \iota_{\mathcal{A}} = E_{\mathcal{B}} \circ \iota_{\mathcal{A}}$, where $\iota_{\mathcal{A}} : \mathcal{A} \to \mathcal{C}$ and $\iota_{\mathcal{B}} : \mathcal{B} \to \mathcal{C}$ are NCPS embeddings, and $E_{\mathcal{B}} : \mathcal{C} \to \mathcal{B}$ is the conditional expectation. Since conditional expectations respects $*$-operation and NCPS embeddings are $*$-homomorphisms, we have

$$\Phi(a^*) = E_{\mathcal{B}}(\iota_{\mathcal{A}}(a^*)) = E_{\mathcal{B}}(\iota_{\mathcal{A}}(a)^*) = E_{\mathcal{B}}(\iota_{\mathcal{A}}(a))^* = \Phi(a)^*,$$

and hence $a = a^*$ implies

$$\Phi(a) = \Phi(a^*) = \Phi(a)^*.$$

□

2.4 Miscellaneous notations

We supply a list of other notations/definitions used in this thesis.

Notation 2.19 (trace)**.** We denote by Tr the unnormalized trace and tr the normalized trace of a matrix algebra. In case of ambiguity, we will add a subscript to Tr or tr to distinguish traces of different matrix algebras.

Notation 2.20 (matrix transpose)**.** We denote by $M_n(\mathbb{C}) \to M_n(\mathbb{C}), a \mapsto a^t$ the matrix transpose $(a_{jk})_{j,k=1}^n \mapsto (a_{kj})_{j,k=1}^n$.

Notation 2.21 (bounded linear operators)**.** Let V be a normed vector space. We denote by $\mathcal{B}(V) = \{T \textit{ is a linear transformation } V \to V : \|T\|_{\text{op}} < \infty\}$ the bounded linear operators on V.

Notation 2.22 (commutant)**.** Let $\mathcal{A}$ be an algebra, and $B \subseteq \mathcal{A}$ be a subset. We denote by $B' = \{a \in \mathcal{A} : ab = ba \quad \forall b \in B\}$ the commutant of B. We denote by $B'' = (B')'$ the double commutant of B.

Notation 2.23 (automorphism group)**.** We denote by $\text{Aut}(\mathcal{A})$ the automorphism group of an algebra $\mathcal{A}$, which contains all the isomorphisms $\mathcal{A} \to \mathcal{A}$. In the case when $\mathcal{A}$ is a $*$-algebra, we also require the algebra isomorphisms to respect the $*$-operation.

Notation 2.24 (convex hull)**.** We denote by $\text{conv}(A)$ the convex hull of a subset A inside a vector space V, which contains all the convex combinations $c_1a_1 + \cdots + c_na_n$ such that $n \in \mathbb{N}$, $c_1, \ldots, c_n \geq 0$, $c_1 + \cdots + c_n = 1$, and $a_1, \ldots, a_n \in A$.

Notation 2.25. Let $\mathcal{A}$ be a $*$-algebra, and $u \in \mathcal{A}$. We denote by Ad_u the map $\mathcal{A} \to \mathcal{A}, x \mapsto uxu^*$.

Notation 2.26 (unitary group)**.** We denote by $\text{U}(n) = \{U \in M_n(\mathbb{C}) : UU^* = 1_n = U^*U\}$ the unitary group on $M_n(\mathbb{C})$.

Notation 2.27 (special unitary group)**.** We denote by $\text{SU}(n) = \{U \in M_n(\mathbb{C}) : UU^* = 1_n = U^*U, \det(U) = 1\}$ the special unitary group on $M_n(\mathbb{C})$.

Notation 2.28 (special orthogonal group)**.** We denote by $\text{SO}(n) = \{U \in M_n(\mathbb{R}) : UU^t = 1_n = U^tU, \det(U) = 1\}$ the special orthogonal group on $M_n(\mathbb{R})$.

Notation 2.29 (matrix units)**.** We denote by e_{jk} the matrix in $M_n(\mathbb{C})$ that has 1 at entry (j,k) and 0 at all other entries. We denote by

$$\mathcal{B}_n = \{e_{jk}\}_{1\leq j,k\leq n}$$

an orthonormal basis of $M_n(\mathbb{C})$ as a $\mathbb{C}$-vector space, and we denote by

$$\mathcal{B}_{n,\text{sa}} = \{e_{jj}\}_{1\leq j\leq n} \cup \{\frac{1}{\sqrt{2}}(e_{jk} + e_{kj})\}_{1\leq j<k\leq n} \cup \{\frac{i}{\sqrt{2}}(e_{jk} - e_{kj})\}_{1\leq j<k\leq n}$$

an orthonormal basis of $M_n(\mathbb{C})_{\text{sa}}$ as a $\mathbb{R}$-vector space (both equipped with the inner product induced by Tr_n).

3 Optimal couplings and unitary conjugation

In this section, we always assume we are given fixed $X = (X_1, \dots, X_m), Y = (Y_1, \dots, Y_m) \in M_n(\mathbb{C})^m_{\mathrm{sa}}$. Recall the simple example we gave in Definition 2.17 of factorizable maps: $\mathrm{Ad}_U : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ where $U \in \mathrm{U}(n)$. In fact, they are the only kind of $\Phi \in \mathcal{FM}(M_n(\mathbb{C}))$ that factors through the same algebra $M_n(\mathbb{C})$. We will first demonstrate why this is true, and then use this proposition to study optimal couplings given by unitary conjugation.

Proposition 3.1. *The automorphisms on $M_n(\mathbb{C})$ are completely described by unitary conjugation on $M_n(\mathbb{C})$. Moreover, they are exactly the factorizable maps $M_n(\mathbb{C}) \to M_n(\mathbb{C})$ that factorize through $M_n(\mathbb{C})$:*

$$\{\mathrm{Ad}_U : U \in \mathrm{U}(n)\} = \mathrm{Aut}(M_n(\mathbb{C})) = \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C})).$$

Proof. We first show $\{\mathrm{Ad}_U : U \in \mathrm{U}(n)\} \subseteq \mathrm{Aut}(M_n(\mathbb{C}))$. Fix $U \in \mathrm{U}(n)$, then for arbitrary $A, B \in M_n(\mathbb{C})$,

$$\mathrm{Ad}_U(A+B) = U(A+B)U^* = UAU^* + UBU^* = \mathrm{Ad}_U(A) + \mathrm{Ad}_U(B),$$
$$\mathrm{Ad}_U(AB) = U(AB)U^* = (UAU^*)(UBU^*) = \mathrm{Ad}_U(A)\,\mathrm{Ad}_U(B),$$
$$\mathrm{Ad}_U(A^*) = UA^*U^* = (UAU^*)^* = \mathrm{Ad}_U(A)^*,$$

and Ad_U is bijective with inverse Ad_{U^*}, so Ad_U is a $*$-isomorphism $M_n(\mathbb{C}) \to M_n(\mathbb{C})$. Hence $\mathrm{Ad}_U \in \mathrm{Aut}(M_n(\mathbb{C}))$.

Conversely, fix $\rho \in \mathrm{Aut}(M_n(\mathbb{C}))$. We quote the Skolem–Noether theorem to conclude that ρ is inner, so $\rho(X) = PXP^{-1}$ for some invertible $P \in M_n(\mathbb{C})$. ρ respects the $*$-operation, so

$$PX^*P^{-1} = \rho(X^*) = \rho(X)^* = (PXP^{-1})^* = (P^{-1})^*X^*P^* = (P^*)^{-1}X^*P^*,$$

where in the last equality, complex conjugation commutes with inverse because

$$P^*(P^{-1})^* = (P^{-1}P)^* = 1_n^* = 1_n = 1_n^* = (PP^{-1})^* = (P^{-1})^*P^*.$$

Hence for all $X \in M_n(\mathbb{C})$, $PX^*P^{-1} = (P^*)^{-1}X^*P^*$, so $P^*PX^* = X^*P^*P$. From this we conclude $P^*P \in M_n(\mathbb{C})'$. We have $M_n(\mathbb{C})' = \mathbb{C}1_n$ because by direct computation, if $D \in M_n(\mathbb{C})$ satisfies (by linearity it suffices to hold for matrix units e_{jk}) $De_{jk} = e_{jk}D$, then D can only have non-zero entries in its diagonal and $[D]_{jj} = [D]_{kk}$. By iterating through all different pairs $1 \le j \ne k \le n$, we have $D = \lambda 1_n$ for some $\lambda \in \mathbb{C}$. Moreover, we can assume $P^*P = \lambda 1_n$ for some $\lambda \ge 0$ because $\lambda = \mathrm{tr}_n(\lambda 1_n) = \mathrm{tr}_n(P^*P) \ge 0$ since tr_n is positive. In addition, we have $\lambda \ne 0$ because if otherwise $P^*P = 0$, then P has to be the zero matrix, which is not invertible, contradictory to our assumption. Finally, we let $Q = \frac{1}{\sqrt{\lambda}}P$, then $Q^{-1} = \sqrt{\lambda}P^{-1}$, and $Q^* = \frac{1}{\sqrt{\lambda}}P^* = \frac{1}{\sqrt{\lambda}}\lambda P^{-1}$. Putting this together, $\rho(X) = PXP^{-1} = QXQ^{-1}$ with $Q^*Q = 1_n = QQ^*$, so $\rho = \mathrm{Ad}_Q$ with $Q \in \mathrm{U}(n)$.

Then we show $\mathrm{Aut}(M_n(\mathbb{C})) \subseteq \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C}))$. Fix $\rho \in \mathrm{Aut}(M_n(\mathbb{C}))$. Note ρ and the identity map $\mathrm{id}_{M_n(\mathbb{C})}$ are both NCPS embeddings $M_n(\mathbb{C}) \to M_n(\mathbb{C})$, and the conditional expectation adjoint $\mathrm{id}^*_{M_n(\mathbb{C})} = \mathrm{id}_{M_n(\mathbb{C})}$. Hence $\rho = \mathrm{id}_{M_n(\mathbb{C})} \circ \rho = \mathrm{id}^*_{M_n(\mathbb{C})} \circ \rho \in \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C}))$.

Conversely, fix $\Phi \in \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C}))$. By definition, $\Phi = \rho_2^* \circ \rho_1$, where ρ_1, ρ_2 are NCPS embeddings $M_n(\mathbb{C}) \to M_n(\mathbb{C})$. By Proposition 2.1, ρ_1, ρ_2 are isometric, so in particular they are injective. Since $M_n(\mathbb{C})$ is finite dimensional, the injective $*$-homomorphisms ρ_1, ρ_2 are thus isomorphisms $M_n(\mathbb{C}) \to M_n(\mathbb{C})$. By our previous results, we can assume $\rho_1 = \mathrm{Ad}_{U_1}$ and $\rho_2 = \mathrm{Ad}_{U_2}$ for some $U_1, U_2 \in \mathrm{U}(n)$. For arbitrary $X, Y \in M_n(\mathbb{C})$,

$$\langle \Phi(X), Y \rangle = \langle \rho_1(X), \rho_2(Y) \rangle = \langle U_1XU_1^*, U_2YU_2^* \rangle = \langle U_2^*U_1XU_1^*U_2, Y \rangle = \left\langle \mathrm{Ad}_{U_2^*U_1}(X), Y \right\rangle.$$

Hence $\Phi = \mathrm{Ad}_{U_2^*U_1} \in \mathrm{Aut}(M_n(\mathbb{C}))$. ◻

By the previous proposition, the reduced problem we are considering in this section is finding $U \in \mathrm{U}(n)$ that achieves (because $\mathrm{U}(n)$ is compact and $U \mapsto \langle UXU^*, Y \rangle_{\mathrm{tr}}$ is continuous in U) the supremum:

$$\sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} = \sup_{U \in \mathrm{U}(n)} \sum_{j=1}^m \langle UX_jU^*, Y_j \rangle_{\mathrm{tr}}.$$

Proposition 3.2. *Moreover, we have the following:*

$$\sup_{\Phi \in \operatorname{conv}(\operatorname{Aut}(M_n(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} = \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} = \sup_{U \in \mathrm{SU}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}.$$

Proof. Fix $\Phi \in \operatorname{conv}(\operatorname{Aut}(M_n(\mathbb{C})))$, write $\Phi = \sum_{j=1}^k c_j \operatorname{Ad}_{U_j}$ where $k \in \mathbb{N}$, $\sum_{j=1}^k c_j = 1$, and $U_1, \dots, U_k \in \mathrm{U}(n)$. Then,

$$\begin{aligned}\langle \Phi(X), Y \rangle_{\mathrm{tr}} &= \left\langle \sum_{j=1}^k c_j \operatorname{Ad}_{U_j}(X), Y \right\rangle_{\mathrm{tr}} = \left\langle \sum_{j=1}^k c_j U_j X U_j^*, Y \right\rangle_{\mathrm{tr}} = \sum_{j=1}^k c_j \langle U_j X U_j^*, Y \rangle_{\mathrm{tr}} \\ &\le \sum_{j=1}^k c_j \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} = \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} \sum_{j=1}^k c_j = \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}.\end{aligned}$$

This implies

$$\sup_{\Phi \in \operatorname{conv}(\operatorname{Aut}(M_n(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} \le \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}.$$

Conversely, let $Q \in \mathrm{U}(n)$ be the maximizer of $\langle UXU^*, Y \rangle_{\mathrm{tr}}$. Then $\operatorname{Ad}_Q \in \operatorname{conv}(\operatorname{Aut}(M_n(\mathbb{C})))$, so

$$\sup_{\Phi \in \operatorname{conv}(\operatorname{Aut}(M_n(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} \ge \sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}.$$

For arbitrary $U \in \mathrm{U}(n)$, let $\lambda = \det(U) \neq 0$ because U is invertible. Now consider $Q = \lambda^{-1/n} U$, then

$$\det(Q) = \det(\lambda^{-1/n} U) = (\lambda^{-1/n})^n \det(U) = \lambda^{-1} \lambda = 1.$$

We also have

$$|\lambda|^2 = \lambda \overline{\lambda} = \det(U) \overline{\det(U)} = \det(U) \det(\overline{U}) = \det(U) \det(\overline{U}^t) = \det(UU^*) = \det(1_n) = 1.$$

Writing $\lambda = e^{i\theta}$, we have

$$QQ^* = e^{-i\theta/n} U e^{i\theta/n} U^* = 1_n = e^{i\theta/n} U^* e^{-i\theta/n} U = Q^* Q.$$

Hence $Q \in \mathrm{SU}(n)$ with $\langle QXQ^*, Y \rangle_{\mathrm{tr}} = \left\langle e^{-i\theta/n} U X e^{i\theta/n} U^*, Y \right\rangle_{\mathrm{tr}} = \langle UXU^*, Y \rangle_{\mathrm{tr}}$, and this implies

$$\sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} \le \sup_{U \in \mathrm{SU}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}.$$

The converse

$$\sup_{U \in \mathrm{U}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}} \ge \sup_{U \in \mathrm{SU}(n)} \langle UXU^*, Y \rangle_{\mathrm{tr}}$$

is clear because $\mathrm{SU}(n) \subseteq \mathrm{U}(n)$. $\square$

3.1 Reduction to rank-one matrices

We have two tools to further reduce the problem for easier computations and possible connections to vector geometry ideas. One tool is diagonalizing the self-adjoint X_j, Y_j, and the other is shifting by elements in $\mathbb{R}1_n$ (Proposition 2.3).

If we consider the spectral decomposition of $X_j = V_j A_j V_j^*$ and $Y_j = W_j B_j W_j^*$, we will have

$$V_j = \begin{bmatrix} v_{j1} & \cdots & v_{jn} \end{bmatrix}, \ A_j = \begin{bmatrix} \alpha_{j1} & & \\ & \ddots & \\ & & \alpha_{jn} \end{bmatrix},$$

where $\{v_{j1}, \dots, v_{jn}\}$ form an orthonormal basis of $\mathbb{C}^n$ and $\alpha_{j1}, \dots, \alpha_{jn} \in \mathbb{R}$ (thanks to X_j being self-adjoint); similarly for

$$W_j = \begin{bmatrix} w_{j1} & \cdots & w_{jn} \end{bmatrix}, \ B_j = \begin{bmatrix} \beta_{j1} & & \\ & \ddots & \\ & & \beta_{jn} \end{bmatrix}.$$

Therefore, we have the following:

$$X_j = V_j A_j V_j^* = \sum_{k=1}^n \alpha_{jk} v_{jk} v_{jk}^*, \qquad Y_j = W_j B_j W_j^* = \sum_{l=1}^n \beta_{jl} w_{jl} w_{jl}^*,$$

$$\begin{aligned}\langle UXU^*, Y\rangle_{\mathrm{tr}} &= \sum_{j=1}^m \operatorname{tr}(UX_jU^*Y_j) = \sum_{j=1}^m \operatorname{tr}(U(\sum_{k=1}^n \alpha_{jk} v_{jk} v_{jk}^*)U^*(\sum_{l=1}^n \beta_{jl} w_{jl} w_{jl}^*)) \\ &= \sum_{j=1}^m \sum_{k,l=1}^n \alpha_{jk}\beta_{jl} \operatorname{tr}(Uv_{jk}v_{jk}^*U^*w_{jl}w_{jl}^*) = \sum_{j=1}^m \sum_{k,l=1}^n \alpha_{jk}\beta_{jl} \operatorname{tr}((Uv_{jk})^*w_{jl}w_{jl}^*(Uv_{jk})) \\ &= \sum_{j=1}^m \sum_{k,l=1}^n \alpha_{jk}\beta_{jl} \langle w_{jl}, Uv_{jk}\rangle\langle Uv_{jk}, w_{jl}\rangle \operatorname{tr}(1_n) = \sum_{j=1}^m \sum_{k,l=1}^n \alpha_{jk}\beta_{jl} \overline{\langle Uv_{jk}, w_{jl}\rangle}\langle Uv_{jk}, w_{jl}\rangle \\ &= \sum_{j=1}^m \sum_{k,l=1}^n \alpha_{jk}\beta_{jl} |\langle Uv_{jk}, w_{jl}\rangle|^2.\end{aligned}$$

Note that the inner product in the last equality is just the usual inner product in $\mathbb{C}^n$.

We can combine these two tools. First we add real multiples of 1_n to each X_j and Y_j to make their eigenvalues strictly positive. Then we rename $v_{jk} = \sqrt{\alpha_{jk}} v_{jk}$ and $w_{jl} = \sqrt{\beta_{jl}} w_{jl}$ (this makes them no longer an orthonormal basis of $\mathbb{C}^n$ but still an orthogonal basis). Hence it suffices to optimize

$$\langle UXU^*, Y\rangle_{\mathrm{tr}} = \sum_{j=1}^m \sum_{k,l=1}^n |\langle Uv_{jk}, w_{jl}\rangle|^2.$$

3.2 Details of case $n = 2$

By the discussion in the previous section, we let $a, b \in \mathbb{R}^m$ where a_j and b_j are the smaller eigenvalue of X_j and Y_j, respectively. Then we rename $X = X - a$ and $Y = Y - b$, so X and Y are rank-one matrix tuples. Hence it suffices to consider $X = (x_1x_1^*, \ldots, x_mx_m^*)$ and $Y = (y_1y_1^*, \ldots, y_my_m^*)$, where $0 \neq x_j, y_j \in \mathbb{C}^2$. Then similar to the computation we did before, we have the following:

$$\langle UXU^*, Y\rangle_{\mathrm{tr}} = \sum_{j=1}^m \langle Ux_jx_j^*U^*, y_jy_j^*\rangle_{\mathrm{tr}} = \sum_{j=1}^m \operatorname{tr}(Ux_jx_j^*U^*y_jy_j^*) = \sum_{j=1}^m |\langle Ux_j, y_j\rangle|^2.$$

3.2.1 Conjugating SO(2) in $M_2(\mathbb{R})^m_{\mathrm{sa}}$

Perhaps conjugating SU(2) in $M_2(\mathbb{C})^m_{\mathrm{sa}}$ is still harder to visualize and compute, because an element in SU(2) involves three real parameters. However, more explicit computation and intuitive conclusion are possible in the real analogue when we conjugate SO(2) in $M_2(\mathbb{R})^m_{\mathrm{sa}}$.

In addition to the simplification above, we have the following lemma because SO(2) is abelian.

Lemma 3.3. *In this case, it suffices to assume $x_1 = \cdots = x_m = e_1 \in \mathbb{R}^2$.*

Proof. For each x_j, there exists $V_j \in \mathrm{SO}(2)$ such that $x_j = \|x_j\|V_je_1$. (Set the first column in V_j to be $x_j/\|x_j\|$, and then, for example, do Gram–Schmidt process). Now:

$$\begin{aligned}\langle UXU^*, Y\rangle_{\mathrm{tr}} &= \sum_{j=1}^m |\langle Ux_j, y_j\rangle|^2 = \sum_{j=1}^m |\langle U(\|x_j\|V_je_1), y_j\rangle|^2 \\ &= \sum_{j=1}^m |\langle \|x_j\|V_jUe_1, y_j\rangle|^2 = \sum_{j=1}^m |\langle Ue_1, \|x_j\|V_j^*y_j\rangle|^2\end{aligned}$$

Hence it suffices to rename $y_j = \|x_j\|V_j^*y_j$ and assume $x_j = e_1$. $\square$

We use the following bijective parameterization to do the computation explicitly: for each $1 \leq j \leq m$, let $x_j = e_1$, $y_j = c_j \begin{bmatrix} \cos\alpha_j \\ \sin\alpha_j \end{bmatrix}$, and $U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, with $\alpha_j, \theta \in [0, 2\pi]$, $c_j > 0$. Therefore,

$$\begin{aligned}
\langle UXU^*, Y\rangle_{\mathrm{tr}} &= \sum_{j=1}^m |\langle Ux_j, y_j\rangle|^2 = \sum_{j=1}^m c_j^2(\cos\theta\cos\alpha_j + \sin\theta\sin\alpha_j)^2 = \sum_{j=1}^m c_j^2\cos^2(\theta-\alpha_j) \\
=& \sum_{j=1}^m c_j^2 \frac{\cos(2\theta - 2\alpha_j) + 1}{2} = \frac{1}{2}\sum_{j=1}^m c_j^2(\cos 2\theta\cos 2\alpha_j + \sin 2\theta \sin 2\alpha_j + 1) \\
=& A\cos 2\theta + B\sin 2\theta + C_2,
\end{aligned}$$

for constants

$$A = \frac{1}{2}\sum_{j=1}^m c_j^2\cos 2\alpha_j, \quad B = \frac{1}{2}\sum_{j=1}^m c_j^2 \sin 2\alpha_j, \quad C_2 = \frac{1}{2}\sum_{j=1}^m c_j^2.$$

If $A \neq 0$ or $B \neq 0$, denote $C_1 = \sqrt{A^2+B^2}$, and let $\phi \in [0, 2\pi]$ be such that $\sin\phi = \frac{A}{\sqrt{A^2+B^2}}$ and $\cos\phi = \frac{B}{\sqrt{A^2+B^2}}$. Then,

$$\begin{aligned}
\langle UXU^*, Y\rangle_{\mathrm{tr}} &= C_1(\sin\phi\cos 2\theta + \cos\phi\sin 2\theta) + C_2 \\
&= C_1\sin(2\theta+\phi) + C_2.
\end{aligned}$$

The computation above reduces maximizing $U \mapsto \langle UXU^*, Y\rangle_{\mathrm{tr}}$ over $U \in \mathrm{SO}(2)$ to maximizing $\theta \mapsto A\cos 2\theta + B\sin 2\theta + C_2$ (π-periodic) over $\theta \in [0, 2\pi]$. Hence we can conclude, in the non-degenerate case ($A \neq 0$ or $B \neq 0$), we will have exactly two maximizers $0 \leq \theta_{max} < \theta_{max} + \pi < 2\pi$, and $U = \pm \begin{bmatrix} \cos\theta_{max} & -\sin\theta_{max} \\ \sin\theta_{max} & \cos\theta_{max} \end{bmatrix}$. In the degenerate case ($A = 0$ and $B = 0$), $\langle UXU^*, Y\rangle_{\mathrm{tr}}$ is constant, and any $U \in \mathrm{SO}(2)$ is a maximizer.

More intuitively, when $m = 2$ and $c_1 = c_2 = c$, the problem is equivalent to maximizing $P(\theta) = \cos^2(\alpha_1 - \theta) + \cos^2(\alpha_2 - \theta)$ over $[0, 2\pi]$. Setting $dP/d\theta = 0$ and $d^2P/d\theta^2 < 0$, we have the following:

$$\begin{cases} \sin(2\alpha_1 - 2\theta_{max}) + \sin(2\alpha_2 - 2\theta_{max}) = 0 \\ \cos(2\alpha_1 - 2\theta_{max}) + \cos(2\alpha_2 - 2\theta_{max}) > 0 \end{cases}$$

The solutions will be $\theta_{max} = \alpha_1/2 + \alpha_2/2$ and $\theta_{max} + \pi$ or $\theta_{max} - \pi$. Geometrically, the corresponding $U \in \mathrm{SO}(2)$ will send e_1 to be the angle bisector of y_1 and y_2.

3.3 Parameterization of SU(2) and SU(3)

At this point, we also want to use Matlab to numerically compute $\sup_{U\in\mathrm{SU}(n)}\langle UXU^*, Y\rangle_{\mathrm{tr}}$ for concrete examples of matrix tuples $X, Y \in M_n(\mathbb{C})^m_{\mathrm{sa}}$, at least for $n = 2, 3$. To achieve this, we need to parameterize SU(2) and SO(2) by real variables.

In [Ham15] (Section IV), the author describes a bijective parameterization of SU(2):

$$U = \begin{bmatrix} e^{-i\alpha}\cos\theta & -e^{i\beta}\sin\theta \\ e^{-i\beta}\sin\theta & e^{i\alpha}\cos\theta \end{bmatrix},$$

where $(\alpha, \beta, \theta) \in D = D_1 \cup D_2 \cup D_3$, with

$$D_1 = [-\pi, \pi) \times [-\pi, \pi) \times (0, \pi/2), D_2 = [-\pi, \pi) \times \{0\} \times \{0\}, D_3 = \{0\} \times [-\pi, \pi) \times \{\pi/2\}.$$

In [Bro88] (Section II), the author describes a parameterization of SU(3):

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{bmatrix},$$

where

$$\begin{aligned}
u_{11} &= \cos\theta_1 \cos\theta_2 e^{i\phi_1}, \\
u_{12} &= \sin\theta_1 e^{i\phi_3}, \\
u_{13} &= \cos\theta_1 \sin\theta_2 e^{i\phi_4}, \\
u_{21} &= \sin\theta_2 \sin\theta_3 e^{-i(\phi_4+\phi_5)} - \sin\theta_1 \cos\theta_2 \cos\theta_3 e^{i(\phi_1+\phi_2-\phi_3)}, \\
u_{22} &= \cos\theta_1 \cos\theta_3 e^{i\phi_2}, \\
u_{23} &= -\cos\theta_2 \sin\theta_3 e^{-i(\phi_1+\phi_5)} - \sin\theta_1 \sin\theta_2 \cos\theta_3 e^{i(\phi_2-\phi_3+\phi_4)}, \\
u_{31} &= -\sin\theta_1 \cos\theta_2 \sin\theta_3 e^{i(\phi_1-\phi_3+\phi_5)} - \sin\theta_2 \cos\theta_3 e^{-i(\phi_2+\phi_4)}, \\
u_{32} &= \cos\theta_1 \sin\theta_3 e^{i\phi_5}, \\
u_{33} &= \cos\theta_2 \cos\theta_3 e^{-i(\phi_1+\phi_2)} - \sin\theta_1 \sin\theta_2 \sin\theta_3 e^{-i(\phi_3-\phi_4-\phi_5)},
\end{aligned}$$

with $0 \le \theta_1, \theta_2, \theta_3 \le \pi/2$, $0 \le \phi_1, \ldots, \phi_5 \le 2\pi$.

4 Optimally coupling matrix tuples in a larger algebra

In this section, we remove the requirement that $\Phi \in \mathcal{FM}_{M_n(\mathbb{C})}(M_n(\mathbb{C}))$. In fact, we are looking for examples of finite-dimensional matrix tuples whose optimal coupling requires embedding into a matrix algebra of larger dimension, or even an infinite dimensional W^*-algebra. In other words, we want to find matrix tuples $X, Y \in M_n(\mathbb{C})^m_{\mathrm{sa}}$ such that

$$\sup_{\Phi \in \mathrm{conv}(\mathrm{Aut}(M_n(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} < \sup_{\Phi \in \mathcal{FM}_{\mathrm{mat}}(M_n(\mathbb{C}))} \langle \Phi(X), Y \rangle_{\mathrm{tr}},$$

or even

$$\sup_{\Phi \in \mathcal{FM}_{\mathrm{fin}}(M_n(\mathbb{C}))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} < \sup_{\Phi \in \mathcal{FM}(M_n(\mathbb{C}))} \langle \Phi(X), Y \rangle_{\mathrm{tr}}.$$

We will start with cases $n = 2$ and $n = 3$. They are easier to analyze and to compute in an optimization program. It turns out that they already produce some interesting results.

4.1 The case $n = 2$

By Theorem 5.10 in [Kü85], maps in $\mathcal{FM}(M_2(\mathbb{C}))$ fall in a larger class of maps (unital completely positive trace-preserving maps) given by an average of unitary conjugation in the sense of integrating against some measure. Hence $\mathcal{FM}(M_2(\mathbb{C})) \subseteq \overline{\mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C})))}$. Note $\mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C})))$ is already closed in the norm topology on $\mathcal{B}(M_2(\mathbb{C}))$, since $\mathcal{B}(M_2(\mathbb{C}))$ is finite-dimensional, by an argument in [HM11] (in between Corollary 2.3 and Proposition 2.4). Therefore $\mathcal{FM}(M_2(\mathbb{C})) \subseteq \mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C})))$.

Moreover, we showed in Proposition 3.1 $\mathrm{Aut}(M_2(\mathbb{C})) = \mathcal{FM}_{M_2(\mathbb{C})}(M_2(\mathbb{C})) \subseteq \mathcal{FM}(M_2(\mathbb{C}))$, and by Proposition 2.4 $\mathcal{FM}(M_2(\mathbb{C}))$ is convex, so $\mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C}))) \subseteq \mathcal{FM}(M_2(\mathbb{C}))$. Hence we have

$$\mathcal{FM}(M_2(\mathbb{C})) = \mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C}))).$$

This means for matrix tuples $X, Y \in M_2(\mathbb{C})^m_{\mathrm{sa}}$, their optimal coupling is always achieved by conjugating some unitary matrix $U \in \mathrm{SU}(2)$ because, by Proposition 3.2,

$$\sup_{\Phi \in \mathcal{FM}(M_2(\mathbb{C}))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} = \sup_{\Phi \in \mathrm{conv}(\mathrm{Aut}(M_2(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} = \sup_{U \in \mathrm{SU}(2)} \langle UXU^*, Y \rangle_{\mathrm{tr}},$$

which can be directly optimized in a Matlab program.

4.2 The case $n = 3$

In this case, there exist matrix tuples $X, Y \in M_3(\mathbb{C})^m_{\mathrm{sa}}$ such that

$$\sup_{\Phi \in \mathrm{conv}(\mathrm{Aut}(M_3(\mathbb{C})))} \langle \Phi(X), Y \rangle_{\mathrm{tr}} < \sup_{\Phi \in \mathcal{FM}_{\mathrm{mat}}(M_3(\mathbb{C}))} \langle \Phi(X), Y \rangle_{\mathrm{tr}}.$$

We will provide a concrete example of $X, Y \in M_3(\mathbb{C})_{\text{sa}}^9$ based on ideas from Lemma 5.7 in [GJNS21]. The construction can be summarized as follows:

We use a map $\Phi \in \mathcal{FM}_{\text{mat}}(M_3(\mathbb{C})) \setminus \text{conv}(\text{Aut}(M_3(\mathbb{C})))$. Note $\text{conv}(\text{Aut}(M_3(\mathbb{C})))$ is closed in the norm topology on $\mathcal{B}(M_3(\mathbb{C}))$ for the same reason as in the case $n = 2$. Now we consider the real vector space $\mathcal{H} = L_{\mathbb{R}}(M_3(\mathbb{C})_{\text{sa}}, M_3(\mathbb{C})_{\text{sa}})$, the set of $\mathbb{R}$-linear transformations $M_3(\mathbb{C})_{\text{sa}} \to M_3(\mathbb{C})_{\text{sa}}$. In particular, $\mathcal{H}$ contains maps in $\mathcal{FM}(M_3(\mathbb{C}))$ restricted to $M_3(\mathbb{C})_{\text{sa}}$, because factorizable maps will take self-adjoint elements to self-adjoint elements by Lemma 2.5. Moreover, we equip $\mathcal{H}$ with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ induced by the trace

$$\text{Tr}_{\mathcal{H}}(\Psi) = \sum_{b \in \mathcal{B}_{3,\text{sa}}} \langle \Psi b, b \rangle_{\text{Tr}_3}.$$

This makes $\mathcal{K} = \text{conv}(\text{Aut}(M_3(\mathbb{C})))$ (also restricted to $M_3(\mathbb{C})_{\text{sa}}$) a closed convex subset of the topological vector space $\mathcal{H}$. Then we find the projection of $\Phi \in \mathcal{H}$ onto $\mathcal{K}$ in light of the Hilbert projection theorem (see A.1). Now we are at a good position to apply the Hahn-Banach separation theorem (see A.3), which will give us a strict inequality, because we are separating a closed convex set and a point outside. Finally, we make use of a linear isomorphism $\rho : \mathcal{H} \to (M_3(\mathbb{C})_{\text{sa}} \otimes_{\mathbb{R}} M_3(\mathbb{C})_{\text{sa}})^*$ to obtain our desired matrix tuples.

We first write down relevant definitions. We will use those to find a map

$$\Phi \in \mathcal{FM}_{\text{mat}}(M_3(\mathbb{C})) \setminus \text{conv}(\text{Aut}(M_3(\mathbb{C})))$$

and its projection onto $\mathcal{K}$.

Definition 4.1 (Holevo–Werner channels)**.** ([HM15] Equation 4.10 and 4.11) For each integer $n \geq 2$, the Holevo–Werner channels $W_n^+, W_n^- \in \mathcal{B}(M_n(\mathbb{C}))$ are defined to be:

$$W_n^+(x) = \frac{1}{n+1}\Big(\text{Tr}_n(x)1_n + x^t\Big), \qquad W_n^-(x) = \frac{1}{n-1}\Big(\text{Tr}_n(x)1_n - x^t\Big), \qquad x \in \mathcal{B}(M_n(\mathbb{C})).$$

Alternatively, they can be expressed as:

$$W_n^+(x) = \frac{1}{2n+2}\sum_{j,k=1}^{n}(e_{jk} + e_{kj})x(e_{jk} + e_{kj})^*, \qquad W_n^-(x) = \frac{1}{2n-2}\sum_{j,k=1}^{n}(e_{jk} - e_{kj})x(e_{jk} - e_{kj})^*.$$

The word channel in this definition refers to quantum channel in quantum information theory, which is a communication channel transmitting quantum information. Mathematically, quantum channels are completely positive trace-preserving maps between two operator algebras. We summarize some useful facts about Holevo–Werner channels stated in [HM15]:

Proposition 4.1.

- *[HM15, Corollary 5.7]: For $0 \leq \lambda \leq 1$, $\lambda W_3^+ + (1-\lambda)W_3^- \in \mathcal{FM}(M_3(\mathbb{C}))$ if and only if $\lambda \geq 2/27$.*
- *[HM15, Corollary 4.11]: For $0 \leq \lambda \leq 1$, $\lambda W_3^+ + (1-\lambda)W_3^- \in \text{conv}(\text{Aut}(M_3(\mathbb{C})))$ if and only if $\lambda \geq 1/3$.*

Notation 4.2. We denote

$$\Phi = \frac{2}{27}W_3^+ + \frac{25}{27}W_3^-.$$

By Proposition 4.1, $\Phi \in \mathcal{FM}(M_3(\mathbb{C})) \setminus \text{conv}(\text{Aut}(M_3(\mathbb{C})))$. Moreover by [HM15] (Theorem 5.6), it has an exact factorization through $M_3(\mathbb{C}) \otimes M_3(\mathbb{C})$. Hence it is our candidate for a map in $\mathcal{FM}_{\text{mat}}(M_3(\mathbb{C})) \setminus \text{conv}(\text{Aut}(M_3(\mathbb{C})))$.

Definition 4.3 (twirling map)**.** ([HM15] Definition 4.1) The map $F : \mathcal{B}(M_n(\mathbb{C})) \to \mathcal{B}(M_n(\mathbb{C}))$

$$T \mapsto \int_{\text{U}(n)} \text{Ad}_U\, T\, \text{Ad}_{U^t}\ dU$$

is called the twirling map, where dU denotes the Haar measure on the unitary group $\text{U}(n)$. (For background information of Haar measure on a compact group, see Chapter 5 Section 11 in [Con85].)

Although the name twirling map is an intuitive reflection of the nature of the map F, it is entirely not obvious that the dimension of the image of F is significantly reduced to a constant two. Essentially, this is a consequence of the Schur-Weyl duality in representation theory, on which we will not expand in this thesis.

Proposition 4.2. *We summarize some useful facts about the twirling map F stated in [HM15]:*

- $F(\mathrm{conv}(\mathrm{Aut}(M_n(\mathbb{C})))) \subseteq \mathrm{conv}(\mathrm{Aut}(M_n(\mathbb{C})))$ *(Proposition 4.2);*
- *F is a projection of $\mathcal{B}(M_n(\mathbb{C}))$ onto the subspace spanned by W_n^+ and W_n^-;*
- *F maps factorizable maps to the line segment spanned by W_n^+ and W_n^- (after Lemma 4.4 before Theorem 4.5).*

Proof. We will only provide proof for the second item in this proposition (see Lemma 4.4). $\square$

Definition 4.4 (Jamiolkowski transform)**.** ([HM15] Lemma 4.3) The map

$$\begin{aligned} \mathcal{B}(M_n(\mathbb{C})) &\to M_n(\mathbb{C}) \otimes M_n(\mathbb{C}) \\ T &\mapsto \widehat{T} = \frac{1}{n}\sum_{j,k=1}^{n} T(e_{jk}) \otimes e_{jk} \end{aligned}$$

is called the Jamiolkowski transform.

It turns out that if we normalize traces carefully, the Jamiolkowski transform will preserve the induced inner products in $\mathcal{B}(M_n(\mathbb{C}))$ and $M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$:

Lemma 4.3. *For arbitrary $S, T \in \mathcal{B}(M_n(\mathbb{C}))$, we have*

$$\left\langle \widehat{S}, \widehat{T} \right\rangle_{\mathrm{Tr}_n \otimes \mathrm{Tr}_n} = \langle S, T\rangle_{\mathrm{tr}_{\mathcal{B}(M_n(\mathbb{C}))}} = \frac{1}{n^2}\sum_{b \in \mathcal{B}_n} \langle S(b), T(b)\rangle_{\mathrm{Tr}_n}.$$

Proof. We simply do the explicit computation:

$$\begin{aligned} \left\langle \widehat{S}, \widehat{T} \right\rangle_{\mathrm{Tr}_n \otimes \mathrm{Tr}_n} &= \left\langle \frac{1}{n}\sum_{j,k=1}^{n} S(e_{jk}) \otimes e_{jk}\ ,\ \frac{1}{n}\sum_{j',k'=1}^{n} T(e_{j'k'}) \otimes e_{j'k'} \right\rangle_{\mathrm{Tr}_n \otimes \mathrm{Tr}_n} \\ =&\frac{1}{n^2}\sum_{j,k,j',k'=1}^{n} \langle S(e_{jk}), T(e_{j'k'})\rangle_{\mathrm{Tr}_n} \langle e_{jk}, e_{j'k'}\rangle_{\mathrm{Tr}_n} = \frac{1}{n^2}\sum_{j,k=1}^{n} \langle S(e_{jk}), T(e_{jk})\rangle_{\mathrm{Tr}_n}, \end{aligned}$$

where the last equality is because $\mathcal{B}_n = \{e_{jk}\}_{j,k=1}^n$ is an orthonormal basis of $M_n(\mathbb{C})$ with $\langle\cdot,\cdot\rangle_{\mathrm{Tr}_n}$. Also,

$$\begin{aligned} \langle S, T\rangle_{\mathrm{tr}_{\mathcal{B}(M_n(\mathbb{C}))}} &= \mathrm{tr}_{\mathcal{B}(M_n(\mathbb{C}))}(T^*S) \\ =&\frac{1}{n^2}\sum_{j,k=1}^{n} \langle T^*S(e_{jk}), e_{jk}\rangle_{\mathrm{Tr}_n} = \frac{1}{n^2}\sum_{j,k=1}^{n} \langle S(e_{jk}), T(e_{jk})\rangle_{\mathrm{Tr}_n}. \end{aligned}$$

Hence the lemma is proved. $\square$

Moreover, we observe the following: if we plug in the same operator $T \in \mathcal{B}(M_n(\mathbb{C}))$, we get that the Jamiolkowski transform is an isometry between $\mathcal{B}(M_n(\mathbb{C}))$ and $M_n(\mathbb{C}) \otimes M_n(\mathbb{C})$ with the induced norms. In particular, this implies it is injective. Because it is also a linear transformation between vector spaces of the same dimension n^4, the Jamiolkowski transform is a linear isomorphism. We proceed to prove that the twirling map F is an orthogonal projection.

Lemma 4.4. *If we denote $p^+ = \frac{n(n+1)}{2}\widehat{W_n^+}$ and $p^- = \frac{n(n-1)}{2}\widehat{W_n^-}$, E the conditional expectation (trace-preserving orthogonal projection) $M_n(\mathbb{C}) \otimes M_n(\mathbb{C}) \to \mathbb{C}p^+ + \mathbb{C}p^-$, then the following diagram commutes:*

$$\begin{array}{ccc}\mathcal{B}(M_n(\mathbb{C})) & \xrightarrow{\ \widehat{\ }\ } & M_n(\mathbb{C})\otimes M_n(\mathbb{C}) \\ \Big\downarrow F & & \Big\downarrow E \\ \mathcal{B}(M_n(\mathbb{C})) & \xrightarrow{\ \widehat{\ }\ } & M_n(\mathbb{C})\otimes M_n(\mathbb{C})\end{array}$$

and hence, F is the orthogonal projection of $\mathcal{B}(M_n(\mathbb{C}))$ onto the linear span of W_n^+ and W_n^-.

Proof. We first point out that p^+, p^-, and E were introduced and defined differently in [HM15] (Equation 4.2 and 4.4) but later proved to correspond with $\widehat{W_n^+}$, $\widehat{W_n^-}$, and a conditional expectation, respectively in the way stated above. In particular, p^+ and p^- are projections that add up to the identity, so E is indeed a conditional expectation. In this thesis, we choose to define those objects directly so that it is easiest to build upon their results and suit our needs.

The fact that $\widehat{F(T)} = E(\widehat{T})$ is proved in [HM15] (Lemma 4.4). We proceed to prove F is an orthogonal projection, which comes from the fact that E is an orthogonal projection and the Jamiolkowski transform is a linear isomorphism that keeps inner products. More specifically: let $T \in \mathcal{B}(M_n(\mathbb{C}))$ and $S \in \mathbb{C}W_n^+ + \mathbb{C}W_n^-$ be arbitrary, then by the previous Lemma 4.3,

$$\langle T, S\rangle_{\mathrm{tr}_{\mathcal{B}(M_n(\mathbb{C}))}} = \left\langle \widehat{T}, \widehat{S}\right\rangle_{\mathrm{Tr}_n\otimes \mathrm{Tr}_n} = \left\langle E(\widehat{T}), \widehat{S}\right\rangle_{\mathrm{Tr}_n\otimes \mathrm{Tr}_n} = \left\langle \widehat{F(T)}, \widehat{S}\right\rangle_{\mathrm{Tr}_n\otimes \mathrm{Tr}_n} = \langle F(T), S\rangle_{\mathrm{tr}_{\mathcal{B}(M_n(\mathbb{C}))}}.$$

Hence F is the orthogonal projection $\mathcal{B}(M_n(\mathbb{C})) \to \mathbb{C}W_n^+ + \mathbb{C}W_n^-$. $\square$

Proposition 4.5. *Let $N = W_3^- - W_3^+$, $\Phi = \frac{2}{27}W_3^+ + \frac{25}{27}W_3^-$, then we have*

$$\sup_{K\in\mathcal{K}} \langle K, N\rangle_{\mathcal{H}} < \langle \Phi, N\rangle_{\mathcal{H}}.$$

Proof. Note that $\Phi \in \mathbb{C}W_3^+ + \mathbb{C}W_3^-$, and by Proposition 4.1, $\Phi \in \mathcal{FM}_{\mathrm{mat}}(M_3(\mathbb{C}))\setminus \mathrm{conv}(\mathrm{Aut}(M_3(\mathbb{C})))$, so $\Phi \in \mathcal{H}\setminus\mathcal{K}$. We claim the unique closest point to Φ in the closed convex subset $\mathcal{K}$ is

$$P_{\mathcal{K}}(\Phi) = \frac{1}{3}W_3^+ + \frac{2}{3}W_3^-.$$

To see this, we first compute, for arbitrary $K \in \mathcal{K}$,

$$\begin{aligned}\|K-\Phi\|_{\mathcal{H}}^2 &= \|(K - F(K) + F(K)) - (\Phi - F(\Phi) + F(\Phi))\|_{\mathcal{H}}^2 \\ &= \|(K - F(K)) - (\Phi - F(\Phi))\|_{\mathcal{H}}^2 + \|F(\Phi) - F(K)\|_{\mathcal{H}}^2 \\ &= \|K - F(K)\|^2 + \|\Phi - F(K)\|_{\mathcal{H}}^2,\end{aligned}$$

because F is an orthogonal projection onto $\mathbb{C}W_3^+ + \mathbb{C}W_3^-$ and $\Phi \in \mathbb{C}W_3^+ + \mathbb{C}W_3^-$. Hence we know $P_{\mathcal{K}}(\Phi)$ is in the line segment spanned by W_3^+ and W_3^-, because if otherwise, then

$$\|F(P_{\mathcal{K}}(\Phi)) - \Phi\|_{\mathcal{H}}^2 < \|P_{\mathcal{K}}(\Phi) - F(P_{\mathcal{K}}(\Phi))\|_{\mathcal{H}}^2 + \|F(P_{\mathcal{K}}(\Phi)) - \Phi\|_{\mathcal{H}}^2 = \|P_{\mathcal{K}}(\Phi) - \Phi\|_{\mathcal{H}}^2.$$

Since we know $P_{\mathcal{K}}(\Phi) \in \mathcal{K}$ therefore factorizable, then by Proposition 4.2, $F(P_{\mathcal{K}}(\Phi)) \in \mathcal{K}$, and $F(P_{\mathcal{K}}(\Phi))$ is in the line segment spanned by W_3^+ and W_3^-. Hence the above inequality will contradict $P_{\mathcal{K}}(\Phi)$ being closest to Φ in $\mathcal{K}$. Finally by Proposition 4.1, in the line segment spanned by W_3^+ and W_3^- ($\lambda W_3^+ + (1-\lambda)W_3^-$ as λ goes from 0 to 1), $\frac{1}{3}W_3^+ + \frac{2}{3}W_3^-$ is the first map in $\mathcal{K}$, hence proving the claim.

Now by a simplified version of the Hahn-Banach separation Theorem A.3, we have

$$\sup_{K\in\mathcal{K}} \langle K, \Phi - P_{\mathcal{K}}(\Phi)\rangle_{\mathcal{H}} < \langle \Phi, \Phi - P_{\mathcal{K}}(\Phi)\rangle_{\mathcal{H}},$$

but note that $N = W_3^- - W_3^+$ is just a positive scalar multiple of $\Phi - P_{\mathcal{K}}(\Phi) = \frac{7}{27}W_3^- - \frac{7}{27}W_3^+$ (to make later calculations easier), so the proposition is proved. $\square$

Before finally constructing our matrix tuples, we need this lemma giving us a linear isomorphism $\rho : \mathcal{H} \to (M_3(\mathbb{C})_{\mathrm{sa}} \otimes_{\mathbb{R}} M_3(\mathbb{C})_{\mathrm{sa}})^*$.

Lemma 4.6. *Let V and W be finite-dimensional $\mathbb{R}$-vector spaces, equipped with inner products $\langle\cdot,\cdot\rangle_V$ and $\langle\cdot,\cdot\rangle_W$ respectively. We denote $\dim(V) = n$ and $\dim(W) = m$. Then the map $\rho : L(V,W) \to (V \otimes W)^*$*

$$\rho(\Psi)(v \otimes w) = \langle \Psi(v), w\rangle_W$$

is a linear isomorphism.

Proof. ρ is clearly a linear homomorphism because $\langle\cdot,\cdot\rangle_W$ is bilinear, so linear in the first component. Note that since $L(V,W)$ and $(V \otimes W)^*$ both have dimension nm, it suffices to show ρ is an injective.

Let $\rho(\Psi) \in (V \otimes W)^*$ be the zero map, so $\rho(\Psi)(v \otimes w) = \langle \Psi(v), w\rangle_W = 0$ for all $v \in V$ and $w \in W$. Toward contradiction, assume Ψ is not the zero map in $L(V,W)$, then there exists $v \in V$ such that $\Psi(v) \neq 0$. This implies $\rho(\Psi)(v \otimes \Psi(v)) = \langle \Psi(v), \Psi(v)\rangle_W = \|\Psi(v)\|_W^2 > 0$, contradictory to $\rho(\Psi) \in (V \otimes W)^*$ being the zero map. This proves injectivity. $\square$

Theorem 4.7. *We have a concrete example of matrix tuples $X, Y \in M_3(\mathbb{C})^9_{\mathrm{sa}}$ such that*

$$\sup_{\Psi \in \mathrm{conv}(\mathrm{Aut}(M_3(\mathbb{C})))} \langle \Psi(X), Y\rangle_{\mathrm{tr}_3} < \langle \Phi(X), Y\rangle_{\mathrm{tr}_3} = \sup_{\Psi \in \mathcal{FM}(M_3(\mathbb{C}))} \langle \Psi(X), Y\rangle_{\mathrm{tr}_3},$$

with

$$\Phi = \frac{2}{27} W_3^+ + \frac{25}{27} W_3^-,$$

as defined in Notation 4.2. They are

$$\begin{aligned}
X_1 &= e_{11}, & Y_1 &= \frac{1}{3}1_3 - e_{11},\\
X_2 &= e_{22}, & Y_2 &= \frac{1}{3}1_3 - e_{22},\\
X_3 &= e_{33}, & Y_3 &= \frac{1}{3}1_3 - e_{33},\\
X_4 &= \frac{1}{\sqrt{2}}(e_{12} + e_{21}), & Y_4 &= -\frac{1}{\sqrt{2}}(e_{12} + e_{21}),\\
X_5 &= \frac{1}{\sqrt{2}}(e_{13} + e_{31}), & Y_5 &= -\frac{1}{\sqrt{2}}(e_{13} + e_{31}),\\
X_6 &= \frac{1}{\sqrt{2}}(e_{23} + e_{32}), & Y_6 &= -\frac{1}{\sqrt{2}}(e_{23} + e_{32}),\\
X_7 &= \frac{i}{\sqrt{2}}(e_{12} - e_{21}), & Y_7 &= \frac{i}{\sqrt{2}}(e_{12} - e_{21}),\\
X_8 &= \frac{i}{\sqrt{2}}(e_{13} - e_{31}), & Y_8 &= \frac{i}{\sqrt{2}}(e_{13} - e_{31}),\\
X_9 &= \frac{i}{\sqrt{2}}(e_{23} - e_{32}), & Y_9 &= \frac{i}{\sqrt{2}}(e_{23} - e_{32}).
\end{aligned}$$

Proof. Define $v : \mathcal{H} \to M_3(\mathbb{C})_{\mathrm{sa}} \otimes_{\mathbb{R}} M_3(\mathbb{C})_{\mathrm{sa}}$,

$$S \mapsto \sum_{b \in \mathcal{B}_{3,\mathrm{sa}}} b \otimes S(b).$$

Then we have

$$\begin{aligned}
&\sup_{\Psi \in \mathcal{K}} \rho(\Psi)(v(N)) = \sup_{\Psi \in \mathcal{K}} \sum_{b \in \mathcal{B}_{3,\mathrm{sa}}} \langle \Psi(b), N(b)\rangle_{\mathrm{Tr}_3} = \sup_{\Psi \in \mathcal{K}} \sum_{b \in \mathcal{B}_{3,\mathrm{sa}}} \langle N^*\Psi(b), b\rangle_{\mathrm{Tr}_3} = \sup_{\Psi \in \mathcal{K}} \langle \Psi, N\rangle_{\mathcal{H}}\\
<&\langle \Phi, N\rangle_{\mathcal{H}} = \sum_{b \in \mathcal{B}_{3,\mathrm{sa}}} \langle N^*\Phi(b), b\rangle_{\mathrm{Tr}_3} = \sum_{b \in \mathcal{B}_{3,\mathrm{sa}}} \langle \Phi(b), N(b)\rangle_{\mathrm{Tr}_3} = \rho(\Phi)(v(N)).
\end{aligned}$$

We want to construct $X, Y \in M_3(\mathbb{C})^m_{\text{sa}}$ such that

$$\langle H(X), Y\rangle_{\text{Tr}_3} = \sum_{j=1}^{m} \langle H(X_j), Y_j\rangle_{\text{Tr}_3} = \sum_{b\in\mathcal{B}_{3,\text{sa}}} \langle H(b), N(b)\rangle_{\text{Tr}_3} = \rho(H)(v(N)) \qquad \forall H \in \mathcal{H}.$$

Then by multiplying 1/3 to normalize trace, we get the strict separation

$$\sup_{\Psi\in\text{conv}(\text{Aut}(M_3(\mathbb{C})))} \langle \Psi(X), Y\rangle_{\text{tr}} < \langle \Phi(X), Y\rangle_{\text{tr}}.$$

Hence we let $m = |\mathcal{B}_{3,\text{sa}}| = 9$. Fix an ordering of $\mathcal{B}_{3,\text{sa}} = \{b_1, \dots, b_9\}$, and let $X_j = b_j, Y_j = N(b_j)$ for $1 \le j \le 9$ to make the equality hold for all $H \in \mathcal{H}$. If we do the computation explicitly, we will get a concrete example as claimed (up to a normalization constant to make it look nice). We showcase three representative items:

$$\begin{aligned} N(x) &= W_3^-(x) - W_3^+(x) \\ &= \frac{1}{3-1}\left(\text{Tr}_3(x)1_3 - x^t\right) - \frac{1}{3+1}\left(\text{Tr}_3(x)1_3 + x^t\right) = \frac{1}{4}\text{Tr}_3(x)1_3 - \frac{3}{4}x^t, \end{aligned}$$

$$\begin{aligned} N(e_{jj}) &= \frac{1}{4}1_3 - \frac{3}{4}e_{jj}, \\ N\left(\frac{1}{\sqrt{2}}(e_{jk}+e_{kj})\right) &= -\frac{3}{4\sqrt{2}}(e_{jk}+e_{kj}), \\ N\left(\frac{i}{\sqrt{2}}(e_{jk}-e_{kj})\right) &= \frac{3i}{4\sqrt{2}}(e_{jk}-e_{kj}), \end{aligned}$$

and the normalization is multiplying by 4/3.

Now for the equality, let $\Psi \in \mathcal{FM}(M_3(\mathbb{C}))$ be arbitrary, then since F is the projection onto $\mathbb{C}W_3^+ + \mathbb{C}W_3^-$, and $N \in \mathbb{C}W_3^+ + \mathbb{C}W_3^-$,

$$\begin{aligned} \langle \Phi, N\rangle_{\mathcal{H}} - \langle \Psi, N\rangle_{\mathcal{H}} &= \langle \Phi, N\rangle_{\mathcal{H}} - \langle F(\Psi), N\rangle_{\mathcal{H}} \\ &= \left\langle \frac{2}{27}W_3^+ + \frac{25}{27}W_3^-, N\right\rangle_{\mathcal{H}} - \left\langle \lambda W_3^+ + (1-\lambda)W_3^-, N\right\rangle_{\mathcal{H}} \\ &= \left\langle \left(\lambda - \frac{2}{27}\right)W_3^- - \left(\lambda - \frac{2}{27}\right)W_3^+, N\right\rangle_{\mathcal{H}} \\ &= \left(\lambda - \frac{2}{27}\right)\|N\|^2_{\mathcal{H}} \ge 0, \end{aligned}$$

where the last inequality is from [HM15] Theorem 5.6(2). Namely, if we write $F(\Psi) = \lambda W_3^+ + (1-\lambda)W_3^-$, then we have $2/27 \le \lambda \le 1$. Taking sup over $\mathcal{FM}(M_3(\mathbb{C}))$ and using $\Phi \in \mathcal{FM}(M_3(\mathbb{C}))$, we get

$$\langle \Phi(X), Y\rangle_{\text{tr}_3} = \sup_{\Psi\in\mathcal{FM}(M_3(\mathbb{C}))} \langle \Psi(X), Y\rangle_{\text{tr}_3}.$$

□

Acknowledgement

I am greatly indebted to Professor Todd Kemp for teaching me probability and introducing me to random matrix theory and free probability. I also want to offer my sincere gratitude to Dr. David Jekel for teaching me operator algebra. In our numerous meetings, their expertise, guidance, patience, and detailed comments helped me work on this fun research topic and produce this thesis. At last, I thank my family and especially my mom Julie for the constant support.

A Appendix

A.1 Hilbert projection theorem

Theorem A.1. *[Kem21, Lecture 31.1] Let $\mathcal{H}$ be a Hilbert space, $\mathcal{K} \subseteq \mathcal{H}$ be a closed convex subset. Let $x \in \mathcal{H}$, then $\exists! y \in \mathcal{K}$ such that $\|x-y\|^2 = d(x,\mathcal{K})^2 := (\inf_{z\in\mathcal{K}}\|x-z\|)^2$. Moreover, if $\mathcal{K}$ is also a Hilbert subspace, then y is also the unique element in $\mathcal{K}$ such that $x-y \in \mathcal{K}^\perp$.*

Proof. For arbitrary $y, z \in \mathcal{K}$,
$\|y-z\|^2 = \|(y-x)-(z-x)\|^2 = 2(\|y-x\|^2 + \|z-x\|^2) - \|(y-x)+(z-x)\|^2$ by parallelogram law. Note $\|(y-x)+(z-x)\|^2 = 4\left\|\frac{y+z}{2}-x\right\|^2$, and $\frac{y+z}{2} \in \mathcal{K}$ so $\left\|\frac{y+z}{2}-x\right\|^2 \geq d(x,\mathcal{K})^2$. Therefore $\|y-z\|^2 + 4d(x,\mathcal{K})^2 \leq \|y-z\|^2 + 4\left\|\frac{y+z}{2}-x\right\|^2 = 2(\|y-x\|^2 + \|z-x\|^2)$.

Existence: let $y_n \in \mathcal{K}$ with $\|x-y_n\|^2 \leq d(x,\mathcal{K})^2 + 1/n$, then $\|y_n - y_m\|^2 + 4d(x,\mathcal{K})^2 \leq 2(\|y_n-x\|^2 + \|y_m - x\|^2)$ so $\|y_n-y_m\|^2 \leq 2(1/n+1/m) \to 0$ as $n, m \to \infty$. Hence $\{y_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in the Hilbert space $\mathcal{H}$, so $y_n \to y \in \mathcal{H}$. y is also in $\mathcal{K}$ because $\mathcal{K}$ is closed. Finally, $d(x,\mathcal{K}) \leq \|x-y\| \leq \|x-y_n\| + \|y_n - y\| \to d(x,\mathcal{K})$, so $\|x-y\| = d(x,\mathcal{K})$.

Uniqueness: if $d(x,\mathcal{K}) = \|x-y\|^2 = \|x-z\|^2$, then $\|y-z\|^2 \leq 0$ so $y=z$.

Now for the moreover part, we let $y \in \mathcal{K}$ be the unique closest point to $x \in \mathcal{H}$. For arbitrary $z \in \mathcal{K}$, consider $t \mapsto \alpha(t) := \|x-(y+tz)\|^2 = \|x-y\|^2 - 2t\operatorname{Re}\langle x-y, z\rangle + t^2\|z\|^2$ for $t \in \mathbb{R}$. By definition of y, $\min\alpha = \alpha(0)$ so $0 = \alpha'(0) = -2\operatorname{Re}\langle x-y,z\rangle$, hence $\operatorname{Re}\langle x-y,z\rangle = 0$. Similarly, by considering $t \mapsto \beta(t) := \|x-(y+itz)\|^2 = \|x-y\|^2 - 2t\operatorname{Im}\langle x-y,z\rangle + t^2\|z\|^2$ for $t \in \mathbb{R}$, we have $\operatorname{Im}\langle x-y,z\rangle = 0$. Therefore $\langle x-y,z\rangle = 0$. Conversely, let $y \in \mathcal{K}$ with $x - y \in \mathcal{K}^\perp$, then for all $z \in \mathcal{K}$, $\|x-z\|^2 = \|(x-y)+(y-z)\|^2 = \|x-y\|^2 + \|y-z\|^2 \geq \|x-y\|^2$ since $x-y \in \mathcal{K}^\perp$ and $y - z \in \mathcal{K}$ implies $\langle x-y, y-z\rangle = 0 = \langle y-z, x-y\rangle$. Note equality holds if and only if $z = y$. Therefore $y \in \mathcal{K}$ is the unique closest point to $x \in \mathcal{H}$. $\square$

The consequence of this theorem is $\mathcal{H} = \mathcal{K} \oplus \mathcal{K}^\perp$ if $\mathcal{K} \subseteq \mathcal{H}$ is a Hilbert subspace (every $x \in \mathcal{H}$ uniquely decomposes into $x_1 + x_2$ where $x_1 \in \mathcal{K}$ and $x_2 \in \mathcal{K}^\perp$).

We conclude with a discussion of orthogonal projections in a Hilbert space.

Theorem A.2. *[Kem21, Lecture 31.1] Let $\mathcal{H}$ be a Hilbert space and $\mathcal{K} \subseteq \mathcal{H}$ be a closed subspace. Then there exists a unique linear surjection $P_\mathcal{K} : \mathcal{H} \to \mathcal{K}$ such that, for $x \in \mathcal{H}$:*

1. *$P_\mathcal{K}(x)$ is the unique element in $\mathcal{K}$ such that $\langle y, P_\mathcal{K}(x)\rangle = \langle y, x\rangle$ for all $y \in \mathcal{K}$.*
2. *$P_\mathcal{K}(x) = x$ if $x \in \mathcal{K}$ and $P_\mathcal{K}(x) = 0$ if $x \in \mathcal{K}^\perp$.*
3. *$P_\mathcal{K}$ is a self-adjoint idempotent: $P_\mathcal{K}^2 = P_\mathcal{K}$ and $P_\mathcal{K}^* = P_\mathcal{K}$.*
4. *$P_\mathcal{K}(x)$ is the unique element in $\mathcal{K}$ that is the closest to x.*
5. *if $\mathcal{L} \subseteq \mathcal{K} \subseteq \mathcal{H}$ are closed subspaces, then $P_\mathcal{L}P_\mathcal{K} = P_\mathcal{K}P_\mathcal{L}$.*
6. *$P_\mathcal{K}$ is Lip_1-continuous.*

We call $P_\mathcal{K}$ the orthogonal projection of $\mathcal{H}$ onto $\mathcal{K}$.

Proof. we prove these one by one:

1. Fix $x \in \mathcal{H}$, then the map $y \mapsto \langle y, x\rangle$ is in $\mathcal{K}^*$, so by Riesz representation theorem, $\exists! P_\mathcal{K}(x) \in \mathcal{K}$ such that $\langle y, x\rangle = \langle y, P_\mathcal{K}(x)\rangle$ for all $y \in \mathcal{K}$. This also proves $P_\mathcal{K}$ is linear since inner product is a sesquilinear form.
2. This follows from the uniqueness in the Riesz representation. This also proves $P_\mathcal{K}$ is surjective.
3. The first equality follows immediately from item 2. For the second equality, let $x, y \in \mathcal{H}$ be arbitrary. We compute $\langle P_\mathcal{K}(x), y\rangle = \langle P_\mathcal{K}(x), P_\mathcal{K}(y)\rangle + \langle P_\mathcal{K}(x), y - P_\mathcal{K}(y)\rangle = \langle P_\mathcal{K}(x), P_\mathcal{K}(y)\rangle = \langle P_\mathcal{K}(x), P_\mathcal{K}(y)\rangle + \langle x - P_\mathcal{K}(x), P_\mathcal{K}(y)\rangle = \langle x, P_\mathcal{K}(y)\rangle$.
4. This follows from $P_\mathcal{K}(x)$ being the unique element in $\mathcal{K}$ such that $x - P_\mathcal{K}(x) \in \mathcal{K}^\perp$ and theorem A.1.

5. We first observe $P_{\mathcal{L}}(x) \in \mathcal{L} \subseteq \mathcal{K}$, so $P_{\mathcal{K}}P_{\mathcal{L}}(x) = P_{\mathcal{L}}(x)$. Then we compute

$$\langle P_{\mathcal{L}}P_{\mathcal{K}}(x), y\rangle = \langle P_{\mathcal{K}}(x), P_{\mathcal{L}}(y)\rangle = \langle x, P_{\mathcal{K}}P_{\mathcal{L}}(y)\rangle = \langle x, P_{\mathcal{L}}(y)\rangle = \langle P_{\mathcal{L}}(x), y\rangle \qquad \forall x, y \in \mathcal{H}.$$

Hence in particular, take $y = P_{\mathcal{L}}P_{\mathcal{K}}(x) - P_{\mathcal{L}}(x)$, then

$$\langle P_{\mathcal{L}}P_{\mathcal{K}}(x) - P_{\mathcal{L}}(x), P_{\mathcal{L}}P_{\mathcal{K}}(x) - P_{\mathcal{L}}(x)\rangle = 0 \qquad \forall x \in \mathcal{H},$$

so $P_{\mathcal{L}}P_{\mathcal{K}}(x) = P_{\mathcal{L}}(x) = P_{\mathcal{K}}P_{\mathcal{L}}(x)$.

6. Let $x \in \mathcal{H}$ be arbitrary and write as $x = x - P_{\mathcal{K}}(x) + P_{\mathcal{K}}(x)$. By othogonality, $\|x\|^2 = \|x - P_{\mathcal{K}}(x)\|^2 + \|P_{\mathcal{K}}(x)\|^2$, so $\|x\|^2 \geq \|P_{\mathcal{K}}(x)\|^2$. Hence for arbitrary $x, y \in \mathcal{H}$,

$$\|P_{\mathcal{K}}(x) - P_{\mathcal{K}}(y)\| = \|P_{\mathcal{K}}(x - y)\| \leq \|x - y\|.$$

$\square$

A.2 Hahn-Banach separation theorem

Theorem A.3. *Let X be a locally convex topological vector space over $\mathbb{R}$. If $a \in X$ and $B \subseteq X$ satisfies:*

- *$a \notin B$,*
- *B is closed and convex,*

then a and B are strictly separated: there exists a continuous linear functional $f : X \to \mathbb{R}$ and $s, t \in \mathbb{R}$ such that $\sup_{b \in B} f(b) < f(a)$ for all $b \in B$. Moreover if X is a Hilbert space with a real inner product, then f can be given by $x \mapsto \langle x, a - P_B(a)\rangle$, where $P_B(a)$ denotes the projection of a onto B: the unique point in B closest to a.

Proof. The general theorem holds for infinite-dimensional topological vector spaces X, following from the well-known Hahn-Banach theorem in functional analysis (for example see [Con85] Chapter 3 Theorem 6.2). Thus in this thesis, we only prove the moreover part, which is what we need in section 4.2.

Since B is a closed and convex subset in the Hilbert space X, by the Hilbert projection theorem A.1, $P_B(a)$ is well-defined with the properties stated above. Now let $b \in B$ be arbitrary and $0 < t \leq 1$,

$$\begin{aligned}\|a - P_B(a)\|^2 &\leq \|a - [(1-t)P_B(a) + tb]\|^2 = \|t(P_B(a) - b) + (a - P_B(a)\|^2 \\ &= \|a - P_B(a)\|^2 + \|P_B(a) - b\|^2 t^2 + 2t\langle P_B(a) - b, a - P_B(a)\rangle,\end{aligned}$$

where the inequality is by definition of $P_B(a)$ and $(1-t)P_B(a) + tb \in B$. This implies

$$\|P_B(a) - b\|^2 t + 2\langle P_B(a) - b, a - P_B(a)\rangle \geq 0,$$

taking the limit as $t \to 0$, we get $\langle P_B(a) - b, a - P_B(a)\rangle \geq 0$, hence

$$\langle b, a - P_B(a)\rangle \leq \langle P_B(a), a - P_B(a)\rangle \quad \forall b \in B.$$

In addition, $a \notin B$ so $\langle a - P_B(a), a - P_B(a)\rangle = \|a - P_B(a)\|^2 > 0$. This gives us the full inequality:

$$\begin{aligned}\langle b, a - P_B(a)\rangle \leq& \langle P_B(a), a - P_B(a)\rangle \\ <& \frac{1}{2}\|a - P_B(a)\|^2 + \langle P_B(a), a - P_B(a)\rangle \\ <& \langle a, a - P_B(a)\rangle \quad \forall b \in B.\end{aligned}$$

Taking sup over B, we have

$$\sup_{b \in B}\langle b, a - P_B(a)\rangle \leq \frac{1}{2}\|a - P_B(a)\|^2 + \langle P_B(a), a - P_B(a)\rangle < \langle a, a - P_B(a)\rangle.$$

$\square$

References

[AD06] Claire Anantharaman-Delaroche. On ergodic theorems for free group actions on noncommutative spaces. *Probab. Theory Rel. Fields*, 135:520–546, 2006.

[Bro88] J. B. Bronzan. Parameterization of su(3). *Phys. Rev. D*, 38(6), 1988.

[BV01] Philippe Biane and Dan-Virgil Voiculescu. A free probability analogue of the Wasserstein metric on the trace-state space. *Geometric and Functional Analysis*, 11:1125–1138, 2001.

[Con85] John B. Conway. *A course in functional analysis.* Springer Science+Business Media, LLC, 1985.

[Dix69] Jacques Dixmier. *Les C^*-algèbres et leurs représentations*, volume 29 of *Cahiers Scientifiques.* Gauthier-Villars, Paris, 2 edition, 1969. Reprinted by Editions Jacques Gabay, Paris, 1996. Translated as C^*-algebras, North-Holland, Amsterdam, 1977. First Edition 1964.

[GJNS21] Wilfrid Gangbo, David Jekel, Kyeongsik Nam, and Dimitri Shlyakhtenko. Duality for optimal couplings in free probability, 2021.

[Ham15] Mitsuru Hamada. On parameterizations of rotations. *Tamagawa University Quantum ICT Research Institute Bulletin*, 5(1):25–28, 2015.

[HM11] Uffe Haagerup and Magdalena Musat. Factorization and dilation problems for completely positive maps on von neumann algebras. *Communications in Mathematical Physics*, 303(2):555–594, 2011.

[HM15] Uffe Haagerup and Magdalena Musat. An asymptotic property of factorizable completely positive maps and the connes embedding problem. *Communications in Mathematical Physics*, 338(2):721–752, 2015.

[Kem21] Todd Kemp. Probability theory and stochastic processes, a full year graduate course, 2021.

[KR83] Richard V. Kadison and John R. Ringrose. *Fundamentals of the Theory of Operator Algebras I*, volume 15 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, 1983.

[Kü85] Burkhard Kümmerer. Markov dilations on W*-algebras. *Journal of Functional Analysis*, 63(2):139–177, 1985.

[MR19] Magdalena Musat and Mikael Rørdam. Non-closure of quantum correlation matrices and factorizable channels that require infinite dimensional ancilla (with an appendix by narutaka ozawa). *Communications in Mathematical Physics*, 375(3):1761–1776, 2019.

[Sak71] Shôichirô Sakai. C**-algebras and* W**-algebras*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete*. Springer-Verlag, Berlin Heidelberg, 1971.

[Tak02] Masamichi Takesaki. *Theory of Operator Algebras I*, volume 124 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin Heidelberg, 2002.

[VDN92] Dan-Virgil Voiculescu, Kenneth J. Dykema, and Alexandru Nica. *Free Random Variables*, volume 1 of *CRM Monograph Series.* American Mathematical Society, Providence, RI, 1992.

[vN27a] John von Neumann. Mathematische begründung der quantenmechanik [mathematical foundation of quantum mechanics]. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 1–57, 1927.

[vN27b] John von Neumann. Thermodynamik quantenmechanischer gesamtheiten [thermodynamics of quantum mechanical quantities]. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 273–291, 1927.

[vN27c] John von Neumann. Wahrscheinlichkeitstheoretischer aufbau der quantenmechanik [probabilistic theory of quantum mechanics]. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 245–272, 1927.

[Wil13] Mark M. Wilde. *Quantum Information Theory.* Cambridge University Press, 2013.

[Wit20] Edward Witten. A mini-introduction to information theory. *La Rivista del Nuovo Cimento*, 43:187–227, 2020.

[Zhu93] Kehe Zhu. *An Introduction to Operator Algebras.* Studies in Advanced Mathematics. CRC Press, Ann Arbor, 1993.